Source Coding and Function Computation: Optimal Rate in Zero-Error and Vanishing Zero-Error Regime

Solmaz Torabi
Dept. of Electrical and Computer Engineering
Drexel University
st669@drexel.edu
Advisor: Dr. John M. Walsh


Outline

- Introduction
  - Motivation
  - Some graph theory quantities

- Zero error source coding (Scalar vs. Block coding)
  - Fixed length coding
  - Variable length coding

- Asymptotically zero probability of error
  - Optimal rate for function compression
  - Achievable coding scheme

- Function compression in zero-error regime

- Compression of specific functions
Outline

- **Introduction**
  - Motivation
  - Some graph theory quantities

- Zero-error source coding (Scalar vs. Block coding)
  - Fixed length coding
  - Variable length coding

- Asymptotically zero probability of error
  - Optimal rate for function compression
  - Achievable coding scheme

- Compression of specific functions

- Function compression in zero-error regime
In big data problems
▶ How to design computational methodologies suited to statistical calculation
    ▶ Scale easily as the data being processed becomes immense
▶ Fundamental problems in massive datasets where some statistical primitives needs to be computed for some tasks.

▶ Users submit machine learning tasks
▶ Analytics servers translate the tasks into statistical primitives calculation
Motivation

- Central receiver wishes to compute the average temperature of the building.
- Bandwidth limited sensors do not communicate with each other.
- Considers the recovery not of the sources but the function of sources.
Outline

- **Introduction**
  - Motivation
  - Some graph theory quantities

- Zero error source coding (Scalar vs. Block coding)
  - Fixed length coding
  - Variable length coding

- Asymptotically zero probability of error
  - Optimal rate for function compression
  - Achievable coding scheme

- Function compression in zero-error regime

- Compression of specific functions
Some graph theory quantities

Quantities to be discussed for undirected graphs: (undirected, finite and contain neither loops nor multiple edges)

⇒ Independent set of the graph
⇒ Coloring of the graph
⇒ Product of the graph
An independent set or stable set is a set of vertices in a graph, no two of which are adjacent.

An independent set that is not the subset of another independent set is called maximal.
An independent set or stable set is a set of vertices in a graph, no two of which are adjacent.

An independent set that is not the subset of another independent set is called maximal.
An independent set or stable set is a set of vertices in a graph, no two of which are adjacent.

An independent set that is not the subset of another independent set is called maximal.
An independent set or stable set is a set of vertices in a graph, no two of which are adjacent.

An independent set that is not the subset of another independent set is called maximal.
An independent set or stable set is a set of vertices in a graph, no two of which are adjacent.

An independent set that is not the subset of another independent set is called maximal.
Probabilistic Graph-Chromatic Number

- $\chi(G)$ is the smallest number of colors needed to color the vertices of $G$ so that no two adjacent vertices share the same color.

- The chromatic number of a graph $\geq$ clique number (maximum clique size).

\[
\begin{align*}
\chi(G) &= 3 \\
\end{align*}
\]
Entropy of the coloring

- Random variable \( X \sim P(x) \), and function \( c : \mathcal{X} \to \mathbb{N} \)

- Entropy of the coloring is defined as

\[
H(c(X)) = - \sum_{i \in c(\mathcal{X})} p_i \log p_i
\]

\[
p_i = P(c^{-1}(i)) = \sum_{x \in V : c(x) = \gamma} P(x)
\]

- \( c^{-1}(i) \) is a color class (a set of vertices assigned the same color \( i \))

- Coloring of \( G \) is any function \( c \) over \( V \) such that \( c^{-1}(.) \) induced a partition of \( V \) into independent sets of \( G \)

\[
H_{\mathcal{X}}(G, X) = \min \{ H(c(X)) : c \text{ is a coloring of } G \}
\]
Chromatic Entropy

Note that $\chi_{H(G,P)} \geq \chi_G$

\[
\begin{align*}
H(0.333, 0.333, 0.333) &= 1.58 \\
H(0.616, 0.333, 0.05) &= 1.17 \\
H(0.85, 0.05, 0.05, 0.05) &= 0.84
\end{align*}
\]
Chromatic Entropy

- Chromatic entropy is known to be NP-hard \(^1\)

- Find independent sets with high probability mass and establish these as color classes,
  - Unbalance the distribution as much as possible

- Add a node to the color class with the highest probability if possible. 
  (concavity of \(f(x) = -x \log(x)\) )

- If color classes \(C_1\) and \(C_2\) can be merged properly,
  \(\Rightarrow C' = \{C_1 \cup C_2, C_3, \ldots, C_k\}\) has strictly lower entropy than the original coloring \(C\).

