On Hierarchical Type Covering[1]

Ertem Tuncel
University of California, Riverside
E-mail: ertem@ee.ucr.edu

Jayanth Nayak
University of California, Santa Barbara
E-mail: jayanth@ece.ucsb.edu

Kenneth Rose
University of California, Santa Barbara
E-mail: rose@ece.ucsb.edu

Central to many rate-distortion theoretic results is the concept of *type covering*, i.e., covering of the set of vectors with the same type P by identical "distortion balls." For example, the achievability of the rate-distortion function $R_P(D)$ can be shown using the fact that it suffices to use $\approx 2^{nR_P(D)}$ balls with radius D to cover a type class T_P^n of length-n sequences [2]. Another example is guessing [1], where $\mathbf{X}$, drawn from a DMS, is guessed using a fixed sequence $\mathbf{y}(1), \mathbf{y}(2), \ldots$ until $d(\mathbf{X}, \mathbf{y}(i)) \leq D$. The optimal guessing list (minimizing expected i) is formed by sorting P in increasing $R_P(D)$, and concatenating centers of balls that cover each T_P^n.

In the *hierarchical* extension introduced in [3], T_P^n is first covered using D_1-balls, and then the portion of each D_1-ball (parent) that lies within T_P^n is covered using D_2-balls (children). We call this *strong* hierarchical type covering to distinguish it from the *weak* version introduced in [5], where the D_2-balls are chosen so that for all $\mathbf{x} \in T_P^n$, there exists a pair of parent D_1- and child D_2-balls both covering $\mathbf{x}$. The main use of hierarchical type covering in both [3] and [5] was the determination of achievable error exponents in scalable source coding. In fact, weak covering is sufficient for that purpose. Strong covering, on the other hand, is necessary for hierarchical guessing [4]: In the first stage, $\mathbf{X}$ is guessed using a fixed $\mathbf{y}_1(1), \mathbf{y}_1(2), \ldots$ until $d_1(\mathbf{X}, \mathbf{y}_1(i)) \leq D_1$, and in the second stage, new fixed guesses $\mathbf{y}_2(1|i), \mathbf{y}_2(2|i), \ldots$ are used until $d_2(\mathbf{X}, \mathbf{y}_2(j|i)) \leq D_2$. If the type $P_{\mathbf{X}}$ of $\mathbf{X}$ is *known* beforehand, the optimal strategy is to find a strong hierarchical covering of $T_{P_{\mathbf{X}}}^n$ achieving a balance point in the tradeoff of the required number of D_1- and D_2-balls.

The distinction between the two covering strategies motivated us to re-derive single-letter characterizations of the rates of D_1- and D_2-balls necessary and sufficient to cover T_P^n both weakly and strongly. Somewhat surprisingly, the two characterizations lead to different rate regions. In fact, the claimed rate region for strong covering that appeared in [3] is precisely the achievability region for weak covering[2].

Denote by $I(Q_1, V)$ the mutual information between Y_1 and X induced by $Q_1(y_1)V(x|y_1)$. Similarly, $I(Q_{2|1}, W|Q_1)$ denotes $I(Y_2; X|Y_1)$ induced by $Q_1(y_1)Q_{2|1}(y_2|y_1)W(x|y_1, y_2)$. For $P(x)$ and $Q_1(y_1)$, define $\mathcal{V}(P, Q_1, D_1)$ as the set of all $V(x|y_1)$ such that $\sum_{y_1} Q_1(y_1)V(x|y_1) = P(x)$ and $E_{Q_1 V}\{d_1(X, Y_1)\} \leq D_1$. Similarly, for $Q_1(y_1)$, $Q_{2|1}(y_2|y_1)$, and $V(x|y_1)$, define $\mathcal{W}(V, Q_{2|1}, Q_1, D_2)$ as the set of all $W(x|y_1, y_2)$ such that $\sum_{y_2} Q_{2|1}(y_2|y_1)W(x|y_1, y_2) = V(x|y_1)$ and $E_{Q_1 Q_{2|1} W}\{d_2(X, Y_1)\} \leq D_2$. Now, define $\mathcal{R}_P(D_1, D_2)$ as the region of all (R_1, R_2) such that $I(Q_1, V) \leq R_1$ and $I(Q_1, V) + I(Q_{2|1}, W|Q_1) \leq R_2$ for some Q_1, $Q_{2|1}$, $V \in \mathcal{V}(P, Q_1, D_1)$, and $W \in \mathcal{W}(V, Q_{2|1}, Q_1, D_2)$. Also define $\mathcal{T}_P(D_1, D_2)$ as the rate pairs (R_1, R_2) such that there exists $Q_1(y_1)$ satisfying

$$I_m(P||Q_1, D_1) \triangleq \inf_{V \in \mathcal{V}(P, Q_1, D_1)} I(Q_1, V) \leq R_1 \ ,$$

and for all $V \in \mathcal{V}(P, Q_1, D_1)$, there exists $Q_{2|1}(y_2|y_1)$ with

$$I_m(V||Q_{2|1}, Q_1, D_2) \triangleq \inf_{W \in \mathcal{W}(V, Q_{2|1}, Q_1, D_2)} I(Q_{2|1}, W|Q_1) \leq R_2 - R_1.$$

The next lemma shows that $\mathcal{T}_P(D_1, D_2)$ completely characterizes achievable rates for strong covering.

