Lossy Source Coding under a Maximum Distortion Constraint with Decoder Side-Information

Jayanth Nayak, Ertem Tuncel, Kenneth Rose

Abstract

We consider the problem of variable length coding of sources where *every* sample needs to be reconstructed within a certain distortion using side-information available solely at the decoder. We derive a non-single-letter characterization of the minimum asymptotic rate required and bound this quantity. We also derive necessary and sufficient conditions for an upper and lower bound on this rate to coincide. For the exact reconstruction case, these conditions specialize to those for the normality of the source confusablity graph.

I. Introduction

A basic problem in information theory is source coding under a distortion constraint when the decoder has side-information about the source [1]. Traditionally, the constraint imposed is that the expected distortion between the source and its reconstruction averaged over the block being coded not exceed a given value. In certain applications (e.g. medical imaging) however, this constraint is considered too weak: we require that with probability 1 the maximum samplewise distortion in a block not exceed a given value. In other words, we require that the distortion at *each* sample be less than a given level with probability 1. Special cases of this problem have been studied earlier: the case where perfect reconstruction is required was considered in [2], [3], [4] and elsewhere and the case where there is no side-information was considered in [5].

In this paper we shall focus on the problem of variable length coding of a memoryless source under a maximum distortion constraint when there is side-information solely at the decoder. The problem is formally defined in Section II. Using a combinatorial approach, we derive a non-single-letter expression for the minimum asymptotic average rate as well as single-letter bounds on this rate. These quantities reduce to known ones when the reconstruction is required to be exact. In Section III, we derive necessary and sufficient conditions for a certain pair of upper and lower bounds on the rate to coincide.

II. Variable Length Lossy Source Coding with Decoder Side-Information

The key elements of the problem setup are: a sequence of pairs of random variables $\{(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}\}$, drawn independently at each instant from some common distribution $P_{XY}(X, Y)$, a distortion measure $d\colon \mathcal{X} \times \mathcal{Z} \to [0, \infty)$, and a distortion level $D \geq 0$. $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ are the source, side-information and reconstruction alphabets, respectively. These sets are assumed to be finite.

Scalar Code: A scalar maximum distortion side-information (MDSI) code (ϕ, ψ), is an encoder-decoder pair such that $d(X, \psi(\phi(X), Y)) \leq D$ with probability 1. The range of ϕ is the set of binary strings $\{0, 1\}^*$. We further constrain the codes to be *side-information (SI) prefix-free* codes: for all $x_1, x_2 \in \mathcal{X}$, if there is a $y \in \mathcal{Y}$ such that $P_{XY}(x_1, y)P_{XY}(x_2, y) > 0$, $\phi(x_1)$ cannot be a proper prefix of $\phi(x_2)$. The rate of this code is

$$\bar{l}(\phi) = \sum_{x \in \mathcal{X}} P_X(x)|\,\phi(x)|,$$

where P_X is the $\mathcal{X}$-marginal of P_{XY} and $|\cdot|$ denotes the string length. $\bar{L}(P_{XY}, D)$ is the minimum average number of bits required to encode an instance of X:

$$\bar{L}(P_{XY}, D) = \min\{\bar{l}(\phi) : \phi \text{ along with some } \psi \text{ is a valid MDSI code for } (X, Y) \text{ at distortion level } D.\}$$

J. Nayak and K. Rose are with the University of California, Santa Barbara. E-mail: {jayanth, rose}@ece.ucsb.edu. E. Tuncel is with the University of California, Riverside. Email: ertem@ee.ucr.edu

Block Code: Traditionally, the distortion constraint is extended to blocks by averaging: $\frac{1}{n}\sum_{i=1}^{n} d(x_i, z_i) \leq D$. We apply the stronger maximum distortion constraint: $\max_{1\leq i\leq n} d(x_i, z_i) \leq D$. We are interested in the minimum asymptotic average rate:

$$R^*(P_{XY}, D) = \lim_{n\to\infty} \frac{1}{n}\bar{L}_n(P_{XY}, D),$$

where $\bar{L}_n(P_{XY}, D)$ is the minimum average rate required to encode n instances of X subject to maximum distortion D. The limit exists from the subadditivity of $\bar{L}_n(P_{XY}, D)$.

Our analysis of the problem relies on the following sets

$$\begin{aligned} c(z) &= \{x \in \mathcal{X} : d(x, z) \leq D\} \\ a(y) &= \{x \in \mathcal{X} : p(x, y) > 0\} \end{aligned}$$

c are the D-balls while a are the fan-out sets. Let $\mathcal{A}$ denote the family of fan-out sets and $\mathcal{C}$, the family of D-balls. The fan-out sets of blocks are the Cartesian products of the fan-out sets at each co-ordinate. Since we impose a maximum distortion constraint, the same property holds for D-balls as well. We can restate the distortion constraint on the code in terms of the fan-out sets and the D-balls. We present some definitions that will be required.