---

\(^1\)J. Cardinal, S. Fiorini, and G. Joret, "Minimum entropy coloring," in Algorithms and Computation
AND power graph vs OR power graph

**AND-product** graph of $G_1(X_1, E_1)$ and $G_2(X_2, E_2)$ $\Rightarrow G_1 \wedge G_2$

$\Rightarrow$ has vertex set $X_1 \times X_2$

$\Rightarrow (x_1, x_2)$ and $(x'_1, x'_2)$ are connected if \{either $x_i = x'_i$ or $x_i$ is connected to $x'_i$ in graph $G_i$\} for each $i = 1, 2$

**OR-product** graph of $G_1(X_1, E_1)$ and $G_2(X_2, E_2)$ $\Rightarrow G_1 \vee G_2$

$\Rightarrow$ has vertex set $X_1 \times X_2$

$\Rightarrow (x_1, x_2)$ and $(x'_1, x'_2)$ are connected if either $x_1$ is adjacent to $x'_1$ or $x'_2$ is adjacent to $x_2'$ in graphs $G_1$ and $G_2$, respectively.

$\Rightarrow \overline{G_1 \wedge G_2} = \overline{G_1} \vee \overline{G_2} \Rightarrow \overline{G}^\wedge n = \overline{G}^n$
AND product graph vs OR product graph

\[ G \text{ original graph} \]
\[ 2 - \text{AND Power graph} \]
\[ 2 - \text{OR Power graph} \]
Outline

- Introduction

- Zero error source coding (Scalar vs. Block coding)
  - Fixed length coding
  - Variable length coding

- Function compression
  - Optimal rate
    - Achievable coding scheme

- Function compression in zero-error regime

- Compression of specific functions
**Optimal Rate:** Source coding, Function compression

<table>
<thead>
<tr>
<th></th>
<th>Source, side info</th>
<th>Source, side info</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fixed length</td>
<td>variable length</td>
<td></td>
</tr>
<tr>
<td>scalar code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero error</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>block code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero error</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>block code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\epsilon - error)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Zero error source coding

- In information theory: Increasing the number of repetitions decreases $p(e)$

- In some applications no error can be tolerated.

- Sometimes there are only few source instances available.

- The error prob. decreases as the no. of instances increases → Not useful
Zero error- Simple point-to-point communication - Scalar

- Every source instances, a random variable $X$ is generated according to $p(x)$ independently of all other instance.

- for a single instance, the smallest number of bits in the worst case is $\log |\mathcal{X}| \leq \lceil \log |\mathcal{X}| \rceil \leq \log |\mathcal{X}| + 1$

$$X^n = [x(1), x(2), \ldots, x(n)]$$
$$x \in \mathcal{X}$$
$$x \sim p(x)$$
For $n$ independent instances, the number of bits needed in the worst case is $n \log |\mathcal{X}| \leq \lceil n \log |\mathcal{X}| \rceil \leq n \log |\mathcal{X}| + 1$

The asymptotic per-instance

- In worst case is between $\log |\mathcal{X}|$ and $\log |\mathcal{X}| + 1/n$ bits

\[ X^n = [x(1), x(2), ..., x(n)] \]
\[ x \in \mathcal{X} \]
\[ x \sim p(x) \]
\[ p\{X^n \neq \hat{X}^n\} = 0 \]
Zero-error source coding with side information

- The receiver knows some possibly related rv. $Y$ jointly distributed with $X$
- Probability of error must be exactly zero.
- How many bits are needed in the worst case?
- Any advantages over block coding of any independent instances?

