Network coding for multicast relation to compression and generalization of Slepian-Wolf

Overview

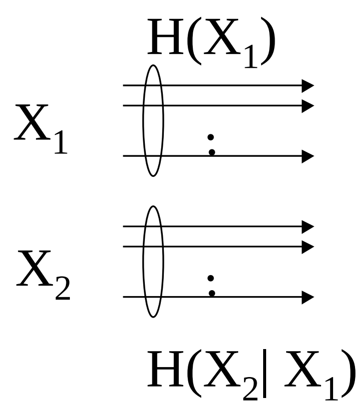
- Review of Slepian-Wolf
- Distributed network compression
- Error exponents Source-channel separation issues
- Code construction for finite field multiple access networks

Distributed data compression

Consider two correlated sources $(X, Y) \sim p(x, y)$ that must be separately encoded for a user who wants to reconstruct both

What information transmission rates from each source allow decoding with arbitrarily small probability of error?

E.g.



Distributed source code

A $((2^{nR_1}, 2^{nR_2}), n)$ distributed source code for joint source (X, Y) consists of encoder maps

$$f_1 : \mathcal{X}^n \rightarrow \{1, 2, \ldots, 2^{nR_1}\}$$
$$f_2 : \mathcal{Y}^n \rightarrow \{1, 2, \ldots, 2^{nR_2}\}$$

and a decoder map

$$g : \{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{nR_2}\} \rightarrow \mathcal{X}^n \times \mathcal{Y}^n$$

- X^n is mapped to $f_1(X^n)$
- Y^n is mapped to $f_2(Y^n)$
- (R_1, R_2) is the rate pair of the code

Probability of error

$$P_e^{(n)} = \mathsf{Pr}\{g(f_1(X^n), f_2(Y^n)) \neq (X^n, Y^n)\}$$

Slepian-Wolf

Definitions:

A rate pair (R_1, R_2) is *achievable* if there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ distributed source codes with probability of error $P_e^{(n)} \to 0$ as $n \to \infty$

achievable rate region - closure of the set of achievable rates

Slepian-Wolf Theorem:

For the distributed source coding problem for source (X, Y) drawn i.i.d. $\sim p(x, y)$, the achievable rate region is

$$\begin{aligned} R_1 &\geq H(X|Y) \\ R_2 &\geq H(Y|X) \\ R_1 + R_2 &\geq H(X, Y) \end{aligned}$$

Proof of achievability

Main idea: show that if the rate pair is in the Slepian-Wolf region, we can use a random binning encoding scheme with typical set decoding to obtain a probability of error that tends to zero

Coding scheme:

- Source X assigns every sourceword $x \in \mathcal{X}^n$ randomly among 2^{nR_1} bins, and source Y independently assigns every $y \in \mathcal{Y}^n$ randomly among 2^{nR_2} bins

- Each sends the bin index corresponding to the message

- the receiver decodes correctly if there is exactly one jointly typical sourceword pair corresponding to the received bin indexes, otherwise it declares an error

Random binning for single source compression

An encoder that knows the typical set can compress a source X to $H(X)+\epsilon$ without loss, by employing separate codes for typical and atypical sequences

Random binning is a way to compress a source X to $H(X)+\epsilon$ with asymptotically small probability of error without the encoder knowing the typical set, as well as the decoder knows the typical set

- the encoder maps each source sequence X^n uniformly at random into one of 2^{nR} bins

- the bin index, which is R bits long, forms the code

- the receiver decodes correctly if there is exactly one typical sequence corresponding to the received bin index

Error analysis

An error occurs if:

a) the transmitted sourceword is not typical, i.e. event

$$E_0 = \{\mathbf{X} \notin A_\epsilon^{(n)}\}$$

b) there exists another typical sourceword in the same bin, i.e.event

$$E_1 = \{\exists \mathbf{x}' \neq \mathbf{X} : f(\mathbf{x}') = f(\mathbf{X}), \mathbf{x}' \in A_\epsilon^{(n)}\}$$

Use union of events bound:

$$\begin{aligned} P_e^{(n)} &= \Pr(E_0 \cup E_1) \\ &\leq \Pr(E_0) + \Pr(E_1) \end{aligned}$$