Lemma 1: For any $\{\mathbf{y}_1(i)\}_{i=1}^{M_1}$ and $\{\mathbf{y}_2(j|i)\}_{j=1}^{M_2}$ that strongly covers T_P^n, there exists $Q_1(y_1)$ such that

$$\begin{aligned} \frac{1}{n} \log M_1 + \epsilon(n) &\geq I_m(P||Q_1, D_1) \\ \frac{1}{n} \log M_2 + \epsilon(n) &\geq \max_{V \in \mathcal{V}(P, Q_1, D_1)} \min_{Q_{2|1}} I_m(V||Q_{2|1}, Q_1, D_2) \end{aligned}$$

where $\epsilon(n) \to 0$ as $n \to \infty$. Conversely, if $(R_1, R_2) \in \mathcal{T}_P(D_1, D_2)$, then there exist $\{\mathbf{y}_1(i)\}_{i=1}^{M_1}$ and $\{\mathbf{y}_2(j|i)\}_{j=1}^{M_2}$ with

$$\begin{aligned} \frac{1}{n} \log M_1 &\leq R_1 + \epsilon \\ \frac{1}{n} \log M_2 &\leq R_2 - R_1 + \epsilon \end{aligned}$$

strongly (D_1, D_2)-covering T_P^n, for any $\epsilon > 0$ and large n.

We next derive achievable rates for weak covering.

Lemma 2: For any $\{\mathbf{y}_1(i)\}_{i=1}^{M_1}$ and $\{\mathbf{y}_2(j|i)\}_{j=1}^{M_2}$ that covers T_P^n weakly,

$$\left(\frac{1}{n} \log M_1 + \epsilon(n), \frac{1}{n} \log M_1 M_2 + \epsilon(n) \right) \in \mathcal{R}_P(D_1, D_2) \ ,$$

where $\epsilon(n) \to 0$ as $n \to \infty$. Conversely, if $(R_1, R_2) \in \mathcal{R}_P(D_1, D_2)$, there exist $\{\mathbf{y}_1(i)\}_{i=1}^{M_1}$ and $\{\mathbf{y}_2(j|i)\}_{j=1}^{M_2}$ with

$$\begin{aligned} \frac{1}{n} \log M_1 &\leq R_1 + \epsilon \\ \frac{1}{n} \log M_1 M_2 &\leq R_2 + 2\epsilon \end{aligned}$$

weakly (D_1, D_2)-covering T_P^n, for any $\epsilon > 0$ and large n.

By definition, we have $\mathcal{T}_P(D_1, D_2) \subseteq \mathcal{R}_P(D_1, D_2)$. However, we were able to build an example where strict inclusion $\mathcal{T}_P(D_1, D_2) \subset \mathcal{R}_P(D_1, D_2)$ holds.

References

[1] E. Arikan and N. Merhav. Guessing subject to distortion. *IEEE Trans. on Info. Theory*, 44(3):1041–1056, May 1998.

[2] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic, NY, 1982.

[3] A. Kanlis and P. Narayan. Error exponents for successive refinement by partitioning. *IEEE Trans. on Info. Theory*, 42(1):275–282, Jan. 1996.

[4] N. Merhav, R. M. Roth, and E. Arikan. Hierarchical guessing with a fidelity criterion. *IEEE Trans. on Info. Theory*, 45(1):330–337, Jan. 1999.

[5] E. Tuncel and K. Rose. Error exponents in scalable source coding. *IEEE Trans. on Info. Theory*, 49(1):289–296, Jan. 2003.

[1]This work was supported in part by the NSF under grants EIA-9986057 and EIA-0080134, the UC MICRO program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Microsoft Corporation, Mindspeed Technologies, and Qualcomm, Inc.

[2]Private communication with Tamas Linder and Prakash Narayan revealed that if covering with D_1-balls is replaced by covering with V-shells, which are subsets of D_1-balls, then a variant of strong covering can achieve the weak covering rates. However, this variant is not useful in hierarchical guessing as defined in [4].