1 Lossy Distributed Source Coding Problem

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.

1.1 Berger Tung Inner Bound

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.

1.1 Berger Tung Inner Bound

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.

1.1 Berger Tung Inner Bound

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.

1.1 Berger Tung Inner Bound

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.

1.1 Berger Tung Inner Bound

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.

1.1 Berger Tung Inner Bound

The Berger Tung Inner bound uses the familiar random binning technique. A series of conditional distributions for a collection of message random variables $U_1, \ldots, U_M$ and reconstructions $\hat{Z}_1, \ldots, \hat{Z}_L$ are created according to the Markov chain relationships $(X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L, n) \sim p_{X_1, \ldots, X_M, Y, Z_1, \ldots, Z_L}$.

Figure 1: The general multiterminal lossy source coding problem.

The general multiterminal lossy source coding problem depicted in Fig. 1 is not solved, however we have inner and outer bounds. As we will discuss in section 1.3, these inner and outer bounds do not match on even very simple variants of this structure.

Important special cases of this problem include when $L = M$ and $Z_m = X_m$ (the lossy variant of Slepian-Wolf), as well as the case $L = 1$ which is known as the CEO problem.
• $\mathcal{E}^2$: For every $m \in \{1, \ldots, M\}$ there exists an $i_m \in \{1, \ldots, 2^N\bar{R}_m\}$ such that $(U^N_m(i_m), X^N_m) \in T^N_e(U_m, X_m)$ however $(\hat{U}^N_{i_m}(i), X^N_{i_m}, Y^N_m) \notin T^N_e(U_{1:M}, X_{1:M}, Y, Z_{1:L})$.

• $\mathcal{E}^3$: For every $m \in \{1, \ldots, M\}$ there exists an $i_m' \in \{1, \ldots, 2^N\bar{R}_m\}$ such that $(U^N_m(i_m'), X^N_m) \in T^N_e(U_m, X_m)$ and $(\hat{U}^N_{i_m}(i), X^N_{i_m}, Y^N_m) \in T^N_e(U_{1:M}, X_{1:M}, Y, Z_{1:L})$. However, for the set $A \subseteq \{1, \ldots, M\}$, $(U^N_A(i_A'), X^N_A) = (U^N_{i_m'}, X^N_{i_m'}) \in T^N_e(U_A, U_{A^c}, Y)$ for some $i_m' \neq i_m$, $\forall m \in A$ with $i_m' \in B_m(j_m)$ for each $m \in A$.

The first error event involves the same style of analysis as in the rate distortion proof:

$$
P[\mathcal{E}^1_m] = \frac{2^N\bar{R}_m}{\prod_{i=1}^M P[(U^N_m(i), X^N_m) \notin T^N_e(U_m, X_m)]} = \left(1 - \sum_{i \neq i_m} P[(U^N_m(i), X^N_m) \notin T^N_e(U_m, X_m)]\right)^{2^N\bar{R}_m}
$$

From this we observe that the probability of error will be arbitrarily small as $N \to \infty$ provided that

$$
\hat{R}_i > I(X_i; \hat{U}_i)
$$

Error $\mathcal{E}_2$ has a probability that is arbitrarily small as $N \to \infty$ via an extension of the Markov Lemma. Finally, turning our attention to the third collection of error events, we have

$$
P[\mathcal{E}^3] = \sum_{i_A', i_m' \neq i_m, \forall m \in A} \sum_{i_A', i_m' \neq i_m, \forall m \in A} P[(U^N_A(i_A'), U^N_{i_A'}(i_{A'}), Y^N) \in T^N_e(U_A, U_{A^c}, Y), i_m' \in B_m(j_m) \forall m \in A] \left(U^N_{i_m}(i), X^N_m, Y^N \right) \in T^N_e(U_{1:M}, X^N_{1:M}, Y^N)
$$

$$
\leq \sum_{i_A', i_m' \neq i_m, \forall m \in A} \sum_{i_A', i_m' \neq i_m, \forall m \in A} 2^{-N(I(U_{A^c}; Y|U_{A^c}) - \epsilon \epsilon)} \prod_{m \in A} 2^{-N\bar{R}_m} \leq 2^{-N(I(U_{A^c}; Y|U_{A^c}) - \epsilon \epsilon)} \prod_{m \in A} 2^{-N\bar{R}_m} (2^{N\bar{R}_m} - 1)
$$

$$
\leq 2 \left(\sum_{m \in A} \hat{R}_m - R_m - I(U_A; Y|U_{A^c}) + \epsilon \epsilon\right)
$$