Given a set family $\mathcal{T}$ over the alphabet $\mathcal{X}$, we define its completion $\tilde{\mathcal{T}}$ as follows:

$$\tilde{\mathcal{T}} = \{\tilde{t} \subset \mathcal{X} : \tilde{t} \subset t \in \mathcal{T}\}$$

Also, the complement of one set family $\mathcal{T}$ with respect to another set family $\mathcal{U}$ defined over the same alphabet $\mathcal{X}$ is the family

$$\bar{\mathcal{T}}(\mathcal{U}) = \{b \subset \mathcal{X} : \forall t \in \mathcal{T}, t \cap b \in \tilde{\mathcal{U}}\}.$$

Note that in general, $\bar{\bar{\mathcal{T}}}(\mathcal{U}) \neq \mathcal{T}$, but if $\mathcal{V} \triangleq \bar{\mathcal{T}}(\mathcal{U})$, $\mathcal{V} = \bar{\bar{\mathcal{V}}}(\mathcal{U}) = \tilde{\mathcal{V}}$.

The distortion constraint on a MDSI code can now be stated as:

$$\phi^{-1}(h) \in \overline{\mathcal{A}^n}(\mathcal{C}^n), \forall h \in \{0, 1\}^*,$$

where $\mathcal{T}^n$ denotes the n-fold Cartesian product of the set family $\mathcal{T}$. Note that for strings h that are not valid codewords $\phi^{-1}(h)$ is the empty set, which is certainly in $\overline{\mathcal{A}^n}(\mathcal{C}^n)$.

To derive an expression for $R^*(P_{XY}, D)$ we need a few more definitions. Given a set $\mathcal{X}$ and a non-degenerate ($\cup_{t\in\mathcal{T}} t = \mathcal{X}$) set family $\mathcal{T}$ on $\mathcal{X}$, a $\mathcal{T}$-partition is a partition of $\mathcal{X}$ into sets from $\tilde{\mathcal{T}}$. The $\mathcal{T}$-partition entropy of a distribution P on the set $\mathcal{X}$ is defined as:

$$H_{\chi}(P, \mathcal{T}) = \min\{H(c(X)) : c\colon \mathcal{X} \to \mathcal{T} \text{ is a } \mathcal{T}\text{-partition }\},$$

where $H(.)$ is the Shannon entropy [1].

Let $\mathcal{B}_n = \overline{\mathcal{A}^n}(\mathcal{C}^n)$ and $\mathcal{B} = \mathcal{B}_1$. The following lemma bounds $\bar{L}_n(P_{XY}, D)$.

Lemma 1: $H_{\chi}(P_X^n, \mathcal{B}_n) - \log\{H_{\chi}(P_X^n, \mathcal{B}_n) + 1\} - \log e \leq \bar{L}_n(P_{XY}, D) \leq H_{\chi}(P_X^n, \mathcal{B}_n) + 1$ (1)

Proof: Any MDSI encoding of $\mathcal{X}^n$ can be viewed as a partition of $\mathcal{X}^n$ into sets from $\mathcal{B}_n$ followed by a one-to-one encoding of the sets in the partition. The lower bound in the lemma follows from the bounds on the rates of one-to-one codes established in [6]. On the other hand a partition of $\mathcal{X}^n$ followed by a prefix-free (in the conventional sense) code is a valid MDSI code. The upper bound then follows from results on the rate of prefix-free codes [7]. ■

Lemma 1 leads to the following characterization of $R^*(P_{XY}, D)$.

Lemma 2:

$$R^*(P_{XY}, D) = \lim_{n\to\infty} \frac{H_{\chi}(P_X^n, \overline{\mathcal{A}^n}(\mathcal{C}^n))}{n} \tag{2}$$

[1] All information quantities are in bits and all logarithms are to base 2.

The proof of this lemma is similar to that of Lemma 6 in [3] and uses the fact that $H_\chi(P_X^n, \overline{\mathcal{A}^n}(\mathcal{C}^n)) \leq nH_\chi(P_X, \bar{\mathcal{A}}(\mathcal{C}))$. The limit above exists from the subadditivity of $H_\chi(P_X^n, \overline{\mathcal{A}^n}(\mathcal{C}^n))$. The argument in the two lemmas does not change if we only require *SI unique decodability* of the code, that is any concatenation of codewords corresponding to a block of the source must have a unique parsing into codewords given the side-information. Therefore, we see that codes satisfying the weaker requirement of SI unique decodability do not offer any asymptotic rate advantage over SI prefix-free codes.

In the noiseless coding with side-information scenario [3][4] where we require the decoder output to be a perfect copy of X, $\mathcal{C}$ is the family of singletons from $\mathcal{X}$. For such a scenario, the confusability graph G_X is defined as a graph on $\mathcal{X}$ where two nodes are connected if they both belong to the same fan-out set. $\mathcal{B}$ is the family of independent sets of this graph. In this case the minimum asymptotic achievable rate depends only on P_X and G_X: the $\mathcal{B}$-partition entropy is the chromatic entropy of the confusability graph and the limit in Lemma 2 is the complementary graph entropy $\bar{H}(P_X, G_X)$ [4]. Since in our case too, $R^*(P_{XY}, D)$ depends only on the marginal P_X and the set families $\mathcal{A}$ and $\mathcal{C}$, we shall denote it by $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$.