$X^n \rightarrow \text{Encoder} \rightarrow \hat{X}^n$

$Y^n \rightarrow \text{Decoder} \rightarrow \hat{X}^n$

$P(\hat{X}^n \neq X^n) = 0$
Zero-error source coding, Fixed length code

- Receiver wants to know $X$ without error. (No error is tolerated!)

- $\mathcal{X} = \{1, 2, ..., 5\}$ and $\mathcal{Y} = \{1, 2, ..., 5\}$

<table>
<thead>
<tr>
<th>$y$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
</tr>
</tbody>
</table>

$P(\hat{X}^n \neq X^n) = 0$
Confusability graph

- $x, x'$ are confusable if $\exists y \in Y$ such that $p(x, y)p(x', y) > 0$
- Given $y = 1$, then $x$ can be either 1 or 2 (1 and 2 are confusable).
- This confusability relation can be captured by the characteristic graph $G$.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
</tr>
</tbody>
</table>

$(x, x') \in E \iff$ if $x, x'$ are confusable
The smallest number of possible messages is $\chi(G)$.

In worst case the smallest number of bits needed is $\log(\chi(G))$.
Assign different codewords to the confusable vectors

$x^n = (x_1, \ldots, x_n)$ and $x'^n = (x'_1, \ldots, x'_n)$ are confusable if $\exists y^n \in \mathcal{Y}^n$ such that $p(x^n, y^n)p(x'^n, y^n) > 0$

Confusability of the vectors can be captured by the AND power graph. $G^{\wedge n} = (\mathcal{V}^n, E^n)$

$(x^n, x'^n) \in E^n$ if $(x_i, x'_i) \in E$ or $x_i = x'_i$ for all $i \in \{1, \ldots, n\}$

The number of bits needed in worst case is $\log \chi(G^{\wedge n})$

In general $\chi(G^{\wedge n}) \leq (\chi(G))^n \Rightarrow$ benefit of block encoding
The five individual pentagons represent the confusability relation for $x_2$, each for fixed $x_1$

- meta-pentagon represents the confusability relation for $x_1$ alone

- $\log_2 \chi(G^\wedge 2) = \log_2 5 < \log_2 \chi^2(G) = \log_2 9$

- Benefit from blocking can be obtained
Outline

- Introduction
- Zero error source coding (Scalar vs. Block coding)
  - Fixed length coding
  - Variable length coding
- Asymptotically zero probability of error
  - Optimal rate for function compression
  - Achievable coding scheme
- Compression of specific functions
Variable length coding, Simple point-to-Point

- **Scalar code**: the smallest expected number of bits needed is between $H(X)$ and $H(X) + 1$

- **Block encoding**: the number of bits needed on average is between $nH(X)$ and $nH(X) + 1$

- The asymptotic per-instance on average is between $H(X)$ and $H(X) + \frac{1}{n}$ bits

$$X^n = [x(1), x(2), \ldots, x(n)]$$

$$x \in \mathcal{X}$$

$$x \sim p(x)$$
Variable length coding, Side information

- Variable length codes assigns shorter codes to the most frequent symbols and vice versa.

- Alon and Orlitsky considered two family of codes.\(^2\)

- No single letter characterization exists for optimal variable length code ⇒ Complementary graph entropy

- For a subclass of variable length codes the optimal rate is ⇒ Graph entropy.

\[ P(\hat{X}^n \neq X^n) = 0 \]

---

Coding method for unrestricted inputs (UI)

- Encoding function $\phi : \mathcal{X} \rightarrow \{0, 1\}^*$
  
  \[
  \begin{align*}
  \text{for all } x, x' \in V & \implies \phi(x) \neq \phi(x') \\
  (x, x') \in E & \implies \phi(x) \neq \phi(x')
  \end{align*}
  \]

- The decoder learns $x$ only if $p(x, y) > 0$. o.w., the decoder may be wrong on the value of $x$. 

