ANALYSIS OF SLEPIAN WOLF CODING

Jagadish Venkataraman,
University of Notredame

February 4, 2003

Definition of the problem

Two correlated information sources X and Y are obtained from a bivariate distribution *p(x,y)* . Encoders for X and Y do not have any knowledge of the other. The decoders have full information as regards the other bit stream. We try to determine the minimum number of bits per source character required for the two encoded message streams to ensure accurate reconstruction by the decoder of the two outputs.

We know how to encode a source X. A rate $R \geq H(X)$ is sufficient for accurate reconstruction of X at the decoder. Now, suppose we have two sources $(X, Y) \sim p(x, y)$. A rate $H(X, Y)$ is sufficient if we are encoding them together. Consider the scenario where X and Y have to be encoded separately. Clearly, a rate $R(= R_X + R_Y) \geq H(X) + H(Y)$ is sufficient. Slepian and Wolf, however, went on to show that a rate $R \geq H(X, Y)$ would be sufficient to accurately reconstruct both X and Y at the decoder. The proof for the same will be discussed in detail in the following sections.

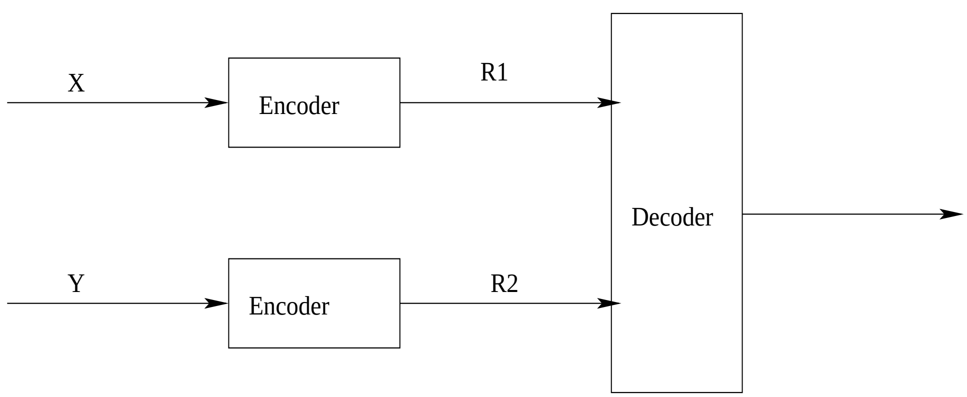

Figure 1: Distributed Source Code

Entropy, Typical Sequences and related concepts

Consider X to be a discrete random variable taking values in the set $\mathcal{X}$ = {1,2,...,M}. Denote the probability distribution of X by $p_X(x)$ = Pr[X = x], x $\in \mathcal{X}$. Now, let **X** $= \{X_1, X_2, ..., X_n\}$ be a sequence of n realizations of X so that the probability distribution for the random n - vector **X** is given by ,

$$\mathbf{P}_X(\mathbf{x}) = \mathbf{Pr}[\mathbf{X} = \mathbf{x}] = \prod_{i=1}^{n} p_X(x_i)$$

$$\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathcal{X}^n$$

We regard **X** as a block of n successive characters from the output of an information source producing characters independently with letter distribution $p_X(x)$. In a typical long block, we have letter 1 occurring $np_X(1)$ times, letter 2 occurring $np_X(2)$ times etc. The probability of such a long typical sequence is, therefore,

$$p_T = p_X(1)^{np_X(1)} ... p_X(M)^{np_X(M)}$$

$$= \exp[np_X(1)\ \log p_X(1)] .... \exp[np_X(M)\ \log p_X(M)]$$

$$= \exp[-nH(X)]$$

where

$$H(X) = -\sum_{1}^{M} p_X(i)\ \log p_X(i)$$

is called the entropy of the random variable X.

We define these $\exp[nH(X)]$ to be the typical sequences and these constitute the set of sequences, $A_\epsilon^{(n)}$ that is most likely to occur. Each of these typical sequences is equally likely and occur with probability $\exp[-nH(X)]$. The conclusion drawn from this is that we can transmit the information source over the channel with a rate $R = H(X)$ and that at least this rate is required for accurate reconstruction at the decoder.