Error analysis continued

$\Pr(E_0) \to 0$ by the Asymptotic Equipartition Property (AEP)

$$
\begin{aligned}
\Pr(E_1) &= \sum_{\mathbf{x}} \Pr\{\exists \mathbf{x}' \neq \mathbf{x} : f(\mathbf{x}') = f(\mathbf{x}), \\
&\qquad\qquad \mathbf{x}' \in A_\epsilon^{(n)}\} \\
&\leq \sum_{\mathbf{x}} \sum_{\substack{\mathbf{x}' \neq \mathbf{x} \\ \mathbf{x}' \in A_\epsilon^{(n)}}} \Pr(f(\mathbf{x}') = f(\mathbf{x})) \\
&= \sum_{\mathbf{x}} |A_\epsilon^{(n)}| \, 2^{-nR} \\
&\leq 2^{-nR} \, 2^{n(H(X)+\epsilon)} \\
&\to 0 \text{ if } R > H(X)
\end{aligned}
$$

For sufficiently large n,

$$\begin{aligned} & \Pr(E_0), \Pr(E_1) < \epsilon \\ \Rightarrow \; & P_\epsilon^{(n)} < 2\epsilon \end{aligned}$$

Jointly typical sequences

The set $A_\epsilon^{(n)}$ of jointly typical sequences is the set of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with probability:

$$2^{-n(H(X)+\epsilon)} \leq p_{\mathbf{X}}(\mathbf{x}) \leq 2^{-n(H(X)-\epsilon)}$$

$$2^{-n(H(Y)+\epsilon)} \leq p_{\mathbf{Y}}(\mathbf{y}) \leq 2^{-n(H(Y)-\epsilon)}$$

$$2^{-n(H(X,Y)+\epsilon)} \leq p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq 2^{-n(H(X,Y)-\epsilon)}$$

for $(\mathbf{X}, \mathbf{Y})$ sequences of length n IID according to $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^n p_{X,Y}(x_i, y_i)$

Size of typical set:

$$|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$$

Proof:

$$\begin{aligned} 1 &= \sum p(\mathbf{x}, \mathbf{y}) \\ &\geq \sum_{A_\epsilon^{(n)}} p(\mathbf{x}, \mathbf{y}) \\ &\geq |A_\epsilon^{(n)}| 2^{-n(H(X,Y)+\epsilon)} \end{aligned}$$

Conditionally typical sequences

The conditionally typical set $A_\epsilon^{(n)}(X|\mathbf{y})$ for a given typical $\mathbf{y}$ sequence is the set of $\mathbf{x}$ sequences that are jointly typical with the given $\mathbf{y}$ sequence.

Size of conditionally typical set:

$$|A_\epsilon^{(n)}(X|\mathbf{y})| \leq 2^{n(H(X|Y)+\epsilon)}$$

Proof:

For $(\mathbf{x}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y)$,

$$\begin{aligned} p(\mathbf{y}) &\doteq 2^{-n(H(Y)\pm\epsilon)} \\ p(\mathbf{x}, \mathbf{y}) &\doteq 2^{-n(H(X,Y)\pm\epsilon)} \\ \Rightarrow p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \\ &\doteq 2^{-n(H(X|Y)\pm 2\epsilon)} \end{aligned}$$

Hence

$$\begin{aligned} 1 &\geq \sum_{\mathbf{x} \in A_\epsilon^{(n)}(X|\mathbf{y})} p(\mathbf{x}|\mathbf{y}) \\ &\geq |A_\epsilon^{(n)}| 2^{-n(H(X|Y)+2\epsilon)} \end{aligned}$$

Proof of achievability – error analysis

Errors occur if:

a) the transmitted sourcewords are not jointly typical, i.e. event

$$E_0 = \{(X, Y) \notin A_\epsilon^{(n)}\}$$

b) there exists another pair of jointly typical sourcewords in the same pair of bins, i.e. one or more of the following events

$$\begin{aligned} E_1 &= \{\exists \mathbf{x}' \neq \mathbf{X} : f_1(\mathbf{x}') = f_1(\mathbf{X}), (\mathbf{x}', \mathbf{Y}) \in A_\epsilon^{(n)}\} \\ E_2 &= \{\exists \mathbf{y}' \neq \mathbf{Y} : f_2(\mathbf{y}') = f_2(\mathbf{Y}), (\mathbf{X}, \mathbf{y}') \in A_\epsilon^{(n)}\} \\ E_{12} &= \{\exists (\mathbf{x}', \mathbf{y}') : \mathbf{x}' \neq \mathbf{X}, \mathbf{y}' \neq \mathbf{Y}, f_1(\mathbf{x}') = f_1(\mathbf{X}), \\ &\quad f_2(\mathbf{y}') = f_2(\mathbf{Y}), (\mathbf{x}', \mathbf{y}') \in A_\epsilon^{(n)}\} \end{aligned}$$

