On Hierarchical Type Covering

Ertem Tuncel\textsuperscript{1}, Jayanth Nayak\textsuperscript{2}, and Kenneth Rose\textsuperscript{2}

\textsuperscript{1}University of California, Riverside
\textsuperscript{2}University of California, Santa Barbara
Types and Typical Sequences

- Type class $T^P_n$: Collection of data points $\mathbf{x}$ with empirical distribution $P$, i.e.,
  $$\frac{1}{n} N(a \mid \mathbf{x}) = P(a)$$

- Conditional type class $T^n_V(y)$ (a.k.a. $V$-shell): Collection of data points $\mathbf{x}$ with empirical conditional distribution $V(x \mid y)$ w.r.t. $y$, i.e.,
  $$\frac{1}{n} N(a, b \mid \mathbf{x}, y) = \frac{1}{n} N(b \mid y)V(a \mid b)$$
Type Covering

- Covering of $T_P^n$ using “distortion spheres” about codevectors $y(i), i = 1, 2, \ldots, M$, i.e.,

$$T_P^n \subset \bigcup_{i=1}^{M} S(y(i), D)$$

where

$$S(y(i), D) = \{ x : d(x, y(i)) \leq D \}$$

- Type covering lemma (Csiszár and Körner) states that there exists a covering strategy with

$$\frac{1}{n} \log M \leq R_P(D) + o(n)$$
Uses of Type Covering

- **Source coding:**
  - Proof of achievability of the rate-distortion curve $R_p(D)$
  - Proof of universal achievability of Marton’s error exponent $E_p(D, R)$ in lossy source coding
  - Results on rate redundancy

- **Guessing subject to distortion:** Alice presents Bob fixed guesses $y(1), y(2), \ldots$ until $d(x, y(i)) \leq D$
  - Minimum guessing exponent $\frac{1}{n} \log E\{i(X)^\rho\}$ is achieved by the concatenation of type-covering codebooks in a certain order.
2-Stage Type Covering

- Two possibilities:
  - **Strong covering:** Introduced by Kanlis and Narayan, is the covering of $T^n_P$ using spheres $S(y_1(i), D_1), \ 1 \leq i \leq M_1$ followed by a covering of each $T^n_P \cap S(y_1(i), D_1)$ using spheres $S(y_2(j \mid i), D_2), \ 1 \leq j \leq M_2$.
  
  - **Weak covering:** Implicit in Tuncel and Rose’s work on error exponents in scalable source coding, is the building of a tree of spheres $S(y_1(i), D_1)$ and $S(y_2(j \mid i), D_2)$ such that for each $x$, at least one parent-child pair $(i, j)$ simultaneously covers $x$. 

TUNCEL, NAYAK, AND ROSE: ON HIERARCHICAL TYPE COVERING
Illustration of the Difference

Every \( x \in T^n_p \) is covered by at least one parent-child pair
Need for Strong vs. Weak

- The difference was previously overlooked.
- Even though strong covering was introduced in the context of scalable source coding, weak covering suffices for that purpose.
- Strong covering, on the other hand, is useful in the generalization of the guessing game, introduced by Merhav, Arikan, and Roth:
  - Alice presents guesses $y_1(1), y_1(2), \ldots$ until $d_1(x, y_1(i)) \leq D_1$ and then guesses $y_2(1 \mid i), y_2(2 \mid i), \ldots$ until $d_2(x, y_2(j \mid i)) \leq D_2$
  - Minimum guessing exponent $\frac{1}{n} \log E\{[i(X) + j(X)]^\rho\} = ?$
Main Results

- We derive necessary and sufficient conditions for achievability of covering rates $\frac{1}{n} \log M_1$ and $\frac{1}{n} \log M_2$ for both strategies.
  - The claimed rate region in Kanlis and Narayan’s work for strong covering is in fact the rate region for weak covering.
    - All the type covering based results regarding error exponents remain valid, as weak covering is sufficient for those analyses.
  - Strong covering rates can be strictly higher than weak covering rates (shown by example).
    - The minimum achievable hierarchical guessing exponent result must be re-examined, as it relies on strong covering.
Weak Covering Rates

Let $\mathcal{R}_P(D_1, D_2)$ denote the region of all $(R_1, R_2)$ satisfying the following:

$\exists \ Q_1(y_1), \ Q_{2|1}(y_2|y_1), \ V(x|y_1), \text{ and } W(x|y_1, y_2) \text{ such that}$

- $V \in \mathcal{V}(P, Q_1, D_1)$, i.e., $V$ and $Q_1$ are consistent with $P$ and induce distortion $\leq D_1$,
- $W \in \mathcal{W}(V, Q_{2|1}, Q_1, D_2)$, i.e., $W$ and $Q_{2|1}$ are consistent with $V$, and (together with $Q_1$) induce distortion $\leq D_2$,