Diagram:
- Encoder
- Decoder
- Lookup table
- $x_1 x_2 x_3 x_4$
- $y_1 y_2 y_3 y_4$
- $\phi(x_1) = \phi(x_3)$
- $p(x_1, y_1) > 0 \implies x_1$ is detected
- $p(x_3, y_1) = 0$
Coding method for unrestricted inputs (UI)

- Encoding function \( \phi : \mathcal{X} \rightarrow \{0, 1\}^* \)

\[
\begin{align*}
\{ & \text{for all } x, x' \in V \rightarrow \phi(x) \neq_P \phi(x') \\
& (x, x') \in E \Rightarrow \phi(x) \neq \phi(x') \}
\end{align*}
\]

- UI code is a coloring of \( G \) followed by a prefix-free coding of the colors.

\[
\phi(x_1), \phi(x_2), \phi(x_3), \phi(x_4)
\]

\[
\phi(x_1) \phi(x_2) \phi(x_3) \phi(x_4)
\]

\[
\begin{align*}
\phi(x_1) &= \phi(x_3) \\
p(x_1, y_1) &> 0 \Rightarrow x_1 \text{ is detected} \\
p(x_3, y_1) &= 0
\end{align*}
\]
The expected number of bits transmitted is

\[ \ell(\psi) = \sum_{x \in V} P(x) |\psi(x)| \]

The minimum Rate of UI codes are:

\[ L(G) = \min \{ \ell(\psi) : \psi \text{ is an UI code} \} \]

call the codes attaining these minima the optimal code
$H_\chi(G, X) \leq L \leq H_\chi(G, X) + 1$

- **Upper bound**: Take a coloring of $G$ achieving $H_\chi(G, X)$.

  ⇒ There is a prefix free encoding of the colors whose expected length is at most $H_\chi(G, X) + 1$

  Data compression result ⇒ $H(X) \leq l_{p.f}(X) \leq H(X) + 1$

- **Lower bound**: Every UI protocol can be viewed as a coloring of $G$ and prefix free encoding of it’s colors.

  ⇒ The entropy of a coloring is at least $H_\chi(G, X)$

  ⇒ The expected length is at least $H_\chi(G, X)$
Decoder needs to learn \( x_i \) for exactly those \( i \)'s where \( p(x_i, y_i) > 0 \) and can be wrong about \( x_i \) for the other \( i \)'s. This may require more bits.

\( x^n, x'^n \Rightarrow \) need to be distinguished if \( \exists i \) such that \( p(x_i, y_i)p(x'_i, y_i) > 0 \)

\( (x^n, x'^n) \in E^n \) if \( (x_i, x'_i) \in E \) for some \( i \in \{1, \ldots, n\} \)

\( \Rightarrow \) This is the definition of OR-Power graph of \( G \)
Optimal rate for UI (variable length code)

Remember for one instance the optimal rate is

$$H_X(G, X) \leq L \leq H_X(G, X) + 1$$

for n-instances

$$H_X(G^\lor n, X^{(n)}) \leq L_n \leq H_X(G_n^\lor n, X^{(n)}) + 1$$

Asymptotic per instance$^3$:

$$L_{am} = \lim_{n \to \infty} \frac{1}{n} H_X(G^\lor n, X^{(n)}) = H_G(X)$$

Graph entropy

- Was originally defined by Körner (1973).

\[ H_G(X) = \min_{X \in W \in \Gamma(G)} I(W, X) \]

- \( \Gamma(G) \) is the set of independent sets of the confusability graph \( G \)

- \( W \) is an independent set of \( G \)

- The minimization can be restricted to \( W \) ranging over the independent sets.
<table>
<thead>
<tr>
<th></th>
<th>Source, side info fixed length</th>
<th>Source, side info variable length</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>scalar code</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zero error</strong></td>
<td>$\log \chi(G)$</td>
<td>$H_\chi(G, X)$</td>
<td>?</td>
</tr>
<tr>
<td><strong>block code</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zero error</strong></td>
<td>$\log \chi(G^{\wedge n})$</td>
<td>$H_G(X)$</td>
<td>?</td>
</tr>
<tr>
<td><strong>block code</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>$\epsilon$ – error</strong></td>
<td></td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>

$$H_G(X|Y) \leq H(X|Y) \leq H_G(X) \leq H_\chi(G, X) \leq H(X) \leq \log \chi(G)$$
Outline

- Introduction

- Zero error source coding (Scalar vs. Block coding)
  - Fixed Length coding
  - Variable Length coding

- Asymptotically zero error
  - Optimal rate for function compression
  - Achievable coding scheme

- Function compression in zero-error regime

- Compression of specific functions
Function computing

- Compute a function $f(X, Y)$ at the decoder, and we want this encoding to be as efficient as possible.