Use union of events bound:

$$
\begin{aligned}
P_e^{(n)} &= \Pr(E_0 \cup E_1 \cup E_2 \cup E_{12}) \\
&\leq \Pr(E_0) + \Pr(E_1) + \Pr(E_2) + \Pr(E_{12})
\end{aligned}
$$

Error analysis continued

$\Pr(E_0) \to 0$ by the AEP

$$\begin{aligned}
\Pr(E_1) &= \sum_{(\mathbf{x},\mathbf{y})} \Pr\{\exists \mathbf{x}' \neq \mathbf{x} : f_1(\mathbf{x}') = f_1(\mathbf{x}), \\
&\qquad\qquad (\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}\} \\
&\le \sum_{(\mathbf{x},\mathbf{y})} \sum_{\substack{\mathbf{x}' \neq \mathbf{x} \\ (\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}}} \Pr(f_1(\mathbf{x}') = f_1(\mathbf{x})) \\
&= \sum_{(\mathbf{x},\mathbf{y})} |A_\epsilon^{(n)}(X|\mathbf{y})| \, 2^{-nR_1} \\
&\le 2^{-nR_1} \, 2^{n(H(X|Y)+2\epsilon)} \\
&\to 0 \text{ if } R_1 > H(X|Y)
\end{aligned}$$

Similarly,

$$
\begin{aligned}
\Pr(E_2) &\le 2^{-nR_2}\, 2^{n(H(Y|X)+2\epsilon)} \\
&\to 0 \text{ if } R_2 > H(Y|X) \\
\Pr(E_{12}) &\le 2^{-n(R_1+R_2)}\, 2^{n(H(X,Y)+\epsilon)} \\
&\to 0 \text{ if } R_1 + R_2 > H(X,Y)
\end{aligned}
$$

Error analysis continued

Thus, if we are in the Slepian-Wolf rate region, for sufficiently large n,

$$\begin{aligned} & \Pr(E_0), \Pr(E_1), \Pr(E_2), \Pr(E_{12}) < \epsilon \\ \Rightarrow \ & P_\epsilon^{(n)} < 4\epsilon \end{aligned}$$

Since the average probability of error is less than 4ϵ, there exist at least one code (f_1^*, f_2^*, g^*) with probability of error $< 4\epsilon$.

Thus, there exists a sequence of codes with $P_\epsilon^{(n)} \to 0$.

Model for distributed network compression

- arbitrary directed graph with integer capacity links
- discrete memoryless source processes with integer bit rates
- randomized linear network coding over vectors of bits in $\mathbb{F}_2$
- coefficients of overall combination transmitted to receivers
- receivers perform minimum entropy or maximum a posteriori probability decoding

Distributed compression problem

Consider

- two sources of bit rates r_1, r_2, whose output values in each unit time period are drawn i.i.d. from the same joint distribution Q

- linear network coding in $\mathbb{F}_2$ over vectors of nr_1 and nr_2 bits from each source respectively

Define

- m_1 and m_2 the minimum cut capacities between the receiver and each source respectively

- m_3 the minimum cut capacity between the receiver and both sources

- L the maximum source-receiver path length

Theorem 1 *The error probability at each receiver using minimum entropy or maximum a posteriori probability decoding is at most* $\sum_{i=1}^{3} p_e^i$*, where*

$$\begin{aligned}
p_e^1 &\leq \exp\Bigg\{ -n \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_1(1 - \frac{1}{n}\log L) - H(X_1|X_2) \right|^+ \Bigg) + 2^{2r_1+r_2}\log(n+1) \Bigg\} \\
p_e^2 &\leq \exp\Bigg\{ -n \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_2(1 - \frac{1}{n}\log L) - H(X_2|X_1) \right|^+ \Bigg) + 2^{r_1+2r_2}\log(n+1) \Bigg\} \\
p_e^3 &\leq \exp\Bigg\{ -n \min_{X_1,X_2} \Bigg(D(P_{X_1X_2}||Q) \\
&\quad + \left| m_3(1 - \frac{1}{n}\log L) - H(X_1X_2) \right|^+ \Bigg) + 2^{2r_1+2r_2}\log(n+1) \Bigg\}
\end{aligned}$$

Distributed compression

- Redundancy is removed or added in different parts of the network depending on available capacity

- Achieved without knowledge of source entropy rates at interior network nodes

- For the special case of a Slepian-Wolf source network consisting of a link from each source to the receiver, the network coding error exponents reduce to known error exponents for linear Slepian-Wolf coding [Csi82]