This can be made arbitrarily small as $N \to \infty$ if

$$
\sum_{i \in A} \hat{R}_i - R_i - I(Y; U_A|U_{A^c})< 0
$$

Putting the two sets of inequalities together, we find that the requirements on the rates are

$$
\sum_{i \in A} R_i > I(X_A; U_{A^c}|Y, U_{A^c})
$$

When its union over all possible choices of auxiliary variables $U_{1:M}$ obeying the constraints is taken, this region may not be convex, but any two achievable strategies can be time shared to get a convex combination in $(R, D)$ space, hence can take a convex hull. One way to show this convex hull in the math is to introduce a time sharing variable $Q$, as shown in the text.

### 1.2 Berger Tung Outer Bound

The Berger Tung outer bound involves the same rate and distortion expressions as the Berger Tung inner bound, but uses changes the message Markov Chain conditions: $(X_{\gamma}, Y, Z_{1:L}) \leftrightarrow X_m \leftrightarrow U_m$. (Note that the other messages $U_{\gamma}$ are absent in the condition.)

### 1.3 Tightness & Other Bounds

As noted in the text, both the Berger Tung Inner and the Berger Tung Outer bounds have cases where they are not tight. There is a gap between the Berger Tung inner bound and the Berger Tung Outer Bound for the scalar Gaussian squared error CEO problem, and the Berger Tung inner bound was shown to give the rate distortion region in this case (Oohama ’05 and Prabhakaran, Tse, Ramachandran ’04). As discussed in 12.3, the Berger Tung inner bound is also tight for two Gaussian source and squared error distortions. Wagner and Anantharam have given a tighter outer bound than the Berger Tung Outer bound (2008). These problems are an area of active work in the community.
2 Multiple Descriptions Problem

Surprisingly, this problem is not solved. We presented a pair of achievable regions and discussed when one of them was tight.

2.1 El Gamal Cover Region

The El Gamal cover region splits each of the two descriptions $S_1, S_2$ up into two parts $S_1 = (S_{10}, S_{11})$ and $S_2 = (S_{20}, S_{22})$. The parts $S_{10}$ and $S_{20}$ are discarded by the decoders receiving only one description, and are concatenated into $S_0 = (S_{10}, S_{20})$ by the decoder receiving both descriptions.

- **Codebook Generation**: Generate a $2^{NR_1} \times N$ matrix with elements IID according to $p_{X_1}$, denote the $i$th row of this by $X_1^N(i)$. Similarly, generate a $2^{NR_2} \times N$ matrix with elements IID according to $p_{X_2}$, denote the $j$th row of this by $X_2^N(j)$. For each $(i,j) \in \{1, \ldots, 2^{NR_1}\} \times \{1, \ldots, 2^{NR_2}\}$ generate a $2^{NR_0} \times N$ matrix, with $k$th row $X_0^N(i,j,k)$, and with elements distributed according to the conditional distribution

$$\prod_{n=1}^{N} p_{X_0|X_1,X_2}(x_{0,n}(i,j,k)|x_{1,n}(i),x_{2,n}(j)) \quad (8)$$

Share these matrices with the encoder and decoder.

- **Encoder**: Select an $i, j, k$ such that $(X_1(i)^N, X_2^N(j), X_0^N(i,j,k)) \in T_e^N(X_1, X_2, X_0, X)$. Split the bits in the index $k$ up into two parts $k = (k_1, k_2)$, and transmit the messages $S_1 = (i,k_1)$ and $S_2 = (j,k_2)$.

- **Decoder**: The decoder receiving only $S_1 = (i,k_1)$ discards $k_1$ and reproduces $X_1^N(i)$. The decoder receiving only $S_2 = (j,k_2)$ discards $k_2$ and reproduces $X_2^N(j)$. The decoder receiving both $S_1, S_2$ reproduces $X_0^N(i,j,k)$.