Single-letter expressions for the complementary graph entropy have eluded researchers for many years and it is unlikely that such characterizations can be easily found in the more general case that we present. We content ourselves with providing bounds on $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$.

A. *Upper bound on* $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$

Given any nondegenerate set family $\mathcal{T}$ and probability distribution P on $\mathcal{X}$, consider a rate-distortion problem with distortion measure $d\colon \mathcal{X} \times \mathcal{T} \to [0, \infty)$ such that

$$d(x,t) = \begin{cases} 0 & \text{if } x \in t \\ 1 & \text{otherwise.} \end{cases}.$$

The rate distortion function $R_d(P, D)$ [7] for any distortion level $D \geq 0$ is the asymptotic rate both necessary and sufficient for a uniquely decodable code that represents every sequence within *average* distortion D with probability 1 to exist. If $D = 0$ we obtain the average rate required to represent vectors in $\mathcal{X}^n$ by sets in $\mathcal{T}^n$ for large n. We denote this rate by $H_\kappa(P, \mathcal{T})$. It can easily be shown that [8]

$$H_\kappa(P, \mathcal{T}) = \min_{\substack{X \sim P \\ X \in T \in \tilde{\mathcal{T}}}} I(X; T). \tag{3}$$

where $I(\cdot;\cdot)$ is the mutual information. The notation $X \in T \in \tilde{\mathcal{T}}$ signifies that the random variable T ranges over the family of sets $\tilde{\mathcal{T}}$ such that if $P_{XT}(x,t) > 0$ then $x \in t$. If $\mathcal{T}$ is the set of independent sets of a graph, $H_\kappa(P, \mathcal{T})$ is the graph entropy or Körner entropy of that graph.

Returning to the MDSI problem, since $\mathcal{B}^n \subset \mathcal{B}_n$, one way to design an MDSI code would be to assign each block in $\mathcal{X}^n$ to an element of $\mathcal{B}^n$ that contains it and use a variable length code to represent the assigned sets. We therefore have the following single-letter bound on $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$.

$$\bar{H}(P_X, \mathcal{A}, \mathcal{C}) \leq H_\kappa(P_X, \bar{\mathcal{A}}(\mathcal{C})) \tag{4}$$

B. *Lower bounds on* $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$

We first derive a simple lower bound on $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$. Consider a two step maximum distortion source coding problem for vectors $X^n \in \mathcal{X}^n \sim P_X^n$ where the first layer acceptable distortion balls are the sets in $\mathcal{A}^n$ and the acceptable second layer distortion balls are the sets in $\mathcal{C}^n$. The prefix-free variable length code in the first layer should allow reconstruction from $\mathcal{A}^n$ while the second layer (SI) prefix-free variable length code along with an arbitrary valid first layer reconstruction should lead to a valid reconstruction from $\mathcal{C}^n$. This is a generalization of the problem considered by Körner and Longo [9] who required perfect reconstruction at the second layer. Since the second layer code has to work with an arbitrary first layer reconstruction, the first layer reconstruction can be considered as side-information at the decoder unavailable at the second layer encoder. So the minimum rate is $\bar{H}(P_X, \mathcal{A}, \mathcal{C})$. As in the derivation of the upper bound, a first layer rate of

$H_\kappa(P,\mathcal{A})$ is sufficient. Since the total rate cannot be less than the minimum rate required to encode with sets from $\mathcal{C}$ without the two layer constraint:

$$H_\kappa(P,\mathcal{A})+\bar{H}(P_X,\mathcal{A},\mathcal{C})\geq H_\kappa(P,\mathcal{C}) \text{ or}$$
$$\bar{H}(P_X,\mathcal{A},\mathcal{C})\geq H_\kappa(P,\mathcal{C})-H_\kappa(P,\mathcal{A}) \tag{5}$$

We now derive an improved lower bound using the following result of Wyner and Ziv [1]:

Theorem 1: Let (X,Y) be drawn independently at each instant $\sim P_{XY}(x,y)$ and let (ϕ,ψ) be an encoder-decoder pair satisfying $E\left[\frac{1}{n}\sum_{i=1}^n d(X_i,Z_i)\right]\leq D$, where $Z^n=\psi(\phi(X^n),Y^n)$, $d(\cdot,\cdot)$ is a distortion measure and $D>0$. Then

$$\tfrac{1}{n}H(\phi(X^n))\geq R_d^{wz}(P_{XY},D)\triangleq \min_W I(X;W|Y), \tag{6}$$

where the minimization is over all auxiliary random variables W such that $W\leftrightarrow X\leftrightarrow Y$ is a Markov chain and there is a function f that satisfies $E[d(X,f(W,Y))]\leq D$.

