On Zero-Error Coding of Correlated Sources

Prashant Koulgi, Ertem Tuncel and Kenneth Rose[1]
Dept of ECE, University of California Santa Barbara, CA 93106, e-mail: {prashant,ertem,rose}@ece.ucsb.edu

Let the correlated information sequences $X_0, X_1, \cdots$ and $Y_0, Y_1, \cdots$ be generated by repeated independent drawings of a pair of finite random variables (X, Y) from a given bivariate distribution $P(x, y)$ on the product set $\mathcal{X} \times \mathcal{Y}$. We study the problem of *zero-error* coding of the correlated sources (X, Y). Thus if a code of block length n, comprising the separate encoders $\phi_X : \mathcal{X}^n \to \{0,1\}^*$ and $\phi_Y : \mathcal{Y}^n \to \{0,1\}^*$, and a decoder $\psi : \{0,1\}^* \times \{0,1\}^* \to \mathcal{X}^n \times \mathcal{Y}^n$, is a valid uniquely decodable code for this problem, then

$$\psi(\phi_X(x^n), \phi_Y(y^n)) = (x^n, y^n) \tag{1}$$

for every pair of vectors (x^n, y^n) with $P^n(x^n, y^n) > 0$. Let $G = (\mathcal{X} \cup \mathcal{Y}, E)$ be the *characteristic bipartite graph* of (X, Y), obtained by setting $\{x, y\} \in E$ iff $P(x, y) > 0$. Let $P_X(x)$ and $P_Y(y)$ denote the marginals of $P(x, y)$. We derive bounds for the region of rate pairs achievable by such codes for some length n. Since it can be shown that this region depends only on (G, P_X, P_Y), we shall denote it by $\mathcal{R}(G, P_X, P_Y)$.

Note that exact determination of $\mathcal{R}(G, P_X, P_Y)$ is at least as hard as known long-standing open problems. The special case of zero-error coding when the decoder has access to unknown side-information is already equivalent to Shannon's well-known zero-error capacity problem, which is currently unresolved. On the other hand, the problem of coding (X, Y) with vanishing error probability was completely resolved by Slepian and Wolf in [1].

Preliminary definitions: $\mathcal{C} = (\mathcal{C}_X, \mathcal{C}_Y)$ is a *bipartite cover* of G if $\mathcal{C}_X$ ($\mathcal{C}_Y$) is a partition of $\mathcal{X}$ ($\mathcal{Y}$), and $s \in \mathcal{C}_X, t \in \mathcal{C}_Y \Rightarrow s \cup t$ induces at most one edge in G. $\mathcal{T}(G_X)$ denotes the collection of maximal complete subgraphs of G_X, the graph on $\mathcal{X}$ obtained by connecting x and x' if there exists y such that $\{x, y\} \in E$ and $\{x', y\} \in E$.

$\mathcal{R}^{in}(G, P_X, P_Y)$, an inner bound for $\mathcal{R}(G, P_X, P_Y)$, is given by the closure of the set of all pairs (R_X, R_Y) with

$$R_X > I(X; S|Q), R_Y > I(Y; T|Q), \tag{2}$$

for some choice of the random variables Q, S and T jointly distributed with X and Y according to $p(q)p(s|x,q)p(t|y,q)P_X(x)P_Y(y)$, satisfying: for each q, $p(s|x,q) > 0$ only if $s \in \mathcal{C}_{X,q}$ and $x \in s$, and $p(t|y,q) > 0$ only if $t \in \mathcal{C}_{Y,q}$ and $y \in t$, where each $(\mathcal{C}_{X,q}, \mathcal{C}_{Y,q}) = \mathcal{C}_q$ is a bipartite cover of G. (We may assume that the time-sharing random variable Q is distributed over $\{1, 2, 3\}$.)

