Strong Converse Theorems for Classes of Multimessage Multicast Networks: A Rényi Divergence Approach

Silas Fong

(Joint work with Vincent Tan)

Department of Electrical & Computer Engineering
National University of Singapore

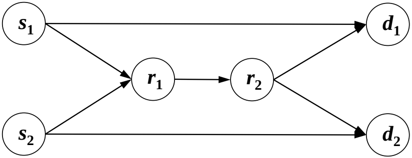
{*silas_fong, vtan*}*@nus.edu.sg*

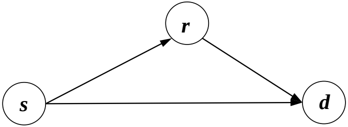
January 21, 2015

Overview

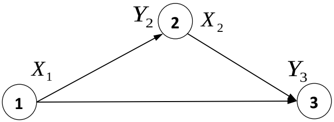
Multimessage Multicast Networks (MMNs)

- Multiple sources transmit messages to multiple destinations.
- Each source transmits 1 message.
- Each destination decodes all the messages.
- Examples include the butterfly network and the relay channel.





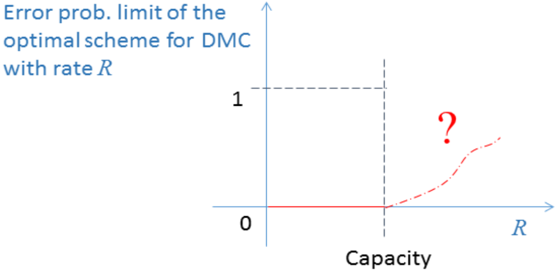
Some MMNs with Known Capacity Regions



- Finite-field linear deterministic network [Avestimehr et al., 2011]
  - $Y_3 = X_1 + X_2 \in \mathrm{GF}(p^n)$ for some finite field $\mathrm{GF}(p^n)$
- MMN consisting of independent DMCs [Köetter et al., 2011]
  - $Y_3 = (X_1 + Z_1, X_2 + Z_2)$ where Z_1 and Z_2 are independent noises.
  - The linear network coding model is a special case when $Z_1 = Z_2 = 0$ [Li et al., 2003].
- Wireless erasure network [Dana et al., 2006]
  - $Y_3 = (\hat{X}_1, \hat{X}_2)$ where $\hat{X}_i = \begin{cases} \text{erasure} & \text{with prob. } \varepsilon_i \\ X_i & \text{with prob. } 1 - \varepsilon_i. \end{cases}$
- The direct part uses random coding. The converse part uses Fano's inequality, which yields a *weak converse*.

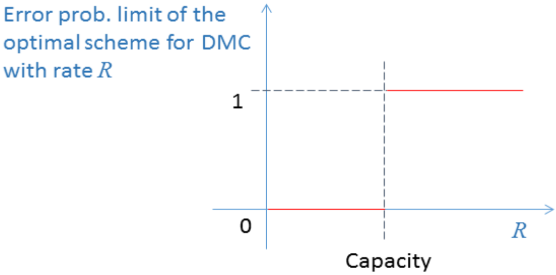
Weak Converse

- Assumption: The probability of decoding error vanishes as the blocklength increases.
- If the rate tuple of a code falls outside the capacity region, the probability of decoding error must be bounded away from 0 as the blocklength increases.
- For the DMC example, consider any sequence of the optimal length-n schemes with rate R which minimize the error probability:



Strong Converse

- Weaker assumption: The asymptotic probability of decoding error is upper bounded by some $\varepsilon \in [0, 1)$ as the blocklength increases.
- If the rate tuple of a code falls outside the capacity region, the probability of decoding error must tend to 1 as the blocklength increases.



Motivation of This Work

- Although the capacity regions of the aforementioned DM-MMNs are well-known, only the *weak converse* has been proved using *Fano's inequality*.
- Therefore we are motivated to prove the *strong converse* using *Rényi divergence*:
  - Powerful technique for establishing strong converse theorems.
  - Has been employed previously to establish strong converse for DMC with output feedback [Polyanskiy and Verdú, 2010], classical-quantum channels [Ogawa and Nagaoka, 1999], and entanglement-breaking quantum channels [Wilde et al., 2013].