Theorem 1 implies that for SI uniquely decodable variable length codes (ϕ,ψ) that satisfy $\frac{1}{n}\sum_{i=1}^n d(X_i,Z_i)\leq D$ with probability 1, we have $\bar{l}(\phi)\geq R_d^{wz}(P_{XY},D)$. This follows from the SI unique decodability and the inequality $H(\phi(X^n))\leq n\bar{l}(\phi)+\log[ne\log|\mathcal{X}|]$ (see [7] Theorem 1.4.1).

To bound $\bar{H}(P_X,\mathcal{A},\mathcal{C})$ using the above, we transform the maximum distortion constraint to an equivalent average distortion constraint. Define a new distortion measure $d'\colon\mathcal{X}\times\mathcal{Z}\to[0,\infty)$ as

$$d'(x,z)=\begin{cases}0 & \text{if } x\in c(z)\\ 1 & \text{otherwise.}\end{cases}$$

So the constraint $\max_i d(x_i,z_i)\leq D$ is equivalent to the constraint $\frac{1}{n}\sum_{i=1}^n d'(x_i,z_i)=0$.

Let (ϕ,ψ) be a SI uniquely decodable variable length code that encodes $\mathcal{X}$-blocks of length n such that the average d' distortion between the encoder observation and the decoder reconstruction is 0. Any such code satisfies $\bar{l}(\phi)\geq R_{d'}^{wz}(P_{XY},0)$, where $\bar{l}(\phi)$ is the per letter average codeword length. This follows from the continuity of $R_{d'}^{wz}(P_{XY},D')$ at $D'=0$ [1].

Therefore, we have $\bar{H}(P_X,\mathcal{A},\mathcal{C})\geq R_{d'}^{wz}(P_{XY},0)$. We can recast this bound in a more intuitive fashion as follows:

Lemma 3:
$$R(P_{XY},d',0)=\min_{\substack{X\in S\in\mathcal{B}\\ S\leftrightarrow X\leftrightarrow Y}} I(X;S|Y) \tag{7}$$

Proof: LHS≤RHS: Obvious. Choose $f(S,Y)=S\cap a(Y)$.

LHS≥RHS: We show that every valid W, say over alphabet $\mathcal{W}$, corresponds to some S with $I(X;W|Y)\geq I(X;S|Y)$. Let f be the function that satisfies $E[d''(X,f(W,Y))]=0$. For all $w\in\mathcal{W}$, let $b(w)$ be the set of those $x\in\mathcal{X}$ such that $p(w|x)>0$. Now the distortion constraint implies that $b(w)\cap a(y)\in\tilde{\mathcal{C}}$ for all $w\in\mathcal{W},y\in\mathcal{Y}$. So every w corresponds to an element $b\in\mathcal{B}$. Let S be the random variable on $\mathcal{B}$ formed by identifying those w that correspond to the same b. We obtain the Markov chain $S\leftrightarrow W\leftrightarrow X\leftrightarrow Y$, which implies $I(X;W|Y)\geq I(X;S|Y)$ and the lemma follows. ■

The bound can be further improved due to the following lemma (Proved in appendix I).

Lemma 4: For all nondegenerate set families $\mathcal{A},\mathcal{C}$ on $\mathcal{X}$ and all $n\in\mathbb{N}$,

$$\overline{\left[\bar{\bar{\mathcal{A}}}(\mathcal{C})\right]^n}(\mathcal{C}^n)=\overline{\mathcal{A}^n}(\mathcal{C}^n) \tag{8}$$

Since the sets available to cover the sequences are the same, for random variables (X',Y') such that $X'\sim P_X$ and Y' ranging over $\bar{\bar{\mathcal{A}}}(\mathcal{C})$ with $P_{X'Y'}(x,a)>0\Rightarrow x\in a$, an asymptotic average rate of $\bar{H}(P_X,\mathcal{A},\mathcal{C})$ is both necessary and sufficient. Therefore the best single-letter bound of this form is

$$\bar{H}(P_X,\mathcal{A},\mathcal{C})\geq \max_{X\in Y\in\bar{\bar{\mathcal{A}}}(\mathcal{C})}\ \min_{\substack{X\in S\in\mathcal{B}\\ S\leftrightarrow X\leftrightarrow Y}} I(X;S|Y)\triangleq H_c(P_X,\mathcal{A},\mathcal{C}) \tag{9}$$