Error events

- $\mathcal{E}_e$: $X^N \notin T_e^N(X)$
- $\mathcal{E}_1$: for every $i$ $(X_1^N(i), X^N) \notin T_e^n(X_1, X)$
- $\mathcal{E}_2$: for every $j$ $(X_2^N(j), X^N) \notin T_e^n(X_2, X)$
- $\mathcal{E}_{12}$: for every $i, j$ $(X_1^N(i), X_2^N(j), X^N) \notin T_e^n(X_1, X_2, X)$
- $\mathcal{E}_0$: there exists $i, j$ such that $(X_1^N(i), X_2^N(j), X^N) \in T_e^n(X_1, X_2, X)$, but for every $k \in \{1, \ldots, 2^{NR_0}\}$, $(X_1^N(i), X_2^N(j), X_0^N(i,j,k), X^N) \notin T_e^n(X_1, X_2, X_0, X)$

Properties of the typical set imply $\mathbb{P}[\mathcal{E}_e] \to 0$ as $N \to \infty$.

A now standard proof bounding the probability that two independent sequences with matching marginals to a joint distribution wind up in the typical set shows that if $R_{11} > I(X_1; X)$ and if $R_{22} > I(X_2; X)$, $\mathbb{P}[\mathcal{E}_1]$ and $\mathbb{P}[\mathcal{E}_2]$ go to 0 as $N \to \infty$. 

![Figure 2: The still generally unsolved multiple descriptions problem.](image-url)
The novel idea in the proof is the bounding of $P[E_{12}]$ via a clever application of Chebyshev’s inequality. In this vein, for each $x \in X$, define the random set

$$G_\mathbf{x} = \{ (i,j) \in \{1, \ldots, 2^{NR_{11}} \} \times \{1, \ldots, 2^{NR_{22}} \} \mid (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X) \}$$

and denote, naturally, the cardinality of this set by $|G_\mathbf{x}|$

Noting that the event $E_{12}$ is equivalent to the event $|G_\mathbf{x}| = 0$, and applying the Chebyshev’s inequality

$$P \left[ |X - E[X]| \geq \epsilon \sqrt{\text{var}(X)} \right] \leq \frac{1}{\epsilon^2},$$

which implies by selecting $\epsilon = \frac{\text{var}(X)}{\sqrt{\text{var}(X)}}$ that $P \left[ |X - E[X]| \geq \text{E}[X] \right] \leq \text{var}(X)/\text{E}[X]^2$, we have

$$P \left[ E_{12} \mid G_\mathbf{x} \right] = P \left[ |G_\mathbf{x}| = 0 \mid G_\mathbf{x} \right] \leq P \left[ \|G_\mathbf{x}\| - E[|G_\mathbf{x}|] \mid G_\mathbf{x} \right] \geq E[|G_\mathbf{x}|] \leq \frac{\text{var}(|G_\mathbf{x}|)}{(\text{E}[|G_\mathbf{x}|])^2}$$

The mean and the variance can be bounded by viewing the cardinality as a sum of indicator functions. For $x \in T_e^N(X)$,

$$\text{E}[|G_\mathbf{x}|] = \sum_{i,j} P \left[ (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X) \right] = \sum_{i,j} \sum_{x_1, x_2 \mid (x_1, x_2, x) \in T_e^N(X_1, X_2, X)} P \left[ X^N(i), X^N(j), x \in T_e^N(X_1, X_2, X) \right]$$

$$\geq (1 - \epsilon)2^{N(R_{11} + R_{22} - H(X_1) - H(X_2) + H(X_1, X_2)|X| - e_{c_f})}$$

while the variance can be bounded via

$$\text{var}(|G_\mathbf{x}|) = \sum_{i,j,i',j'} P \left[ (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X), (X^N(i'), X^N(j'), x) \in T_e^N(X_1, X_2, X) \right]$$

$$- \sum_{i,j,i',j',i \neq i'} P \left[ (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X) \right] P \left[ (X^N(i'), X^N(j'), x) \in T_e^N(X_1, X_2, X) \right]$$

$$+ \sum_{i,j,i',j \neq j'} P \left[ (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X) \right] P \left[ (X^N(i'), X^N(j'), x) \in T_e^N(X_1, X_2, X) \right]$$

$$+ \sum_{i,j,i' \neq i} P \left[ (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X) \right] P \left[ (X^N(i'), X^N(j), x) \in T_e^N(X_1, X_2, X) \right]$$

$$- \sum_{i,j,i' \neq i} P \left[ (X^N(i), X^N(j), x) \in T_e^N(X_1, X_2, X) \right] P \left[ (X^N(i'), X^N(j), x) \in T_e^N(X_1, X_2, X) \right]$$