Extending the above result to a pair of sources, a rate $R > H(X, Y)$ is sufficient to encode a pair of sources (X,Y) together. Before getting to the Slepian Wolf theorem for encoding of correlated sources, a few useful definitions are as follows.

1. A $((2^{nR_1}, 2^{nR_2}), n)$ distributed source code for the joint source (X,Y) consists of two encoder maps,

$$f_1 : \mathcal{X}^n \rightarrow \{1, 2, ..., 2^{nR_1}\}$$

$$f_2 : \Upsilon^n \rightarrow \{1, 2, ..., 2^{nR_2}\}$$

and a decoder map,

$$g : \{1, 2, ..., 2^{nR_1}\} \; X \; \{1, 2, ..., 2^{nR_2}\} \rightarrow \mathcal{X}^n \; X \; \Upsilon^n$$

Here $f_1(X^n)$ is the index corresponding to X^n, $f_2(Y^n)$ is the index corresponding to Y^n and (R_1, R_2) is the rate pair of the code.

2. The probablility of error for a distributed source code is defined as,

$$P_e^{(n)} = P(g(f_1(X^n), f_2(Y^n)) \neq (X^n, Y^n)).$$

3. A rate pair is said to be *achievable* for a distributed source is there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ distributed source codes with probability of error $P_e^{(n)} \rightarrow 0$. The *achievable rate region* is the set of achievable rates.

Having defined these, the Slepian Wolf theorem can now be analyzed in detail.

Theorem: For the distributed source coding problem for the source (X,Y) drawn i.i.d $\sim p(x, y)$, an achievable rate point is given by:

$$R_1 = H(X/Y) + \epsilon_x, \epsilon_x > 0$$

$$R_2 = H(Y) + \epsilon_y, \epsilon_y > 0$$

The proof for achievability of the rates in the Slepian Wolf theorem is presented by introducing a new coding theorem using random bins. The idea behind these random bins is very similar to hash functions: we choose a large random index for each source sequence. If the number of these typical sequences is small enough, then with high probability, different source sequences will have different indices, and we can recover the source sequence from the index.

The procedure for the random binning is as follows: For each sequence X^n, draw an index at random from {1,2,...,2^{nR}}. The set of sequences that have the same index are said to form a bin. For decoding the source from the bin index, we look for a typical X^n sequence in the bin. If there is one and only one typical sequence in the bin, we declare it to be the estimate of the source sequence; otherwise, an error is declared. Thus, an error occurs only if there is more than one typical sequence in this bin. If the source sequence is non–typical, then there will always be an error. The probability of error is arbitrarily small for sufficient R.

Consider the encoding and decoding problem for a single source. The proof for the above coding scheme producing an arbitrarily small probability of error for $R > H(X)$ is as follows,

$$
\begin{aligned}
P_e^{(n)} &= P[g(\mathbf{X}) \neq X] \\
&= P[(\mathbf{X} \notin A_\epsilon^{(n)}) \cup (f(\mathbf{X'}) = f(\mathbf{X}); (\mathbf{X'}, \mathbf{X}) \in A_\epsilon^{(n)}, \mathbf{X'} \neq \mathbf{X})] \\
&\leq P[\mathbf{X} \notin A_\epsilon^{(n)}] + \sum_x P[\exists \mathbf{x'} \neq \mathbf{x} : \mathbf{x'} \in \mathbf{A}_\epsilon^{(\mathbf{n})}, f(\mathbf{x'}) = f(\mathbf{x})] p(\mathbf{x}) \\
&\leq \epsilon + \sum_x \sum_{\mathbf{x'} \in A_\epsilon^{(n)}, \mathbf{x'} \neq \mathbf{x}} P(f(\mathbf{x'}) = f(\mathbf{x})) p(\mathbf{x}) \\
&= \epsilon + \sum_{\mathbf{x'} \in A_\epsilon^{(n)}} 2^{-nR} \sum_x p(\mathbf{x}) \\
&\leq \epsilon + 2^{-nR} 2^{n(H(X)+\epsilon)} \\
&\leq 2\epsilon
\end{aligned}
$$