The improvement of bound (9) over (5) follows from the string of inequalities given below:

$$\begin{aligned}\max_{X\in Y\in\bar{\mathcal{A}}(\mathcal{C})}\min_{\substack{X\in S\in\mathcal{B}\\ Y\leftrightarrow X\leftrightarrow S}} I(X;S|Y) &= \max_{X\in Y\in\bar{\mathcal{A}}(\mathcal{C})}\min_{\substack{X\in S\in\mathcal{B}\\ Y\leftrightarrow X\leftrightarrow S}}\{I(X;S,Y)-I(X;Y)\}\\ &\overset{(a)}{\geq} \max_{X\in Y\in\bar{\mathcal{A}}(\mathcal{C})}\min_{\substack{X\in S\in\mathcal{B}\\ Y\leftrightarrow X\leftrightarrow S}}\{I(X;S\cap Y)-I(X;Y)\}\\ &\overset{(b)}{\geq} \min_{X\in Z\in\tilde{\mathcal{C}}} I(X;Z)-\min_{X\in Y\in\bar{\mathcal{A}}(\mathcal{C})} I(X;Y)\\ &\overset{(c)}{\geq} \min_{X\in Z\in\tilde{\mathcal{C}}} I(X;Z)-\min_{X\in Y\in\tilde{\mathcal{A}}} I(X;Y),\end{aligned}$$

where $X\sim P_X$. (a) follows from $X-(S,Y)-(S\cap Y)$, (b) follows from the fact that $S\cap Y$ is a random variable on $\tilde{\mathcal{C}}$ satisfying $X\in S\cap Y$ and (c) follows from the monotonicity of $H_\kappa(P,\mathcal{T})$ ($\tilde{\mathcal{U}}\subset\tilde{\mathcal{T}}\Rightarrow H_\kappa(P,\mathcal{U})\geq H_\kappa(P,\mathcal{T})$).

The quantities $H_c(P,\mathcal{A},\mathcal{C})$ and $H_\kappa(P,\mathcal{C})-H_\kappa(P,\mathcal{A})$ are different in general, and the latter quantity can even be negative: let $\mathcal{A}$ be the family of singletons in $2^{\mathcal{X}}$ and $\mathcal{C}$ contain the set $\mathcal{X}$ itself. $H_c(P,\mathcal{A},\mathcal{C})$ can easily be seen to be 0, but $H_\kappa(P,\mathcal{C})-H_\kappa(P,\mathcal{A})$ is equal to $-H(P)$.

III. Coincidence of Bounds

Given a set $\mathcal{X}$, set families $\mathcal{A},\mathcal{C}$ and $\mathcal{B}=\bar{\mathcal{A}}(\mathcal{C})$, and a distribution P on $\mathcal{X}$, consider the inequality

$$H_\kappa(P,\mathcal{A})\geq\bar{H}(P,\mathcal{A},\mathcal{C})\geq H_\kappa(P,\mathcal{C})-H_\kappa(P,\mathcal{A}). \tag{10}$$

For a graph G, if $\mathcal{A}$ is the set of its cliques and $\mathcal{C}$ is the set of singletons from $2^{\mathcal{X}}$, $\mathcal{B}$ is the set of independent sets of that graph. For this scenario, an investigation into the cases where the bounds in (10) are achieved with equality brought out some remarkable connections between structural properties of graphs and such information theoretic concepts:

- *Strong Splitting:* For a graph G (10) is achieved with equality for all P if and only if the graph is perfect, that is the chromatic and clique numbers are equal for all subgraphs of G [8].
- *Weak Splitting:* For a graph G, there exists a nowhere vanishing P such that both bounds in (10) are tight if and only if G is normal, that is there is a family of cliques $\mathfrak{A}_0$ and a family of independent sets $\mathfrak{B}_0$ such that each covers $\mathcal{X}$ and the intersection $a\cap b$ is nonempty for all $a\in\mathfrak{A}_0$ and $b\in\mathfrak{B}_0$ [9] [10].

We extend the notion of weak splitting in two ways (albeit no longer in a purely structural manner): we allow arbitrary families $\mathcal{A},\mathcal{C}$, and we only require $\mathcal{B}\subset\bar{\mathcal{A}}(\mathcal{C})$.

Theorem 2: Given sets $\mathcal{A},\mathcal{C}$ and $\mathcal{B}\subset\bar{\mathcal{A}}(\mathcal{C})$,

$$H_\kappa(P,\mathcal{A})+H_\kappa(P,\mathcal{B})=H_\kappa(P,\mathcal{C}) \tag{11}$$

for some nowhere vanishing P if and only if

1. There exist nondegenerate families $\mathcal{A}_0\subset\tilde{\mathcal{A}}$ and $\mathcal{B}_0\subset\tilde{\mathcal{B}}$ such that $a\cap b\neq\emptyset,\forall a\in\mathcal{A}_0,b\in\mathcal{B}_0$. Let $\mathcal{C}_0\triangleq\{a\cap b:a\in\mathcal{A}_0,b\in\mathcal{B}_0\}$.
2. There exists a function $r\colon\mathcal{X}\to(0,\infty)$ such that

$$\begin{aligned}\textstyle\sum_{x\in c} r(x)=1,\forall c\in\mathcal{C}_0\\ \textstyle\sum_{x\in c} r(x)\leq 1,\forall c\notin\mathcal{C}_0\end{aligned} \tag{12}$$

Remark 1: The conditions above imply that $\mathcal{A}_0,\mathcal{B}_0,\mathcal{C}_0$ are families of maximal sets in $\mathcal{A},\mathcal{B},\mathcal{C}$. Hence, our theorem extends Lemma 3 in [11], which gives necessary conditions for (11) in the case where $\mathcal{A},\mathcal{B}$ are the families of independent sets of graphs G and G' and $\mathcal{C}$ is $\{a\cap b:a\in\mathcal{A},b\in\mathcal{B}\}$, the family of independent sets of $G\cup G'$. Note that $\mathcal{B}$ can be a proper subset of $\bar{\mathcal{A}}(\mathcal{C})$ ($\mathcal{B}=\bar{\mathcal{A}}(\mathcal{C})$ if and only if the graphs are edge disjoint).