Network Model

- Let $\mathcal{I} \triangleq \{1, 2, \ldots, N\}$ be the index set of the nodes.
- Let $\mathcal{S} \subseteq \mathcal{I}$ and $\mathcal{D} \subseteq \mathcal{I}$ denote the set of source and destination nodes respectively.
- The sources in $\mathcal{S}$ transmit information to the destinations in $\mathcal{D}$ in n time slots:
  - Each source $i \in \mathcal{S}$ chooses W_i to transmit. Message W_i is uniform on $\{1, 2, \ldots, 2^{nR_i}\}$ where R_i denotes the rate of W_i.
  - Each destination $j \in \mathcal{D}$ wants to decode $W_{\mathcal{S}}$.
- Each node i transmits $X_{i,k}$ and receives $Y_{i,k}$ in time slot k.
- $X_{i,k}$ is a function of (W_i, Y_i^{k-1}).
- The channel is characterized by $q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}$: For each $k \in \{1, 2, \ldots, n\}$,

$$\Pr\{Y_{\mathcal{I},k} = y_{\mathcal{I},k} | W_{\mathcal{I}} = w_{\mathcal{I}}, X_{\mathcal{I}}^k = x_{\mathcal{I}}^k, Y_{\mathcal{I}}^{k-1} = y_{\mathcal{I}}^{k-1}\} = q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}(y_{\mathcal{I},k}|x_{\mathcal{I},k})$$

ε-Capacity Region

- Define $R_{\mathcal{I}} \triangleq (R_1, R_2, \ldots, R_N)$ to be a rate tuple.
  - Assume wlog $R_i = 0 \ \forall i \notin \mathcal{S}$.
- A length-n code operating at rate $R_{\mathcal{I}}$ is called an $(n, R_{\mathcal{I}}, \varepsilon_n)$-*code* if the average probability of decoding error is ε_n.
- $R_{\mathcal{I}}$ is ε-achievable if $\exists$ a sequence of $(n, R_{\mathcal{I}}, \varepsilon_n)$-codes such that $\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon$.
- ε-capacity region $\mathcal{C}_\varepsilon \triangleq \{R_{\mathcal{I}} : R_{\mathcal{I}} \text{ is } \varepsilon\text{-achievable}\}$
- Fano's inequality yields an outer bound for only $\mathcal{C}_0$

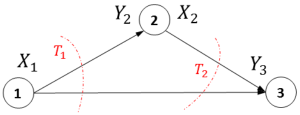
Cut-Set Outer Bound

- A well-known outer bound on the capacity region of DM-MMN developed by El Gamal in 1981.

$$\mathcal{C}_0 \subseteq \bigcup_{p_{X_\mathcal{I}}} \bigcap_{\substack{T \subseteq \mathcal{I}: \\ T^c \cap \mathcal{D} \neq \emptyset}} \left\{ R_\mathcal{I} \,\middle|\, \sum_{i \in T} R_i \leq I_{p_{X_\mathcal{I}} q_{Y_\mathcal{I}|X_\mathcal{I}}}(X_T; Y_{T^c} | X_{T^c}) \right\}$$

- Applying it to the relay channel, we have

$$R \leq \max_{p_{X_1, X_2}} \min \{ \underbrace{I(X_1; Y_2, Y_3 | X_2)}_{\text{cut } T_1}, \underbrace{I(X_1, X_2; Y_3)}_{\text{cut } T_2} \}$$



Simplified Noisy Network Coding (NNC) Inner Bound

- Simplified noisy network coding (NNC) inner bound [Lim et al., 2011]:

$$\mathcal{C}_0 \supseteq \bigcup_{\substack{p_{X_\mathcal{I}}: p_{X_\mathcal{I}} \\ = \prod_{i=1}^N p_{X_i}}} \bigcap_{\substack{T \subseteq \mathcal{I}: \\ T^c \cap \mathcal{D} \neq \emptyset}} \left\{ R_\mathcal{I} \,\middle|\, \sum_{i \in T} R_i \leq I(X_T; Y_{T^c} | X_{T^c}) - H(Y_T | X_\mathcal{I}, Y_{T^c}) \right\}$$

- Similar to the cut-set bound:

$$\mathcal{C}_0 \subseteq \bigcup_{p_{X_\mathcal{I}}} \bigcap_{\substack{T \subseteq \mathcal{I}: \\ T^c \cap \mathcal{D} \neq \emptyset}} \left\{ R_\mathcal{I} \,\middle|\, \sum_{i \in T} R_i \leq I_{p_{X_\mathcal{I}} q_{Y_\mathcal{I}|X_\mathcal{I}}}(X_T; Y_{T^c} | X_{T^c}) \right\}$$