$$\leq \sum_{i,j,i',j} 2^{N(H(X_1) - H(X_2) + H(X_1, X_2|X) + e_{c_f})} 2^{N(H(X_1) - H(X_2) + H(X_1, X_2|X) + e_{c_f})}$$

$$+ \sum_{i,j,i' \neq i} 2^{N(H(X_1) - H(X_2) + H(X_1, X_2|X) + e_{c_f})} 2^{N(H(X_1) - H(X_2) + H(X_1, X_2|X) + e_{c_f})}$$

$$\leq 2^{N(R_{11} + R_{22})} \left( (2^{NR_{11}} - 1) + (2^{NR_{22}} - 1) + 1 \right) (1 - \epsilon)^2 2^{N(H(X_1) - H(X_2) + H(X_1, X_2|X) - e_{c_f})}$$
\[
\begin{align*}
\text{for } R_1 = R_{01} + R_{11} \text{ and } R_2 = R_{02} + R_{22} \text{ the region}
\end{align*}
\]
2.2 Cases where EGC is Tight

- No Excess Rate: Regarding the decoder obtaining both descriptions, it is clear that it can not require less rate than a decoder receiving a single message requiring this distortion. When the rates are selected so that this bound is tight, that is for those rates such that \( R_1 + R_2 = R(D_0) \), the (Ahlswede)

- Gaussian Squared Error. For \( X_n \) scalar Gaussian distributed and \( d \) the squared error, the EGC region is tight.

2.3 Zhang Berger Region

The Zhang Berger region grows the El Gamal cover region by allowing a certain part of each of the two descriptions to be identical, with the remaining descriptions generated conditionally on this part. The achievability construction is as follows

- **Codebook Generation**: Generate a \( 2^{NR_c} \times N \) matrix with elements IID according to \( p_U \), and denote the \( i \)th row of this matrix by \( U_N(i) \). For each \( i \in \{1, \ldots, 2^{NR_c}\} \) generate a \( 2^{NR_1} \times N \) matrix whose rows \( X_1^N(i,j) \) are IID and distributed according to

  \[
  \prod_{n=1}^{N} p_{X_1|U}(x_{1,n}(i,j)|u_n(i))
  \]  

  (38)

  and a \( 2^{NR_2} \times N \) matrix whose rows \( X_2^N(i,k) \) are IID and distributed according to

  \[
  \prod_{n=1}^{N} p_{X_2|U}(x_{2,n}(i,k)|u_n(i))
  \]  

  (39)

  For each \((i,j,k)\) generate a \( 2^{NR_0} \times N \) matrix whose rows \( X_0^N(i,j,k,\ell) \) are IID according to

  \[
  \prod_{n=1}^{N} p_{X_0|X_1,X_2,U}(x_{0,n}(i,j,k,\ell)|x_{1,n}(i,j),x_{2,n}(i,k),u_n(i))
  \]  

  (40)

  Share the codebook matrices with the encoders and decoders.

- **Encoder**: Find \((i,j,k,\ell)\in\{1,\ldots,2^{NR_c}\} \times \{1,\ldots,2^{NR_1}\} \times \{1,\ldots,2^{NR_2}\} \times \{1,\ldots,2^{NR_0}\}\) such that

  \((U_N(i),X_1^N(i,j),X_2^N(i,k),X_0^N(i,j,k,\ell),X^N)\in T_\epsilon^N(U,X_1,X_2,X_0,X)\). The bits of the index \( \ell \) is broken up into two parts \( \ell = (\ell_1,\ell_2) \), and the two descriptions that are sent are \( S_1 = (i,j,\ell_1) \) and \( S_2 = (i,k,\ell_2) \).

- **Decoder**: The decoder receiving only \( S_1 \) discards \( \ell_1 \) and produces \( X_1^N(i,j) \). The decoder receiving only \( S_2 \) discards \( \ell_2 \) and produces \( X_2^N(i,k) \). The decoder receiving both descriptions produces \( X_0^N(i,j,k,\ell) \).