if $R > H(X) + \epsilon$ and n is sufficiently large. Hence, if the rate of the code is greater than the entropy, the probability of error is arbitrarily small and the information sequence is efficiently decoded at the receiving end. The same scheme can be extended to multiple sources as will be illustrated below. Moreover, it is noteworthy that this binning scheme does not require an explicit characterization of the typical set at the encoder; it is only needed at the decoder. This property is exploited in using it for the distributed source case.

A formal proof for the achievability of the rate region specified by the Slepian Wolf theorem will now be presented. The above proof will be used in determining the error probability for the two source case.

Proof of achievability of the rate region specified by the Slepian Wolf theorem:

The basic idea of the proof is to partition the space of $\mathcal{X}^n$ into 2^{nR_1} bins and the space of Υ^n into 2^{nR_2} bins.

Random code generation: Independently assign every $\mathbf{x} \in \mathcal{X}^n$ to one of 2^{nR_1} bins according to a uniform distribution on $\{1, 2, ..., 2^{nR_1}\}$. Similarly, randomly assign every $\mathbf{y} \in \Upsilon^n$ to one of 2^{nR_2} bins. Reveal the assignments f_1 and f_2 to both the encoders and the decoder.

Encoding: Source 1 sends the index of the bin to which $\mathbf{X}$ belongs and source 2 sends the index of the bin to which $\mathbf{Y}$ belongs.

Decoding: Given the index pair (i_0, j_0), declare $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = (\mathbf{x}, \mathbf{y})$, if there is one and only one pair of sequences $(\mathbf{x}, \mathbf{y})$ such that $f_1(\mathbf{x}) = i_0, f_2(\mathbf{y}) = j_0$ and $(\mathbf{x}, \mathbf{y}) \in A_\epsilon^{(n)}$. Otherwise, declare an error.