- For the finite-field linear deterministic network, MMN consisting of independent channels and the wireless erasure network,

NNC inner bound = cut-set bound

Main Theorem

Theorem (Strong Converse Outer Bound)

For each $\varepsilon \in [0,1)$,

$$\mathcal{C}_\varepsilon \subseteq \bigcap_{\substack{T\subseteq\mathcal{I}:\\ T^c\cap\mathcal{D}\neq\emptyset}} \bigcup_{p_{X_\mathcal{I}}} \left\{ R_\mathcal{I} \,\middle|\, \sum_{i\in T} R_i \le I_{p_{X_\mathcal{I}} q_{Y_\mathcal{I}|X_\mathcal{I}}}(X_T; Y_{T^c}|X_{T^c}) \right\}$$

- Compare with cut-set outer bound:

$$\mathcal{C}_0 \subseteq \bigcup_{p_{X_\mathcal{I}}} \bigcap_{\substack{T\subseteq\mathcal{I}:\\ T^c\cap\mathcal{D}\neq\emptyset}} \left\{ R_\mathcal{I} \,\middle|\, \sum_{i\in T} R_i \le I_{p_{X_\mathcal{I}} q_{Y_\mathcal{I}|X_\mathcal{I}}}(X_T; Y_{T^c}|X_{T^c}) \right\}$$

- Reason for the gap:
  - Both proofs first fix an arbitrary T and then find a distribution $p_{X_\mathcal{I}}^{(T)}$ such that $\sum_{i\in T} R_i \le I_{p_{X_\mathcal{I}}^{(T)} q_{Y_\mathcal{I}|X_\mathcal{I}}}(X_T; Y_{T^c}|X_{T^c})$ for the ε-achievable $R_\mathcal{I}$.
  - In the proof of cut-set bound, $p_{X_\mathcal{I}}^{(T)}$ are the same for all T.
  - In the proof of strong converse bound, $p_{X_\mathcal{I}}^{(T)}$ can be different for different T.

Strong Converse for Classes of MMNs

Proposition

For the finite-field linear deterministic network, MMN consisting of independent channels and the wireless erasure network,

$$\text{strong converse bound} = \text{cut-set bound} \overset{\text{known}}{=} \text{NNC inner bound}.$$

Corollary (Strong converse)

Consider a network belonging to one of the above three classes. For each $\varepsilon \in [0, 1)$, our main theorem implies that $\mathcal{C}_\varepsilon \subseteq$ strong converse bound. Combining with the proposition above, we have $\mathcal{C}_\varepsilon \subseteq$ NNC inner bound. Since NNC inner bound $\subseteq \mathcal{C}_\varepsilon$, we have

$$\mathcal{C}_\varepsilon = \text{NNC inner bound}.$$

Rényi Divergence

Definition

The conditional Rényi divergence with parameter $\lambda \in [1, \infty)$ between $p_{X|Z}$ and $q_{X|Z}$ given r_Z is

$$D_\lambda(p_{X|Z} \| q_{X|Z} | r_Z) \triangleq \begin{cases} \frac{1}{\lambda-1} \log \sum_{z \in \mathcal{Z}} r_Z(z) \sum_{x \in \mathcal{X}} \frac{(p_{X|Z}(x|z))^\lambda}{(q_{X|Z}(x|z))^{\lambda-1}} & \text{if } \lambda > 1, \\ D(p_{X|Z} \| q_{X|Z} | r_Z) & \text{if } \lambda = 1 \end{cases}$$

where

$$D(p_{X|Z} \| q_{X|Z} | r_Z) \triangleq \sum_{z \in \mathcal{Z}} r_Z(z) \sum_{x \in \mathcal{X}} p_{X|Z}(x|z) \log \frac{p_{X|Z}(x|z)}{q_{X|Z}(x|z)}$$

is the relative entropy.

Data processing inequality (DPI)

$$D_\lambda(p_X \| q_X) \geq D_\lambda(p_{g(X)} \| q_{g(X)})$$

for any function g. In particular, $D_\lambda(p_{X,Y} \| q_{X,Y}) \geq D_\lambda(p_X \| q_X)$.