- $f(X, Y) = Z$ must be determined for a (single) block of multiple instances.

- Vanishing (bit) block probability of error is allowed.

- $H(f(X, Y) | Y) \leq \text{Rate} = R^* \leq H(X | Y)$

$$p(X, Y) \xrightarrow{\text{encoder}} X \xrightarrow{\text{decoder}} \hat{f}(X, Y)$$

\[
\phi : \mathcal{X}^n \rightarrow \{0, 1\}^k \quad \psi : \{0, 1\}^k \times \mathcal{Y}^n \rightarrow \mathcal{Z}^n
\]

$$p(\psi(\phi(x), y) \neq f(x, y)) < \epsilon$$
Example

- Satellite has 10 parameters to report to the base.
- Base needs to learn only one parameter (only the base knows which one).
- The lower bound gives $H(x_Y | Y) = 1$ bit and the upper bound gives $H(X | Y) = 10$ bits. $R^* = 10$

\[
Y \sim \text{Unif}(1 : 10)
\]

\[
X = (x_1, \ldots, x_{10})
\]
In the case where \( f(X, Y) = (X, Y) \), the well-established rate result is the Slepian-Wolf bound:

\[
R_1 > H(X|Y) \\
R_2 > H(Y|X) \\
R_1 + R_2 > H(X,Y)
\]

When \( Y \) is available at the decoder, the optimal rate for the transmission of \( X \) is the conditional entropy of \( X \) given \( Y \), \( H(X|Y) \).
To achieve this rate $H(X|Y)$, encoder use more efficient encoding by taking advantage of its correlation with $Y$.

Knowing that the decoder can use its knowledge of $Y$ to disambiguate the transmitted data.

This analysis so far has ignored the fact that we are computing $f(X, Y)$ at the decoder, rather than reporting the two values directly!
Can we do better than Slepian Wolf bound?

**Example:** $X$ and $Y$ are integers and $f = X + Y \pmod{4}$.

- Regardless of how large $X$ and $Y$ might be, just encode the last two bits of each.
- The SW bound only takes advantage of the correlation between the two rvs.
- SW doesn’t take advantage of the properties of the function.
- To analyze a given $f$, we construct $f$-confusability graph.
**f-Confusability graph**

- f-confusability graph of $X$ given $Y$, $f$, and $p(x, y)$ is a graph $G = (V, E)$

- $V_x = X$ and, $(x, x') \in E$ if $\exists y \in Y$, such that, both $p(x, y) > 0$ and $p(x', y) > 0$ and $f(x, y) \neq f(x', y)$.

**Negation**

- If two nodes are not confusable, for all possible values of $y$

$$ (x, x') \notin E \quad \text{either} \quad p(x, y) = 0 \quad \text{or} \quad p(x', y) = 0 $$

or

$$ f(x, y) = f(x', y) $$
Encoding scheme 1-instance

- $(x, x') \notin E$ can safely be given the same encoding
- Color the f-confusibility graph $G_f$ with minimum entropy
- Transmit the name of the color class that contains each value
- SW: for asymptotically zero-error, it suffices to send the conditional entropy $H(X|Y)$ bits.
- Use SW encoding on the distribution over the color classes.
- This rate is $H_{\chi}(G_f, X|Y) = \min_c \sum_{y \in Y} p(y) H(c(x)|Y = y)$
Encoding scheme n-instance

- How to define the notion of f-confusability for vector?

  \[ (x^n, x'^n) \text{ are f-confusable, if } \exists y^n \in \mathcal{Y}^n \text{ such that } \]
  \[ p(x^n, y^n)p(x'^n, y^n) > 0 \text{ and } f(x^n, y^n) \neq f(x'^n, y^n) \]

  \[ (x^n, x'^n) \in E^n \text{ if } (x_i, x'_i) \in E \text{ for some } i \in \{1, 2, \ldots, n\} \]

- This is the definition of OR power graph
For 1-instance

\[ H_X(G_f, X|Y) \]