Remark 2: For the noiseless coding with side-information scenario, the first condition reduces to the condition for normality of the confusability graph. The second condition is trivially satisfied by the function $r(x)=1,\forall x\in\mathcal{X}$.

Proof:

Sufficiency: Let $\mathcal{A}_0$, $\mathcal{B}_0$ and $r(\cdot)$ be as above. Let $P(a)$ and $Q(b)$ be distributions on $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ that have $\mathcal{A}_0$ and $\mathcal{B}_0$ as their support. To keep the notation simple we shall represent probabilities $p_T(t)$ by $p(t)$. A similar streamlining of notation is employed for conditional probabilities as well. Let $\mathcal{A}(x) = \{a \in \tilde{\mathcal{A}} : a \ni x\}$. Similarly define $\mathcal{C}(x)$ and $\mathcal{B}(x)$. Define the set of random variables (X, S, Y, Z) on $\mathcal{X} \times \mathcal{B} \times \mathcal{A} \times \mathcal{C}$ such that

$$p(x, b, a, c) = P(a)Q(b)\delta(c, a \cap b)p(x|c), \tag{13}$$

where $\delta(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$, and

$$p(x|c) = \begin{cases} 0 & \text{if } x \notin c \\ r(x) & \text{if } x \in c \in \mathcal{C}_0 \\ \text{arbitrary} & \text{(but leading to valid probability) otherwise} \end{cases}$$

This probability induces

$$p(a, b|x) = \begin{cases} \frac{P(a)Q(b)r(x)}{\sum_{a' \in \mathcal{A}(x) \cap \mathcal{A}_0} \sum_{b' \in \mathcal{B}(x) \cap \mathcal{B}_0} P(a')Q(b')r(x)} & \text{if } x \in a \cap b \in \mathcal{C}_0 \\ 0 & \text{otherwise} \end{cases},$$

which reduces to

$$p(a, b|x) = \frac{P(a)}{\sum_{a' \in \mathcal{A}(x) \cap \mathcal{A}_0} P(a')} \frac{Q(b)}{\sum_{b' \in \mathcal{B}(x) \cap \mathcal{B}_0} Q(b')} = p(a|x)p(b|x).$$

This means that $Y \leftrightarrow X \leftrightarrow S$ is a Markov chain.

For this set of random variables, we now have

$$I(X; Z) - I(X; Y) \stackrel{a}{=} I(X; S|Y) \stackrel{b}{=} I(X; S) - I(S; Y) \stackrel{c}{=} I(X; S),$$

where (a) follows from $Z = Y \cap S$ and $X - Y \cap S - (Y, S)$, (b) follows from $Y \leftrightarrow X \leftrightarrow S$ and (c) follows from the independence of Y and S.

For any non-degenerate family $\mathcal{T}$ and a distribution P on $\mathcal{X}$, the Kuhn-Tucker (KT) conditions necessary and sufficient for a pair of random variables (X, T) to achieve $H_\kappa(P, \mathcal{T})$ are [5]

$$\left.\begin{array}{rcl} \sum_{x \in t} \frac{p(x)}{\sum_{t' \ni x} p(t')} & = & 1 \text{ if } t \in \mathcal{T}_0 \\ \sum_{x \in t} \frac{p(x)}{\sum_{t' \ni x} p(t')} & \leq & 1 \text{ if } t \notin \mathcal{T}_0, \end{array}\right\} \tag{14}$$

where $\mathcal{T}_0$ is the support of T. Using these one can show that the pairs $(X, Y), (X, S)$ and (X, Z) as defined above achieve $H_\kappa(P, \mathcal{A}), H_\kappa(P, \mathcal{B})$ and $H_\kappa(P, \mathcal{C})$ respectively (See appendix II). Therefore equation (11) is satisfied.

Necessity: Let Y, S and Z achieve $H(P, \mathcal{A}), H(P, \mathcal{B})$ and $H(P, \mathcal{C})$ for some nowhere vanishing P when equation (11) is satisfied. We may impose $S \leftrightarrow X \leftrightarrow Y$ since this does not change the quantities above. We now have

$$I(X; Z) = I(X; S) + I(X; Y) = I(Y, S; X) + I(Y; S) = I(X; Y \cap S) + I(X; Y, S|Y \cap S) + I(Y; S).$$

Since $Y \cap S$ is a distribution on $\tilde{\mathcal{C}}$ such that $X \in Y \cap S$ and Z is the minimizer of $I(X; \tilde{Z})$, we have

1. $Y \cap S$ is a minimizer of $I(X; \tilde{Z})$.
2. $X - Y \cap S - (Y, S)$
3. Y and S are independent, which implies $Y \cap S \neq \emptyset$ with probability 1.