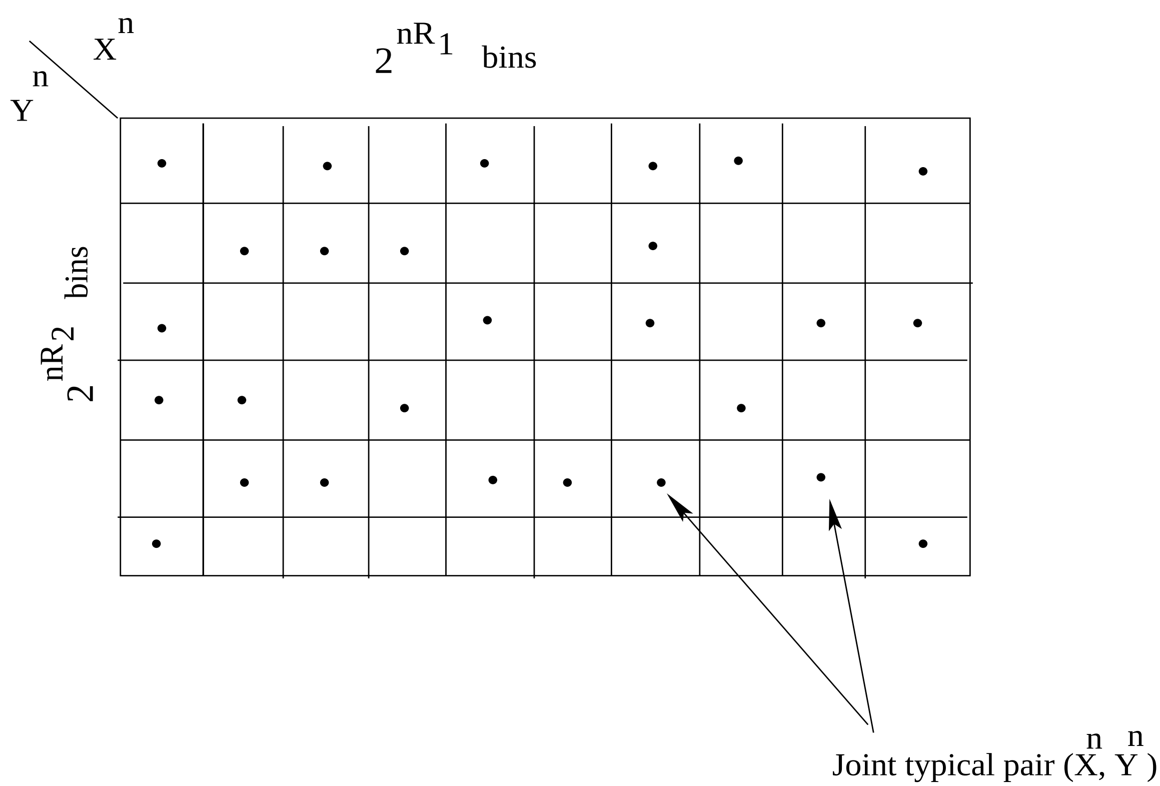

Figure 2: Jointly Typical Sequences

The set of X sequences and the set of Y sequences are divided into bins such a way that the pair of indices specifies a product bin. Having done this, the probability of error at the decoder is defined as the union of the following events:

$$E_0 = \{(\mathbf{X}, \mathbf{Y}) \notin A_\epsilon^{(n)}\}$$

$$E_1 = \{\exists \mathbf{x}' \neq \mathbf{X} : f_1(\mathbf{x}') = f_1(\mathbf{X}) \text{ and } (\mathbf{x}', \mathbf{Y}) \in A_\epsilon^{(n)}\}$$

$$E_2 = \{\exists \mathbf{y}' \neq \mathbf{Y} : f_2(\mathbf{y}') = f_2(\mathbf{Y}) \text{ and } (\mathbf{X}, \mathbf{y}') \in A_\epsilon^{(n)}\}$$

$$E_3 = \{\exists (\mathbf{x}', \mathbf{y}') : \mathbf{x}' \neq \mathbf{X}, \mathbf{y}' \neq \mathbf{Y}, f_1(\mathbf{x}') = f_1(\mathbf{X}), f_2(\mathbf{y}') = f_2(\mathbf{Y}) \\ \text{and } (\mathbf{x}', \mathbf{y}') \in A_\epsilon^{(n)}\}$$

Thus,

$$P_e^{(n)} = P(E_0 \cup E_1 \cup E_2 \cup E_3)$$

$$\leq P(E_0) + P(E_1) + P(E_2) + P(E_3)$$

Extending the result for a single source to two sources, we can say that the cardinality of the set of jointly atypical sequences $(\mathbf{x}, \mathbf{y})$ is very small compared to that of the jointly typical sequences. It follows that the probability measure of that set $\rightarrow 0$ for large n. Hence,

$$P(E_0) = \epsilon$$

Now lets consider $P(E_1)$,

$$P[E_1/(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})] = \bigcup_{(x', y) \in A_\epsilon^{(n)}, (\mathbf{x}' \neq \mathbf{x})} \{f_1(\mathbf{x}') = f_1(\mathbf{x})\}$$

Thus,

$$P[E_1] = \sum_{(\mathbf{x}', \mathbf{y}')} p(\mathbf{x}, \mathbf{y}) P[E_1/(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})]$$

$$\leq \sum_{(\mathbf{x}, \mathbf{y})} p(\mathbf{x}, \mathbf{y}). \sum_{(\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}, (\mathbf{x}' \neq \mathbf{x})} P[f_1(\mathbf{x}') = f_1(\mathbf{x})] \quad \{\text{Union Bound}\}$$

$$\leq \sum_{(\mathbf{x}, \mathbf{y})} p(\mathbf{x}, \mathbf{y}) 2^{-nR_1} |A_\epsilon(\mathbf{X}/\mathbf{y})|$$

$$\leq 2^{-nR_1} 2^{n(H(X/Y)+\epsilon)}$$

which $\rightarrow 0$ if $R_1 > H(X/Y)$ and n is large.