Asymptotic per instance

\[
\lim_{n \to \infty} \frac{1}{n} H_X(G_f^\sqrt{n}, X^{(n)}|Y^{(n)}) = H_{G_f}(X|Y)
\]

- we can compute the function with vanishing probability of error by first coloring a sufficiently large power graph of the characteristic graph
- encode the colors using any code that achieves the entropy bound on the colored source
optimal rate for the zero distortion functional compression problem with side information equals $R^* = H_G(X|Y)$ \(^4\)

$$H_G(X|Y) = \min_{W - X - Y \notin X \in W \in \Gamma(G)} I(W, X|Y)$$

- where $W - X - Y$ indicates that $W, X, Y$ is a Markov chain.
- $W$ should not contain any information about $Y$ that is not available through $X$
- The minimization can be restricted to $W$ ranging over the maximal independent sets.

\(^4\)A. Orlitsky and J. Roche, "Coding for computing," Information Theory, IEEE Transactions
### Summary

<table>
<thead>
<tr>
<th></th>
<th>Source, side info</th>
<th>Source, side info</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fixed length</td>
<td>variable length</td>
<td></td>
</tr>
<tr>
<td>scalar code</td>
<td>log $\chi(G')$</td>
<td>$H_{\chi}(G, X)$</td>
<td>?</td>
</tr>
<tr>
<td>Zero error</td>
<td>log $\chi(G^n)$</td>
<td>$H_G(X)$</td>
<td>?</td>
</tr>
<tr>
<td>block code</td>
<td>log $\chi(G^n)$</td>
<td>$H_G(X)$</td>
<td>?</td>
</tr>
<tr>
<td>block code</td>
<td>$H(X</td>
<td>Y)$</td>
<td>$H_{G_f}(X</td>
</tr>
<tr>
<td>$\epsilon$ error</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$H_{G_f}(X|Y) \leq H(X|Y) \leq H_G(X) \leq H_{\chi}(G, X) \leq H(X) \leq \log \chi(G)$$
Outline

- Introduction
- Zero-error source coding (Scalar vs. Block coding)
  - Fixed Length coding
  - Variable Length coding
- Asymptotically zero-error
  - Optimal rate for function compression
  - Achievable coding scheme
- Function Computation in zero-error regime
- Compression of specific functions
Function Compression, Scalar code, Zero-error

- Compute a function $f(X, Y)$ at the decoder.
- $f(X, Y)$ must be determined for a (single) block instance.
- No error is allowed (zero-error regime)
- Fundamental limit of computing a function with exactly zero probability of error?
- Is motivated by Alon, and Orlitsky (source coding with side information)

```
p(X, Y) ---| X | encoder | decoder |
               | Y |
```

\[ f(X, Y) \]
Achievable scheme

- The protocol that guarantees the zero probability of error, and to compute $f(x, y)$ only when $p(x, y) > 0$

- $(x, x') \notin E_f$ can safely be given the same encoding
  - Either $p(x, y) = 0$ or $p(x', y) = 0$
  - $f(x, y) = f(x', y)$

- Color the $f$-confusibility graph $G_f$ with minimum entropy

- Transmit the name of the color class that contains each value

- Using a prefix free encoding on the distribution over the color classes yields a valid encoding scheme for the problem
Achievable coding scheme

Graph Coloring

ENCODER

DECODER

Lookup Table

$X$ $\rightarrow$ Graph Coloring $\rightarrow$ ENCODER $\rightarrow$ DECODER $\rightarrow$ Lookup Table $\rightarrow$ $f(X, Y)$

Huffman Code

$Y$
Theorem

Optimal rate for computing a function \( f(X, Y) \) with zero probability of error, using scalar variable length code is

\[
H_\chi(G_f, X) \leq R \leq H_\chi(G_f, X) + 1
\]

Proof.

- **Upper bound**: Take a coloring of \( G \) achieving \( H_\chi(G_f, X) \).

  \[\Rightarrow\] There is a prefix free encoding of the colors whose expected length is at most \( H_\chi(G_f, X) + 1 \)

- **Lower bound**: This protocol can be viewed as a coloring of \( G_f \) and prefix free encoding of it’s colors.

  \[\Rightarrow\] The entropy of a coloring is at least \( H_\chi(G_f, X) \)

  \[\Rightarrow\] The expected length is at least \( H_\chi(G_f, X) \)
To protect against loss of synchronization if the side information at the decoder is occasionally wrong:

For a sequence $x_1x_2,..x_n$, the function needs to be computed correctly only for those coordinates that $p(x_i, y_i) > 0$.