$\mathcal{R}^{out}(G, P_X, P_Y)$ is an outer bound for $\mathcal{R}(G, P_X, P_Y)$, and consists of the closure of the set of all pairs (R_X, R_Y) satisfying

$$R_X > \max_{X \in S_X \in \mathcal{T}(G_X)} H(X|S_X), \quad R_Y > \max_{Y \in T_Y \in \mathcal{T}(G_Y)} H(Y|T_Y),$$
$$\text{and } R_X + R_Y > \max_{\{X', Y'\} \in E, X' \sim P_X, Y' \sim P_Y} H(X', Y'). \tag{3}$$

[1] This work is supported in part by the NSF under grants no. EIA-9986057 and EIA-0080134, the University of California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, Inc., and Qualcomm, Inc.

While the outer bound $\mathcal{R}^{out}(G, P_X, P_Y)$ follows quite easily from the results of Slepian and Wolf in [1], the inner bound $\mathcal{R}^{in}(G, P_X, P_Y)$ is one of our main results. In order to derive this bound, we consider the distortion measure $d_E : \mathcal{X} \times \mathcal{Y} \times E \to [0, 1]$, defined by $d_E(x, y, \hat{e}) = 1$ if $\{x, y\} \in E$ and $\{x, y\} \neq \hat{e}$, and $d_E(x, y, \hat{e}) = 0$ otherwise. Then codes (comprising two separate encoders and a common decoder) which guarantee reproduction with zero distortion without error are also codes in the sense of (1). The inner bound is then obtained as a special case of our single-letter characterization of the following general zero-error multi-terminal rate-distortion problem (which is new): given an arbitrary distortion measure $d : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to [0, \infty)$, what is the region of rate pairs achievable by codes which guarantee reproduction of (X, Y) with zero distortion *without error*?

Consider next tightness of the bounds. For a given triple (G, P_X, P_Y), let $R^{low}(G, P_X, P_Y)$ and $R^{up}(G, P_X, P_Y)$ denote, respectively, the minimum of $R_1 + R_2$ over rate pairs (R_1, R_2) of the regions $\mathcal{R}^{out}(G, P_X, P_Y)$ and $\mathcal{R}^{in}(G, P_X, P_Y)$. We first consider coincidence of the numbers $R^{low}(G, P_X, P_Y)$ and $R^{up}(G, P_X, P_Y)$, which is clearly necessary for coincidence of the regions $\mathcal{R}^{out}(G, P_X, P_Y)$ and $\mathcal{R}^{in}(G, P_X, P_Y)$. Fix the bipartite graph G, and consider the family of joint distributions $P(x, y)$ on $\mathcal{X} \times \mathcal{Y}$ such that the characteristic bipartite graph of P is G. We show that there exists at least one member of this family with $R^{low}(G, P_X, P_Y) = R^{up}(G, P_X, P_Y)$ if and only if G has an exact bipartite cover. (The bipartite cover $\mathcal{C} = (\mathcal{C}_X, \mathcal{C}_Y)$ is exact if, for every $s \in \mathcal{C}_X$ and $t \in \mathcal{C}_Y$, $s \cup t$ induces exactly one edge in G.) $R^{low}(G, P_X, P_Y) = R^{up}(G, P_X, P_Y)$ for every member of this family if and only if G is a disjoint collection of complete bipartite graphs. (G is complete if every $\{x, y\}$ is an edge.) Thus we obtain information-theoretic characterizations of the purely combinatorial properties of possession of exact bipartite covers, and of completeness. This is analogous to the information-theoretic characterization of normal graphs and perfect graphs by Körner and others (see [2]).

Consider the following simple encoding scenario, which we term *successive encoding*: ϕ_X performs lossless encoding of X, at rate $H(X)$, and ϕ_Y is designed assuming that the decoder has recovered X. For coding with vanishing error probability, the results of Slepian and Wolf in [1] show that time-sharing of the two possible successive encoding scenarios (obtained by interchanging the roles of X and Y) is optimal, in that all achievable rate pairs can be achieved via this strategy. But time-sharing combined with successive encoding may not be an optimal strategy in the case of zero-error coding. We construct a simple example for this suboptimality phenomenon.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, (no.4), pp. 471-80, July 1973.
[2] G. Simonyi, "Graph entropy: a survey," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol.20, pp.399-441, 1995.

0-7803-7501-7/02/$17.00©2002 IEEE.