The above result follows from the following:

$$P[f_1(\mathbf{x}') = f_1(\mathbf{x})/(f_1(\mathbf{x}) = i_0)] = 2^{-nR_1}$$

This implies that,

$$P[f_1(\mathbf{x}') = f_1(\mathbf{x})] = \sum_{i_0} P[f_1(\mathbf{x}) = i_0].P[f_1(\mathbf{x}') = f_1(\mathbf{x})/(f_1(\mathbf{x}) = i_0)]$$

$$= 2^{-nR_1}.\sum_{i_0} P[f_1(\mathbf{x}) = i_0]$$

$$= 2^{-nR_1}$$

$|A_\epsilon(\mathbf{X}/\mathbf{y})|$ is defined to be the set of $\mathbf{X}$ sequences that are jointly ϵ typical with a particular $\mathbf{Y}$ sequence. The proof for the fact that $|A_\epsilon(\mathbf{X}/\mathbf{y})| \leq 2^{n(H(X/Y)+2\epsilon)}$ is as follows,

$$1 \geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}(X/\mathbf{Y})} p(\mathbf{X}/\mathbf{Y})$$

$$\geq \sum_{\mathbf{x}\in A_\epsilon^{(n)}(X/\mathbf{Y})} 2^{-n(H(X/Y)+2\epsilon)}$$

$$= |A_\epsilon(\mathbf{X}/\mathbf{y})|\, 2^{-n(H(X/Y)+2\epsilon)}$$

Thus, we have

$$|A_\epsilon(\mathbf{X}/\mathbf{y})| \leq 2^{n(H(X/Y)+2\epsilon)}$$

Similarly, we can show that the probabilities of the events E_2 and E_3 get arbitrarily small for large values of n and sufficiently high rates. It follows from the above discussion that the overall probability of error for the joint sequence at the decoder is ,

$$P_e^{(n)} \leq 4\epsilon$$

which is arbitrarily small. It can, therefore, be seen that the condition for achievability of the rate pair has been satisfied by $(R_1, R_2) = (H(X/Y) + \epsilon_x, H(Y) + \epsilon_y)$. Hence, the proof for the theorem is complete.

The rate pair that has been suggested above can change roles i.e, we can have $(R_1, R_2) = (H(X)+\epsilon_x, H(Y/X)+\epsilon_y)$ and the theorem would still hold. This is equivalent to saying that the decoder now has complete information about the source X and is trying to decode Y based on the joint typical sequence set. Thus, the rate region can be expressed as,