For other positions decoder may not be able to compute $f(x_i, y_i)$ correctly.

$(x^n, x'^n) \in E^{(n)}$ if $(x_i, x'_i) \in E$ for some $i \in \{1,..,n\}$

\[
(x^n, x'^n) \in E^{(n)} \Leftrightarrow \phi(x^n) \neq \phi(x'^n) \quad (1)
\]

\[
x^n, x'^n \in V^{(n)} \Leftrightarrow \phi(x^n) \neq \phi(x'^n) \quad (2)
\]
Optimal rate for function compression in zero-error regime

One instance

\[ H_\chi(G_f, X) \leq R \leq H_\chi(G_f, X) + 1 \]

n-instance

\[ H_\chi(G_f^\vee n, X^{(n)}) \leq R^n \leq H_\chi(G_f^\vee n, X^{(n)}) + 1 \]

Asymptotic per instance

\[ \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G_f^\vee n, X^{(n)}) = H_{G_f}(X) \]

- We can compute the function with zero-error by first coloring a sufficiently large power graph of the characteristic graph
- Encode the colors using Huffman codes
Proposition

In exactly zero-error regime, using block code, the optimal rate for computing \( f(X, Y) \) is less than recovering the source \( X \).

Proof.

- \( G_f = (V, E_f) \) and \( G = (V, E) \), and \( E_f \subseteq E \)

- Let \((X, W')\) be random variable achieving \( H_G(X) \)

- \( W' \) is an independent set containing \( X \), and \( G_f \subseteq G \Rightarrow \) it’s also an independent set of \( G_f \).

\[
H_{G_f}(X) = \min_{X \in W \in \Gamma(G_f)} I(X; W) \leq I(X; W') = \min_{X \in W \in \Gamma(G)} I(X; W) = H_G(X)
\]
Proposition

*In exactly zero-error regime, using scalar code, the optimal rate for computing $f(X, Y)$ is less than recovering the source $X$.\*

- $G_f \subseteq G$
- Any coloring of $G$ is a coloring of $G_f$
- The coloring with minimum entropy of $G$ is also a coloring of $G_f$.

$$H_{\chi}(G_f, X) \leq H(c'(G_f)) = H_{\chi}(G, x) \leq H(c(G))$$
Theorem
We can get saving over the block encoding in zero-error function compression regime.

\[ H_{G_f}(X) \leq H_\chi(G_f, X) \]

Proof.
If \( c \) is a coloring of \( G \), then \( c^{-1}(c(X)) \), the color class of \( X \), induced a partition of \( V \) into independent sets of \( G \), and always contains \( X \).

\[ H_\chi(G_f, X) = \min \{ H(c(X) | c^{-1}(c(X)) \text{ are the color classes} \} \]
\[ \overset{a}{=} \min_{X \in W \in \Gamma(G_f)} H(W) | \text{disjoint union of } W\text{'s are the color classes} \]
\[ \overset{b}{=} \min_{X \in W \in \Gamma(G_f)} \{ I(W; X) | \text{disjoint union of } W\text{'s are the color classes} \} \]
\[ \overset{c}{=} \min_{X \in W \in \Gamma(G_f)} \{ I(W; X) \} = H_{G_f}(X) \]
Summary

<table>
<thead>
<tr>
<th>Source, side info</th>
<th>Source, side info</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scalar code</td>
<td>log $\chi(G')$</td>
<td>$H_\chi(G, X)$</td>
</tr>
<tr>
<td>Zero error</td>
<td></td>
<td>$H_\chi(G_f, X)$</td>
</tr>
<tr>
<td></td>
<td>log $\chi(G'^n)$</td>
<td>$H_G(X)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$H_{G_f}(X)$</td>
</tr>
<tr>
<td>block code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>block code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\epsilon - error$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$H(X</td>
<td>Y)$</td>
</tr>
</tbody>
</table>

$$H_G(X|Y) \leq H(X|Y) \leq H_G(X) \leq H_\chi(G, X) \leq H(X) \leq \log \chi(G)$$
Outline