$\begin{align*}
I(X;Y_1) &= I(Q_1, V) \leq R_1 \\
I(X;Y_1) + I(X;Y_2|Y_1) &= I(Q_1, V) + I(Q_{2|1}, W|Q_1) \leq R_2
\end{align*}$
Weak Covering Rates

If \((R_1, R_2) \in \mathcal{R}_P(D_1, D_2)\), then there exists a weak covering with
\[
\frac{1}{n} \log M_1 \leq R_1 + \varepsilon
\]
\[
\frac{1}{n} \log M_1 M_2 \leq R_2 + \varepsilon
\]

Conversely, if there exists a weak covering with codebook sizes \(M_1\) and \(M_2\), then
\[
\left(\frac{1}{n} \log M_1 + \varepsilon, \frac{1}{n} \log M_1 M_2 + \varepsilon\right) \in \mathcal{R}_P(D_1, D_2)
\]
Strong Covering Rates

Let \( T_P(D_1,D_2) \) denote the region of all \((R_1,R_2)\) with \( R_2 \geq R_1 \) satisfying the following:

\[
\exists \ Q_1(y_1) \text{ such that } \min_{V \in \mathcal{V}(P,Q_1,D_1)} I(Q_1,V) \leq R_1,
\]

For all \( V \in \mathcal{V}(P,Q_1,D_1) \), \( \exists \ Q_{2|1}(y_2|y_1) \) such that

\[
\min_{W \in \mathcal{W}(V,Q_{2|1},Q_1,D_2)} I(Q_{2|1},W \mid Q_1) \leq R_2 - R_1
\]

Note that \( T_P(D_1,D_2) \subseteq R_P(D_1,D_2) \)
If \((R_1, R_2) \in \mathcal{T}_P(D_1, D_2)\), then there exists a strong covering with
\[
\frac{1}{n} \log M_1 \leq R_1 + \epsilon \\
\frac{1}{n} \log M_1 M_2 \leq R_2 + \epsilon
\]

Conversely, if there exists a strong covering with codebook sizes \(M_1\) and \(M_2\), then
\[
\left(\frac{1}{n} \log M_1 + \epsilon, \frac{1}{n} \log M_1 M_2 + 2\epsilon\right) \in \mathcal{T}_P(D_1, D_2)
\]
Example of $\mathcal{T}_P(D_1, D_2) \subset R_P(D_1, D_2)$

<table>
<thead>
<tr>
<th>x \ y</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

$P(x) = 1/3$

$D_1 = 0.06$

$D_2 = 0.016$

- When $R_1 = R_P(D_1)$, there exists a unique pair $(Q_1, V)$ such that $V \in \mathcal{V}(P, Q_1, D_1)$ and $I(Q_1, V) \leq R_1$

- Then it suffices to find $V' \in \mathcal{V}(P, Q_1, D_1)$ such that

$$\min_{Q_2 ||, W \in \mathcal{W}(V, Q_2 ||, Q_1, D_2)} I(Q_2 ||, W | Q_1) < \min_{Q_2 ||, W \in \mathcal{W}(V', Q_2 ||, Q_1, D_2)} I(Q_2 ||, W | Q_1)$$

- We found such $V'$

TUNCEL, NAYAK, AND ROSE: ON HIERARCHICAL TYPE COVERING
Consequences on Guessing

- Merhav, Arikan, and Roth proved the following lower bound for the minimum guessing exponent

\[
\liminf_{n \to \infty} \frac{1}{n} \log E\{[i(X) + j(X)]^\rho\} \geq \max_{\rho_1^*} [\rho K(D_1, D_2, P') - D(P' \| P)]
\]

where

\[
K(D_1, D_2, P) = \min_{Q_1, Q_2} \max_{V \in \mathcal{V}(P, Q_1, D_1), W \in \mathcal{W}(V, Q_2, Q_1, D_2)} \{ I(Q_1, V), I(Q_2, W \mid Q_1) \}
\]

- Using strong type covering, they also proved achievability of this lower bound when Alice knows the type of x in advance.
Consequences on Guessing

- Proof assumes every rate pair in $\mathcal{R}_P(D_1,D_2)$ is sufficient for covering.
- What is achievable, instead, is

$$\max_{P'} [\rho L(D_1, D_2, P') - D(P' \| P)]$$

where

$$L(D_1, D_2, P) = \min_{Q_1} \max_{Q_1} \left\{ \min_{V \in V(P,Q_1,D_1)} I(Q_1,V), \max_{V \in V(P,Q_1,D_1)} \min_{Q_2 | 1} I(Q_2 | 1, W | Q_1) \right\}$$

- For the example yielding $\mathcal{T}_P(D_1,D_2) \subset \mathcal{R}_P(D_1,D_2)$, we also show that $L(D_1,D_2,P) > K(D_1,D_2,P)$. 

TUNCEL, NAYAK, AND ROSE: ON HIERARCHICAL TYPE COVERING
To sum up...

- Contrasted two possibilities for hierarchical type covering.
- Presented rate regions for both cases.
  - Rate regions are not identical.
- Investigated the effect of this distinction on the guessing problem.