While we did not go into the details of the proof, we noted that this rate region obtains some extra points that are not in the El Gamal Cover region for the binary source Hamming distortion case.
3 Successive Refinements

In successive refinements, we neglect one of the two decoders receiving an individual description in the multiple descriptions problem, and we obtain the problem shown in Fig. 3. The El Gamal Cover inner bound for the multiple descriptions problem can be specialized to this case by selecting $X_2 = 0$. The region so obtained is

\begin{align*}
R_1 &\geq I(X_1; X) \quad (41) \\
R_1 + R_2 &\geq I((X_1, X_0); X) \quad (42)
\end{align*}

3.1 EGC Tightness: Converse Proof

Successive refinements represent a special case of the more general multiple descriptions problem for which the El Gamal Cover region is tight. We presented the proof in class. Suppose that $R_1$ and $R_2$ are the rates of the successive descriptions achieving distortions $D_1$ and $D_0$, respectively. Following the normal rate distortion converse

\begin{align*}
NR_1 &\geq H(\hat{X}_1^N) = H(\hat{X}_1^N) - H(\hat{X}_1^N | X^N) = I(\hat{X}_1^N; X^N) = H(X^N) - H(X^N | \hat{X}_1^N) \\
&= \sum_{n=1}^{N} H(X_n) - H(X_n | \hat{X}_1^{n-1}, \hat{X}_1^n) \geq \sum_{n=1}^{N} H(X_n) - H(X_n | \hat{X}_1,n) = \sum_{n=1}^{N} I(X_n; \hat{X}_1,n) \quad (43)
\end{align*}

\begin{align*}
N(R_1 + R_2) &\geq H(\hat{X}_1^N, \hat{X}_0^N) = H(\hat{X}_1^N) + H(\hat{X}_0^N | \hat{X}_1^N) = I(\hat{X}_1^N, \hat{X}_0^N; (\hat{X}_1^N, \hat{X}_0^N)) = H(X^N) - H(X^N | \hat{X}_1^N, \hat{X}_0^N) \\
&\geq \sum_{n=1}^{N} H(X_n) - H(X_n | \hat{X}_0^{n-1}, \hat{X}_1^n, \hat{X}_1^n) \geq \sum_{n=1}^{N} H(X_n) - H(X_n | \hat{X}_0,n, \hat{X}_1,n) = \sum_{n=1}^{N} I(X_n; (\hat{X}_0,n, \hat{X}_1,n)) \quad (44)
\end{align*}

Rearranging, we obtain the following inequalities

\begin{align*}
R_1 &\geq \frac{1}{N} \sum_{n=1}^{N} I(X_n; \hat{X}_1,n) \quad (45) \\
R_1 + R_2 &\geq \frac{1}{N} \sum_{n=1}^{N} I(X_n; (\hat{X}_0,n, \hat{X}_1,n)) \quad (46)
\end{align*}

\begin{align*}
D_1 &\geq \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[d(X_n, \hat{X}_1,n)] \quad (47) \\
D_0 &\geq \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[d(X_n, \hat{X}_0)] \quad (48)
\end{align*}

showing that the vector $(R_1, R_2, D_1, D_2)$ is elementwise superior to a point in the convex hull of the rate distortion region given by (40) and (41).

3.2 Successive Refinability of a Source and Distortion Measure

Regarding each of the two decoders individually in Figure 3, it is quite clear that they can not require less rate than a decoder which must only reproduce the source at their associated distortion. Hence, if $R(D)$ is the rate distortion function for the
source $X$ and the distortion metric $d(\cdot, \cdot)$, the rates in the successive refinements problem must obey the bound $R_1 \geq R(D_1)$ and $R_1 + R_2 \geq R(D_0)$. A source is *successively refineable* if these bounds are achievable. That is, for a successively refineable source, the minimum average amount of extra information per symbol necessary to refine a description of the source at distortion $D_1$ to a description of the source at $D_0$ is $R(D_0) - R(D_1)$. A source is successively refineable if and only if for every $D_0 \leq D_1$ there is a conditional distribution $p_{\hat{X}_0, \hat{X}_1|X} = p_{\hat{X}_0|X}p_{\hat{X}_1|\hat{X}_0}$ (i.e. obeying the Markov chain $X \leftrightarrow \hat{X}_0 \leftrightarrow \hat{X}_1$ with $p_{X_1|X}$ and $p_{X_0|X}$ attaining the rate distortion bounds: $I(X; \hat{X}_1) = R(D_1)$ and $I(X; \hat{X}_0) = R(D_0)$