► Introduction

► Zero-error source coding (Scalar vs. Block coding)
  ⇒ Fixed Length coding
  ⇒ Variable Length coding

► Asymptotically zero-error
  ⇒ Optimal rate for function compression
  ⇒ Achievable coding scheme

► Function compression zero-error regime

► Compression of specific functions
Compression of specific functions

▶ In environmental monitoring, a relevant statistic of temperature sensor readings \( x_1, x_2, \ldots, x_n \) may be the mean temperature, or median, or mode.

▶ In “alarm” networks, the quantity of interest might be the maximum, \( \max_{1 \leq i \leq n}(x_i) \) of \( n \) temperature readings.

▶ Frequency histogram function counts the number of occurrences of each argument among measurements. It yields many good results such as mean, variance, maximum, minimum, median and other important statistics.

▶ Function compression on specific functions by graph coloring which provide benefit over compressing each source separately.
We have $K$ users which has access to independent sequences.

Each node (user) periodically makes measurements, which belongs to a fixed finite set $\mathcal{X} = \{1, 2, ..., |\mathcal{X}|\}$. 

$$f : \mathcal{X}^K \rightarrow \{0, ..., r - 1\}$$

![Diagram](image.png)
The characteristic graph of the \(i\)-th user

**Theorem**

\((p, q) \notin E_i \ (p, q \in \mathcal{X}), \text{ if and only if, } p \equiv q \mod r\)

**Proof.**

\(\Rightarrow: \) Since \((p, q) \notin E_i\), we obtain that

\[
\text{for all } x_i \setminus i \in \mathcal{X}^{N-1}, \quad f(X) = f(p, x_i \setminus i) = f(q, x_i \setminus i).
\]

\[
p + \sum_{j \neq i} x_j \equiv q + \sum_{j \neq i} x_j \mod r \Rightarrow p \equiv q \mod r.
\]

\(\Leftarrow: \) If \(p \equiv q \mod r\), and \(\sum_{j \neq i} x_j \equiv \sum_{j \neq i} x_j \mod r\), then for all \(x_i \setminus i \in \mathcal{X}^{N-1}\) we have

\[
p + \sum_{j \neq i} x_j \equiv q + \sum_{j \neq i} x_j \mod r.
\]

Hence, \(p, q \notin E_i\) \(\square\)
The characteristic graph of the $i$-th user

- Note that maximal independent sets form a partition.
  - $W_0 = \{x \in \mathcal{X} | x = 0 \text{ (mod } r)\}$
  - $W_1 = \{x \in \mathcal{X} | x = 1 \text{ (mod } r)\}$
  - ...
  - $W_{r-1} = \{x \in \mathcal{X} | x = r - 1 \text{ (mod } r)\}$

$\Rightarrow H_\chi(G, X)$ can be achieved by assigning different colors to different maximal independent sets.

$\Rightarrow$ The minimum entropy coloring can be achieved by optimal coloring $\chi(G)$
Optimal rate for Mod $r$ sum

\[ H_G(X|Y) = \min I(W, X|Y) \]
\[ = \min \{ H(W|Y) - H(W|XY) \} \]
\[ = \min \{ H(W|Y) - H(W|X) \} \]
\[ = H(W|Y) \]
\[ = H(W) = -(P_0 \log P_0 + \ldots + P_{r-1} \log P_{r-1}) \]

- where $P_i = \sum_{x_i \in W_i} p(x_i)$

- If the distribution over the vertices is uniform, and $|\mathcal{X}| = Cr$, and $C \in \mathbb{Z}^+$, the total benefit is

\[ H(X|Y) - H_G(X|Y) = \log |\mathcal{X}| - \log r = \log C \geq 0 \]
Summary

- Source coding problem with side information in zero-error regime is discussed.
- The optimal rate for function compression in zero-error regime is derived.
- Function computation in asymptoticly zero error regime is discussed.
- The optimal rate for computing mod $r$ sum is derived.
Thank you

Questions?