6.962: Week 2

Presenter: Emin Martinian
Date: September 20, 2000
Topic: The Multiple Descriptions Problem

1 Overview

The problem of multiple descriptions shown in slide 2 was posed by Gersho, Witsenhausen, Wolf, Wyner, Ziv, and Ozarow at the September 1979 IEEE Information Theory Workshop. Over the past decades, researchers have studied the problem by proving various coding theorems, converse theorems, and finding practical constructions.

There is some controversy over whether the multiple descriptions problem is an appropriate model for the Internet. Computers separated by large distances communicate by sending packets which are aggregated and eventually go over a single path through the core. Therefore the model of two independent routes does not necessarily apply and multiple descriptions techniques may be inappropriate.

An alternate application for multiple descriptions is multi-carrier modulation. Different descriptions could be sent along each tone to provide frequency diversity. If either description is undecodable due to noise or fading on one frequency, then a coarse reconstruction can be created from the other description. If both tones are clean and both descriptions are available, a higher quality representation can be created. The multi-stream digital audio broadcasting system developed by Lucent, Lucent Digital Radio, and USA Digital Radio uses this type of structure.

2 EGC Coding Theorem

Slides 3-16 discuss the coding theorem proved by El Gamal and Cover. The coding theorem is based on a two stage encoding structure where the first stage creates two descriptions, i and j, while the second stage creates a refinement, k. The refinement is arbitrarily split between the two descriptions. Thus the side decoders receive either i or j and form a coarse reconstruction while the central decoder receives (i, j, k) and forms a higher quality reconstruction. Later work has shown that this coding theorem is tight in the "no excess rate" case but not tight in other cases. In order to see where the EGC theorem is tight we consider a converse theorem.

3 Converse Theorem

Sher and Feder proved a converse theorem which is summarized in slides 17 and 18. According to the rate-distortion theorem, the rates of each description must exceed $I(\hat{X}_1; X)$ and $I(\hat{X}_2; X)$. If we define $\hat{X}_0$ as an additional encoding which is used only by the the central decoder, then the excess rate required to satisfy the distortion constraint on $\hat{X}_0$ can be bounded as shown on slide 18.

If we define S_1 and S_2 as the two side descriptions, then equation (1) follows because there are 2^{nr_Δ} possible reconstructions in the range of the joint encoding function. Equation (2) follows since

$$H(\hat{X}_0 : S_1, S_2) = I(\hat{X}_0; X|S_1, S_2) + H(\hat{X}_|X, S_1, S_2)$$

Equation (3) follows from the chain rule for mutual informations. Equation (4) follows since

$$I(X_k; \hat{X}_0^n|S_1, S_2, X_1, ..., X_{k-1}) \geq I(X_k; \hat{X}_{0k}|S_1, S_2, X_1, ..., X_{k-1})$$

for each value of k. The final line follows by choosing u_k and v_k appropriately. The final line can be single letterized using convexity arguments. A final manipulation allows the theorem to be stated without the auxiliary random variables. Details are in the paper by Sher and Feder.

The "no excess rate" case is defined as the subproblem where $R_1 + R_2 = R_0(D_0)$. $R_0(D_0)$ is the rate distortion function for the central decoder. Ahlswede has shown that it suffices to consider independent descriptions of the form $p(\hat{x}_1, \hat{x}_2) = p(\hat{x}_1)p(\hat{x}_2)$ for this subproblem. In addition, since the total description $(\hat{X}_1, \hat{X}_2, \hat{X}_0)$ meets the rate-distortion bound, the joint distribution of $(\hat{X}_1, \hat{X}_2)$ can be chosen such that

$$p(\hat{x}_1, \hat{x}_2|x) = p(\hat{x}_1|x)p(\hat{x}_2|x)$$

This follows since if the encodings were not independent given X, the total rate required could be decreased by making the conditionally independent. This would contradict the statement that the original encoding meets the rate distortion bound. Therefore the encodings must be conditionally independent.

Sher and Feder show that the conditions above imply that the achievable region given by the EGC theorem reduces to the converse. Thus both the EGC theorem and the converse are tight in the no excess rate case.

4 Zhang And Berger Coding Theorem

Slides 21-30 discuss an alternate coding theorem proved by Zhang and Berger. This encoding scheme works by first encoding a common core, k. In the next stage, refinements i and j are created based on k. The two descriptions are (k, i) and (k, j). Since k is repeated in both descriptions there is some redundancy in the description received by the central decoder. Consequently, if k has non-zero rate we would expect that this encoding will not be tight in the "no excess rate" case.

In addition to providing an alternative possible encoding structure, the ZB theorem can be used to show that the EGC theorem is not tight for a binary source with Hamming distortion. This indicates that a more sophisticated coding structure might be more effective.

The first 2 equations on slide 28 of the demonstration of the ZB theorem are subtle so we provide a brief explanation. The first equation counts the number of marginally typical sequences once the common codeword, $\hat{X}_0$, has been selected. Since $\hat{X}_1$ and $\hat{X}_2$ are generated separately there are, $2^{nH(\hat{X}_1|\hat{X}_0)}$ and $2^{nH(\hat{X}_2|\hat{X}_0)}$ possible marginally typical sequences for each. Since all the random codebooks are generated independent of the source value, there are also $2^{nH(X)}$ possible marginally typical sequences for the source. Combining these yields the result in the first equation.

The second equation counts the number of jointly typical sequences once a common codeword, $\hat{X}_0$, has been chosen. The number of sequences where the source and all codewords are jointly typical is

$$2^{nH(X,\hat{X}_0,\hat{X}_1,\hat{X}_2)} = 2^{nH(\hat{X}_0)} \times 2^{nH(X,\hat{X}_1,\hat{X}_2|\hat{X}_0)}$$

The line above counts the total number of jointly typical sequences. Since we only want to count the ones which are jointly once $\hat{X}_0$ has been selected, we only need the rightmost term. This is the term that appears in the second equation on slide 28 after splitting off $H(X)$ using the entropy chain rule.

5 Comparison Of Encoders

The EGC encoder and the ZB encoder have notably different structures. Neither is strictly included in the other. One way to build a generalized encoder is to time-share between the two. A more general way is to use the encoder shown in Figure 1. The achievable region for this encoder is no

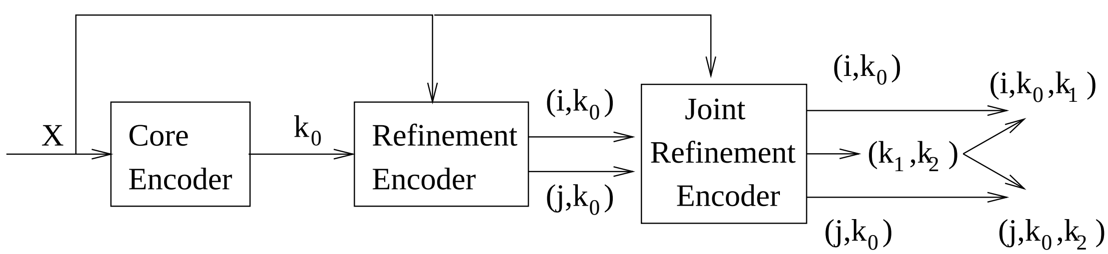

Figure 1: A diagram of a generalized encoder

smaller than the achievable region for the EGC or ZB encoder because it includes them as special cases.