Proof outline

- Error probability $\leq \sum_{i=1}^{3} p_e^i$, where
  - p_e^1 is the probability of correctly decoding X_2 but not X_1,
  - p_e^2 is the probability of correctly decoding X_1 but not X_2
  - p_e^3 is the probability of wrongly decoding X_1, X_2
- Proof approach using method of types similar to that in [Csi82]
- Types $P_{\mathbf{x}_i}$, joint types $P_{\mathbf{xy}}$ are the empirical distributions of elements in vectors $\mathbf{x}_i$

Proof outline (cont'd)

Bound error probabilities by summing over

- sets of joint types

$$\mathcal{P}_n^i = \begin{cases} \{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \mid \tilde{X}_1 \neq X_1, \tilde{X}_2 = X_2\} & i = 1 \\ \{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \mid \tilde{X}_1 = X_1, \tilde{X}_2 \neq X_2\} & i = 2 \\ \{P_{X_1\tilde{X}_1X_2\tilde{X}_2} \mid \tilde{X}_1 \neq X_1, \tilde{X}_2 \neq X_2\} & i = 3 \end{cases}$$

where $X_i, \tilde{X}_i \in \mathbb{F}_2^{nr_i}$

- sequences of each type

$$\mathcal{T}_{X_1X_2} = \left\{ [\ \mathbf{x}\quad \mathbf{y}\ ] \in \mathbb{F}_2^{n(r_1+r_2)} \;\middle|\; P_{\mathbf{xy}} = P_{X_1X_2} \right\}$$

$$\mathcal{T}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{xy}) = \Big\{ [\ \tilde{\mathbf{x}}\quad \tilde{\mathbf{y}}\ ] \in \mathbb{F}_2^{n(r_1+r_2)} \;\Big| \\ P_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}\mathbf{xy}} = P_{\tilde{X}_1\tilde{X}_2X_1X_2} \Big\}$$

Proof outline (cont'd)

- Define
  - $P_i, i = 1, 2$, the probability that distinct $(\mathbf{x}, \mathbf{y}), (\tilde{\mathbf{x}}, \mathbf{y})$, where $\mathbf{x} \neq \tilde{\mathbf{x}}$, at the receiver
  - P_3, the probability that $(\mathbf{x}, \mathbf{y}), (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, where $\mathbf{x} \neq \tilde{\mathbf{x}}$,$\mathbf{y} \neq \tilde{\mathbf{y}}$, are mapped to the same output at the receiver
- These probabilities can be calculated for a given network, or bounded in terms of block length n and network parameters

Proof outline (cont'd)

- A link with ≥ 1 nonzero incoming signal carries the zero signal with probability $\frac{1}{2^{nc}}$, where c is the link capacity

- this is equal to the probability that a pair of distinct input values are mapped to the same output on the link

- We can show by induction on the minimum cut capacities m_i that

$$\begin{aligned} P_i &\leq \left(1-(1-\frac{1}{2^n})^L\right)^{m_i} \\ &\leq \left(\frac{L}{2^n}\right)^{m_i} \end{aligned}$$

Proof outline (cont'd)

We substitute in

- cardinality bounds

$$
\begin{aligned}
|\mathcal{P}_n^1| &< (n+1)^{2^{2r_1+r_2}} \\
|\mathcal{P}_n^2| &< (n+1)^{2^{r_1+2r_2}} \\
|\mathcal{P}_n^3| &< (n+1)^{2^{2r_1+2r_2}} \\
|\mathcal{T}_{X_1X_2}| &\leq \exp\{nH(X_1X_2)\} \\
|\mathcal{T}_{\tilde{X}_1\tilde{X}_2|X_1X_2}(\mathbf{xy})| &\leq \exp\{nH(\tilde{X}_1\tilde{X}_2|X_1X_2)\}
\end{aligned}
$$

- probability of source vector of type $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{X_1X_2}$

$$
Q^n(\mathbf{xy}) = \exp\{-n(D(P_{X_1X_2}||Q) + H(X_1X_2))\}
$$

Proof outline (cont'd)

and the decoding conditions

- minimum entropy decoder:

$$H(\tilde{X}_1\tilde{X}_2) \leq H(X_1X_2)$$

- maximum a posteriori probability decoder:

$$D(P_{\tilde{X}_1\tilde{X}_2}||Q) + H(\tilde{X}_1\tilde{X}_2) \leq D(P_{X_1X_2}||Q) + H(X_1X_2)$$

to obtain the result

Conclusions

- Distributed randomized network coding can achieve distributed compression of correlated sources

- Error exponents generalize results for linear Slepian Wolf coding

- Further work: investigation of non-uniform code distributions, other types of codes, and other decoding schemes

6.441 Information Theory
Spring 2010