$$R_1 \geq H(X/Y)$$

$$R_2 \geq H(Y/X)$$

$$R_1 + R_2 \geq H(X,Y)$$

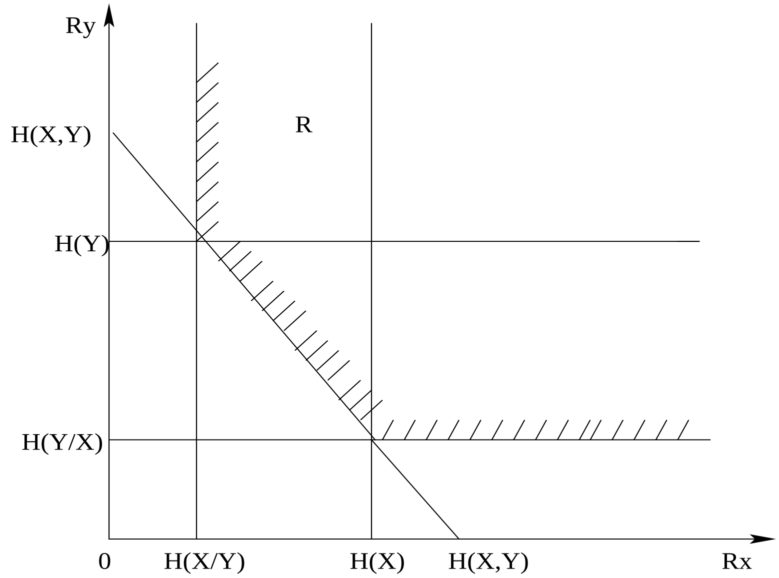

Figure 3: Achievability Region

Interpretation of the Slepian Wolf coding

Consider the rate pair $R_1 = H(X), R_2 = H(Y/X)$. We know from the definition of typical sequences that $nH(X)$ bits are sufficient to encode X for accurate reconstruction at the decoder. But how do we code Y with just $nH(Y/X)$ bits and recover the information at the decoder?

Associated with every typical sequence in $\mathcal{X}^n$, we have a bunch of sequences in Υ^n that form a jointly typical sequence. The encoders of both **X** and **Y** know the joint distribution, $p(x, y)$, and hence know the set of jointly typical sequences. So having sent **X**, all that is required to be done is to send the index of **Y** within the small collection corresponding to this **X**.

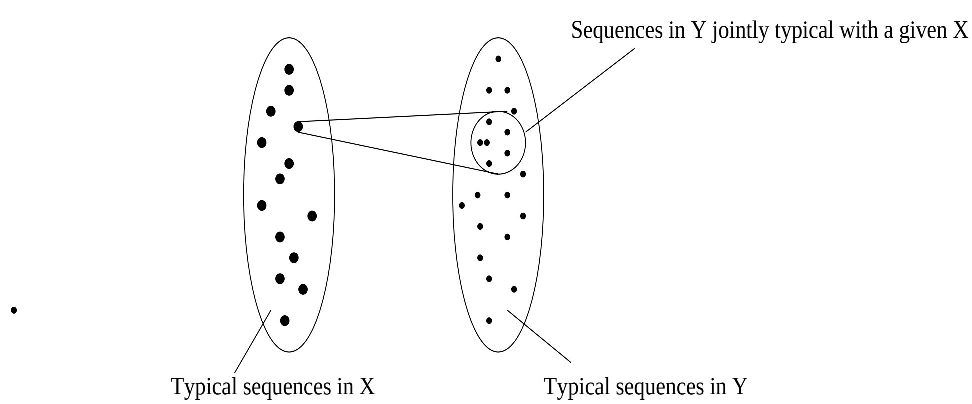

Figure 4: Sequence Association

Since the **Y** encoder does not know which sequence in **X** was sent, an arbitrary index from $\{1, 2, ..., 2^{nR_2}\}$ is assigned to every **Y** and if this index is large enough, every **Y** sequence can be uniquely identified within the collection corresponding to **X**.

$R_2 > H(Y/X) \Rightarrow$ the number of indices is exponentially greater than the number of elements in every collection. This further implies that the probability of error will be exponentially small at the decoding end.

Further Reading

Apart from a rigorous proof for determining the admissible rate region for correlated information sources (as outlined above), Slepian and Wolf[1] also go on to discuss the rate regions for 15 other cases, namely, the encoders have full information of the two sources but the decoders do not, one of the encoders has full information about both sources as does one of the decoders and so on. They also talk about time sharing and Bit stuffing when discussing these cases. The duality between Multiple Access Channels and Slepian Wolf encoding is discussed in Thomas and Cover[2], which makes interesting reading. Furthermore, the paper by Qian Zhao and Michelle Effros[3] addresses the properties of optimal instantaneous multiple access source codes with finite coding dimension ($n < \infty$) and both lossless and near lossless performance.

References

[1] D. Slepian and J.K. Wolf.'Noiseless Coding of Correlated Information Sources'.*IEEE Transcations on Information Theory*,IT-19:471-480,July 1973.

[2] T.M. Cover and J.A. Thomas.'Elements of Information Theory'.*John Wiley and Sons INC.*

[3] Q. Zhao and M. Effros.'Lossless and Near Lossless Coding for Multiple Access Networks'.*Data Compression Conference*,2001: 263-272