These conditions are equivalent to those in the claim with $r(x) = p(x|c)$ for any c in the support of Z. ∎

Appendix

I. Proof of Lemma 4

Let the family of forbidden sets, $\mathcal{F}_\mathcal{C}$, be the family of sets that are minimal among the sets that are not covered by any $c \in \mathcal{C}$.

$$\mathcal{F}_\mathcal{C} = \{b \subset \mathcal{X} : b \notin \tilde{\mathcal{C}}, b' \in \tilde{\mathcal{C}}, \forall b' \subsetneq b\}$$

The proof of Lemma 4 relies on the following:

Lemma 5: $\bar{\bar{\mathcal{A}}}(\mathcal{C}) = \{a \subset \mathcal{X} : \forall a' \subset a, a' \in \mathcal{F}_\mathcal{C} \Rightarrow y' \in \tilde{\mathcal{A}}\}$

Proof: Let $\mathcal{A}_c \triangleq$ RHS , $\mathcal{B} \triangleq \bar{\mathcal{A}}(\mathcal{C})$.

$\boxed{\mathcal{A}_c \subseteq \bar{\bar{\mathcal{A}}}(\mathcal{C})}$: Let $a \in \mathcal{A}_c, b \in \mathcal{B}$. Now

$$\begin{aligned} & a \cap b \notin \tilde{\mathcal{C}} \\ \Rightarrow\ & \exists f \subset a \cap b : f \in \mathcal{F}_\mathcal{C} \\ \Rightarrow\ & \exists a' \in \mathcal{A} : f \subset a' \\ \Rightarrow\ & a' \cap b \notin \tilde{\mathcal{C}}, \end{aligned}$$

which is impossible from the definition of $\mathcal{B}$.

$\boxed{\mathcal{A}_c \supseteq \bar{\bar{\mathcal{A}}}(\mathcal{C})}$: Let $a \in \bar{\bar{\mathcal{A}}}(\mathcal{C})$ contain a forbidden set, say c, that is not a subset of any $a \in \mathcal{A}$. Since c is minimal, it is a member of $\mathcal{B}$. But $a \cap c = c$ is not a member of $\tilde{\mathcal{C}}$, which contradicts the assumption about a. ∎

In $\mathcal{X}^n$, the maximal $\mathcal{C}^n$ subsets are Cartesian products of maximal sets in $\mathcal{C}$.

For $b^n \subset \mathcal{X}^n$, let $\hat{b}^n$ denote the smallest set containing b^n that is a Cartesian product of sets. $\hat{b}^n = b_1 \times \cdots \times b_n$, where $b_i = \{x \in \mathcal{X} : x_i = x,$ for some $x^n \in b\}$. If c^n is the Cartesian product of n subsets of $\mathcal{X}$, $b^n \subset c^n$ if and only if $\hat{b}^n \subset c^n$. So $b^n \notin \widetilde{\mathcal{C}^n} \Leftrightarrow \hat{b}^n \notin \widetilde{\mathcal{C}^n} \Leftrightarrow b_i \notin \tilde{\mathcal{C}}$ for some i.

Proof: [of lemma 4]

Let $a^n \in \left[\bar{\bar{\mathcal{A}}}(\mathcal{C})\right]^n$. Let f^n be any subset of a^n that belongs to $\mathcal{F}_{\mathcal{C}^n}$ ($\Rightarrow \hat{f}^n \subset a^n \Rightarrow f_i \subset a_i$). Since f^n does not belong to $\widetilde{\mathcal{C}^n}$, at least one of the f_i does not belong to $\tilde{\mathcal{C}}$. For all i, f_i is either a singleton or $f_i \in \mathcal{F}_\mathcal{C}$: if $f_i \in \tilde{\mathcal{C}}$ and $f_i' \neq \emptyset \subsetneq f_i$, $f_1 \times \cdots \times f_i' \times \cdots \times f_n \cap f^n$ is a proper subset of f^n that cannot be covered by a set in $\tilde{\mathcal{C}}^n$. On the other hand if $f_i \notin \tilde{\mathcal{C}}$ and $f_i \notin \mathcal{F}_\mathcal{C}$, by replacing f_i with some element of $\mathcal{F}_\mathcal{C}$ that it contains, we can obtain a proper subset of f^n that cannot be covered by an element of $\tilde{\mathcal{C}}^n$. By choosing $a_i' \in \mathcal{A}$ that contains f_i at those co-ordinates i where f_i is some element of $\mathcal{F}_\mathcal{C}$ and by choosing some $a_i' \in \mathcal{A}$ that contains the single element of f_i at the other co-ordinates, we see that the forbidden subsets of a^n are contained in elements of $\mathcal{A}^n$. Therefore $\left[\bar{\bar{\mathcal{A}}}(\mathcal{C})\right]^n \subset \overline{\overline{\mathcal{A}^n}}(\mathcal{C}^n)$ and since $\mathcal{A}^n \subset \left[\bar{\bar{\mathcal{A}}}(\mathcal{C})\right]^n$, we have that all three have the same complement with respect to $\mathcal{C}^n$.