Properties of Rényi Divergence

Lemma 1 (Mutual information approximation)

$$D_\lambda(p_{XY|Z}\|p_{X|Z}p_{Y|Z}|r_Z) \le D(p_{XY|Z}\|p_{X|Z}p_{Y|Z}|r_Z) + 8(\lambda-1)|\mathcal{X}|^5|\mathcal{Y}|^5$$
$$= I_{r_Z p_{XY|Z}}(X;Y|Z) + 8(\lambda-1)|\mathcal{X}|^5|\mathcal{Y}|^5$$

Lemma 2 (Chain rule)

Given $\prod_{k=1}^{n} p_{Y_k|X_k}$, $\prod_{k=1}^{n} q_{Y_k|X_k}$ and $r_{X_{\mathcal{I}}^n}$, we have

$$D_\lambda\left(\prod_{k=1}^{n} p_{Y_k|X_k} \middle\| \prod_{k=1}^{n} q_{Y_k|X_k} \middle| r_{X_{\mathcal{I}}^n}\right) = \sum_{k=1}^{n} D_\lambda(p_{Y_k|X_k}\|q_{Y_k|X_k}|r_{X_k}^{(\lambda)})$$

where $r_{X_k}^{(\lambda)}$ which is determined by λ, $\prod_{m=1}^{k-1} p_{Y_m|X_m}$, $\prod_{m=1}^{k-1} q_{Y_m|X_m}$ and $r_{X_{\mathcal{I}}^k}$

Recalling Proof Steps for Cut-Set Bound

1. Lower bounding the error probability in terms of mutual information using Fano's inequality
   $$n\sum_{i\in T} R_i \leq I(W_T;\hat{W}_{T,d}|W_{T^c}) + 1 + \varepsilon_n n \sum_{i\in T} R$$
2. Using the DPI of the mutual information to introduce the channel output
   $$I(W_T;\hat{W}_{T,d}|W_{T^c}) \leq I(W_T;Y^n_{T^c}|W_{T^c})$$
3. Single-letterizing the mutual information
   $$I(W_T;Y^n_{T^c}|W_{T^c}) \leq \sum_{k=1}^n I(X_{T,k};Y_{T^c,k}|X_{T^c,k})$$
4. Introduction of a time-sharing random variable Q_n
   $$\sum_{k=1}^n I(X_{T,k};Y_{T^c,k}|X_{T^c,k}) \leq nI(X_{T,Q_n};Y_{T^c,Q_n}|X_{T^c,Q_n})$$
5. Combining the above inequalities, using $\lim_{n\to\infty}\varepsilon_n = 0$
   $$\sum_{i\in T} R_i \leq I(X_T;Y_{T^c}|X_{T^c})$$

Proof Steps Using Rényi Divergence

1) Lower bounding the error probability in terms of the Rényi divergence

$$\sum_{i\in T} nR_i \le D_\lambda(p_{W_T,\hat{W}_{T,d}} \| p_{W_T} s_{\hat{W}_{T,d}}) + \lambda(\lambda-1)^{-1}\log\left(\frac{1}{1-\varepsilon_n}\right)$$

for any choice of $s_{\hat{W}_{T,d}}$.

2) Using the DPI of the Rényi divergence to introduce the channel input and output with a proper choice of $s_{X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{T,d}}$

$$D_\lambda(p_{W_T,\hat{W}_{T,d}} \| p_{W_T} s_{\hat{W}_{T,d}}) \le D_\lambda\left(\prod_{k=1}^n q_{Y_{T^c,k}|X_{\mathcal{I},k}} \middle\| \prod_{k=1}^n s_{Y_{T^c,k}|X_{\mathcal{I},k}} \middle| p_{X_{\mathcal{I}}^n}\right)$$

3) Single-letterizing the Rényi divergence using the chain rule

$$D_\lambda\left(\prod_{k=1}^n q_{Y_{T^c,k}|X_{\mathcal{I},k}} \middle\| \prod_{k=1}^n s_{Y_{T^c,k}|X_{\mathcal{I},k}} \middle| p_{X_{\mathcal{I}}^n}\right) = \sum_{k=1}^n D_\lambda(q_{Y_{T^c}|X_{\mathcal{I}}} \| s_{Y_{T^c,k}|X_{T^c,k}} | p^{(\lambda)}_{X_{\mathcal{I},k}}).$$

Proof Steps Using Rényi Divergence

4) Representing distributions in the Rényi divergence by a single distribution
   - Due to the careful choice of $s_{X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{T\times\{d\}}}$, we can define $\tilde{p}^{(\lambda)}_{X_{\mathcal{I},k}, Y_{T^c,k}}$ s.t.