∎

II. KT conditions

Let (X, Y, S, Z) be r.v. distributed as in equation 13. The induced $p(a|x)$ satisfies the KT conditions for $\min_{X \in Y \in \tilde{A}} I(X;Y)$. To see this consider the KT conditions

$$\left.\begin{aligned} \sum_{x \in a} \frac{p(x)}{\sum_{a' \ni x} P(a')} &= 1 \text{ if } a \in \mathcal{A}_0 \\ \sum_{x \in a} \frac{p(x)}{\sum_{a' \ni x} P(a')} &\leq 1 \text{ if } a \notin \mathcal{A}_0. \end{aligned}\right\} \tag{15}$$

Now, $\displaystyle\sum_{x\in a}\frac{p(x)}{\sum_{a'\ni x}P(a')}=\sum_{x\in a}\frac{\sum_{a'\in\mathcal{A}(x)\cap\mathcal{A}_0}\sum_{b'\in\mathcal{B}(x)\cap\mathcal{B}_0}P(a')Q(b')r(x)}{\sum_{a'\in\mathcal{A}(x)\cap\mathcal{A}_0}P(a')}$

$$=\sum_{x\in a}\sum_{b'\in\mathcal{B}(x)\cap\mathcal{B}_0}r(x)Q(b')=\sum_{b\in\mathcal{B}_0}Q(b)\sum_{x\in a\cap b}r(x)$$

By the conditions on $r(x)$, the KT conditions here are satisfied and Y is the optimal reproduction. By symmetry, S is the optimal reproduction for $\min_{X\in S\in\tilde{\mathcal{B}}} I(X;S)$.

Finally, consider the KT conditions for achieving $H(P,\mathcal{C})$:

$$\left.\begin{array}{l}\sum_{x\in c}\frac{p(x)}{\sum_{c'\ni x}p(c')}=1,\forall c:p(c)>0\\ \sum_{x\in c}\frac{p(x)}{\sum_{c'\ni x}p(c')}\leq 1,\forall c:p(c)=0\end{array}\right\} \tag{16}$$

But, $\displaystyle\frac{p(x)}{\sum_{c''\in\mathcal{C}(x)\cap\mathcal{C}_0}p(c')}=\frac{\sum_{c''}p(c'')p(x|c'')}{\sum_{c''\in\mathcal{C}(x)\cap\mathcal{C}_0}p(c')}=\frac{\sum_{c''\in\mathcal{C}(x)\cap\mathcal{C}_0}p(c'')r(x)}{\sum_{c'\in\mathcal{C}(x)\cap\mathcal{C}_0}p(c')}=r(x).$

Therefore, since $p(c)>0$ only if $c\in\mathcal{C}_0$, the KT conditions are satisfied and we have

$$H(P,\mathcal{A})+H(P,\mathcal{B})=H(P,\mathcal{C}).$$

References

[1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Information Theory*, vol.22, no.1, pp.1-10, Jan. 1976.

[2] H. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. on Information Theory*, vol.22, (no. 5), pp.592-593, Sep. 1976.

[3] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. on Information Theory*, vol.42, (no.5), pp. 1329-39, Sep. 1996.

[4] P. Koulgi, E. Tuncel, S. L. Regunathan and K. Rose, "On zero-error source coding with decoder side-information," *IEEE Trans. on Information Theory*, vol.49, no.1, pp.99-111, Jan. 2003.

[5] E. Tuncel, P. Koulgi, S. Regunathan and K. Rose, "Zero-error source coding with maximum distortion criterion," *Proc. Data Compression Conference 2002*

[6] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. on Information Theory*, vol. 40, pp.1670-72, Sep. 1996.

[7] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Academic Press, New York, 1982.

[8] I. Csiszár, J. Körner, L. Lovász, K. Marton and G. Simonyi, "Entropy splitting for anti-blocking corners and perfect graphs," *Combinatorica*, vol. 10, no. 1, 1990.

[9] J. Körner and G. Longo, "Two-step encoding of finite memoryless sources," *IEEE Trans. on Information Theory*, vol.IT-19, (no.6), pp.778-82, Nov. 1973.

[10] J. Körner and K. Marton, "Graphs that split entropies," *SIAM J. Discrete Mathematics*, vol. 1, pp.71-79, 1988

[11] J. Körner, G. Simonyi and Zs. Tuza, "Perfect couples of graphs," *Combinatorica*, vol. 12, no. 2, 1992.