$$\sum_{k=1}^{n} D_\lambda(q_{Y_{T^c}|X_{\mathcal{I}}} \| s_{Y_{T^c,k}|X_{T^c,k}} | p^{(\lambda)}_{X_{\mathcal{I},k}}) \le \sum_{k=1}^{n} D_\lambda(\tilde{p}^{(\lambda)}_{Y_{T^c,k}|X_{\mathcal{I},k}} \| \tilde{p}^{(\lambda)}_{Y_{T^c,k}|X_{T^c,k}} | \tilde{p}^{(\lambda)}_{X_{\mathcal{I},k}}).$$

5) Introduction of a time-sharing variable followed by approximating I with D_λ.
   - Using a time-sharing variable Q_n and letting $\lambda = 1 + \frac{1}{\sqrt{n}}$,

$$\begin{aligned}
&\sum_{k=1}^{n} D_{1+\frac{1}{\sqrt{n}}}(\tilde{p}^{(1+\frac{1}{\sqrt{n}})}_{Y_{T^c,k}|X_{\mathcal{I},k}} \| \tilde{p}^{(1+\frac{1}{\sqrt{n}})}_{Y_{T^c,k}|X_{T^c,k}} | \tilde{p}^{(1+\frac{1}{\sqrt{n}})}_{X_{\mathcal{I},k}}) \\
&\le n D_{1+\frac{1}{\sqrt{n}}}(\tilde{p}^{(1+\frac{1}{\sqrt{n}})}_{Y_{T^c,Q_n}|X_{\mathcal{I},Q_n}} \| \tilde{p}^{(1+\frac{1}{\sqrt{n}})}_{Y_{T^c,Q_n}|X_{T^c,Q_n}} | \tilde{p}^{(1+\frac{1}{\sqrt{n}})}_{X_{\mathcal{I},Q_n}}) \\
&\le nI(X_T; Y_{T^c}|X_{T^c}) + 8|\mathcal{X}|^5|\mathcal{Y}|^5\sqrt{n}
\end{aligned}$$

   - Combining the steps and letting $n \to \infty$,

$$\sum_{i\in T} R_i \le I(X_T; Y_{T^c}|X_{T^c}).$$

Conclusion

- In a multimessage multicast network (MMN), every source node transmits a message and every destination node decodes all the messages.
- A strong converse outer bound for the discrete memoryless MMN have been established using Rényi divergence, i.e., outer bound on $\mathcal{C}_\varepsilon$.
- For any sequence of codes that operate at a rate tuple outside the strong converse bound, the average probability of decoding error must tend to 1.
- Our strong converse bound implies the strong converse some classes of MMNs including
  - The finite-field linear deterministic network.
  - The MMN consisting of independent DMCs.
  - The wireless erasure network.
- For the aforementioned MMNs, we have fully characterized their ε-capacity regions, which coincide with the NNC inner bound and the cut-set bound.
- Open problem: Prove or disprove that the cut-set bound contains $\mathcal{C}_\varepsilon$.

References

Avestimehr, A. S., Diggavi, S. N., and Tse, D. N. (2011).
Wireless network information flow: A deterministic approach.
IEEE Trans. Inf. Theory, 57(4):1872–1905.

Dana, A. F., Gowaikar, R., Palanki, R., Hassibi, B., and Effros, M. (2006).
Capacity of wireless erasure networks.
IEEE Trans. Inf. Theory, 52(3):789–804.

Köetter, R., Effros, M., and Médard, M. (2011).
A theory of network equivalence — Part I: Point-to-point channels.
IEEE Trans. Inf. Theory, 57(2):972–995.

Li, S.-Y. R., Yeung, R. W., and Cai, N. (2003).
Linear network coding.
IEEE Trans. Inf. Theory, 49(2):371–381.

Lim, S. H., Kim, Y.-H., El Gamal, A., and Chung, S.-Y. (2011).
Noisy network coding.
IEEE Trans. on Inform. Th., 57(5):3132–52.

Ogawa, T. and Nagaoka, H. (1999).
Strong converse to the quantum channel coding theorem.
IEEE Trans. on Inform. Th., 45(7):2486–2489.

Polyanskiy, Y. and Verdú, S. (2010).
Arimoto channel coding converse and Rényi divergence.
In *Proc. Allerton Conference on Communication, Control and Computing*, pages 1327 – 1333.

Wilde, M. M., Winter, A., and Yang, D. (2013).
Strong converse for the classical capacity of entanglement-breaking and hadamard channels via a sandwiched Rényi relative entropy.
ArXiv Preprint