Network Information Theory and the Entropy Power Inequality

Sormeh Shadbakht and Babak Hassibi

Department of Electrical Engineering
California Institute of Technology

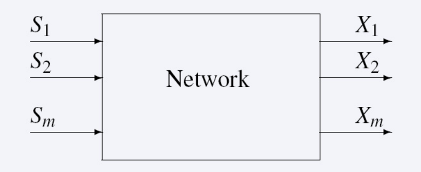
2nd Annual North American School of Information Theory, Evanston, IL
August 10-13, 2009

Motivation

- Growing interest in information transmission over networks and hence *multi-user information theory*
    - wired networks, e.g. the Internet
    - emerging wireless networks, e.g. sensor networks, ad hoc networks, etc.
- Many multi-user information theory problems still open (in contrast to the point-to-point case)
  $\Rightarrow$ Much less deployment of information theory in the design of current networks
- Difficulties in developing *network information theory*
- One of the fundamental difficulties: *Characterizing the Entropy Region*

A Network Problem

Consider an acyclic discrete memory-less network where transmitter S_i needs to communicate with the receiver X_i at rate R_i



The *rate region* for reliable communication is (cf. van der Meulen, Ahlswede)

$$\mathcal{R} = \text{cl}\left\{ R_i, i = 1, \ldots, m \;\middle|\; R_i < \frac{1}{T}\left(H(X_i^T) - H(X_i^T | S_i^T) \right) \right\} \quad \text{as } T \to \infty$$

- ${S_i}^T$, ${X_i}^T$: concatenated over T channel uses
- While not a surprising characterization, it is difficult to compute

- Equivalently we can represent the rate region through its tangent hyperplanes, for which we need to solve,

$$\lim_{T\to\infty} \sup_{p(S_i^T) \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) - H(X_i^T|S_i^T) \right)$$

- An extremely difficult problem, since
  - it is infinite-dimensional (*infinite-letter characterization*)
  - for any T, the problem is highly non-convex in the $p(S_i^T)$ and the "network operations"

Can these issues be resolved?

Entropy Vectors

- Notion of **Entropy Vectors** simplifies such problem
- **Definition:**
  - Consider n discrete random variables with alphabet-size N. For any set $\alpha \subseteq \{1, \ldots, n\}$, the *normalized entropy* is

    $$h_\alpha = \frac{1}{\log N} H(X_i, i \in \alpha).$$

    The $2^n - 1$ dimensional vector obtained from these entropies is called a *normalized entropy vector*.
  - Conversely, any $2^n - 1$ dimensional vector which can be regarded as the normalized entropy vector of some collection of n random variables, for some alphabet size N, is called *normalized-entropic*.
- Similarly we can define non-normalized entropy vectors through

$$H_\alpha = H(X_i, i \in \alpha)$$

- Γ_n^*: The space of all non-normalized entropic vectors
- Ω_n^*: The region of all normalized entropy vectors
- Focusing on *normalized* entropy seems to be more natural
  - it comes up in

$$\sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) + H(S_i^T) - H(X_i^T, S_i^T) \right)$$

  - It makes the entropy region finite,

$$h_\alpha \leq |\alpha|$$

Theorem (Ω_n^* and Γ_n^*)

Let $\mathrm{cone}(S) = \{\alpha X | \alpha \geq 0, X \in S\}$*. Then we have,*

$$\mathrm{cone}(\overline{\Omega}_n^*) = \overline{\Gamma}_n^*$$

Moreover $\overline{\Omega}_n^*$ *and* $\overline{\Gamma}_n^*$ *are convex sets.*

Network Problem Revisited

Thus the network problem can be rewritten as

$$\sup \alpha^T h,$$

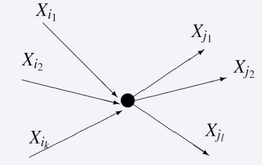
subject to $h \in \Omega_n^*$ (where n is the number of random variables in the network) and subject to the network constraints:

1. topological constraints
2. channel constraints

Let's examine these constraints...

Topological Constraints

- Let i be a non-source node, receiving the signals $X_{i_1}, \ldots, X_{i_k}$ and transmitting the signals $X_{j_1}, \ldots, X_{j_l}$



The following linear constraint is the result of causality:

$$h(X_{j_q}, X_{i_1}, \ldots, X_{i_k}) - h(X_{i_1}, \ldots, X_{i_k}) = 0 \quad q = 1, \ldots l$$

- At source nodes,
  $h(S_i, S_j) - h(S_i) - h(S_j) = 0$ if S_i and S_j are independent and
  $h(S_i, S_j) = h(S_i) = h(S_j)$ if $S_i = S_j$

Channel Constraints

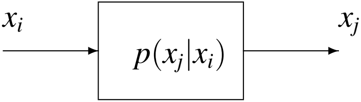
- Channels constrain the joint distribution of all random variables in the network $\Longrightarrow$ admissible entropy vectors:

  $$p(X_i, X_j) = p(X_j|X_i)p(X_i),$$

  or, equivalently,

  $$\int \prod_{k \neq i,j} dX_k \, p(X_1, \ldots, X_n) = p(X_j|X_i) \int \prod_{k \neq i} dX_k \, p(X_1, \ldots, X_n),$$

  which is a linear constraint on the joint distribution



- Channel constraints do not translate directly to entropies
- Let $\Omega^*_{n,C}$ denote the space of entropic vectors constrained by the discrete memoryless channels in the network

Back to the Network Problem

Characterizing $\Omega^*_{n,C}$ is not an easy task in general, however the region of channel constrained entropy vectors remains convex,

Theorem (Network Problem as a Convex Optimization)

Network problem can be cast as a convex optimization over the set of channel constrained entropy vectors:

$$\max_{h \in \Omega^*_{n,C}, Ah=0} \alpha^T h,$$

where $Ah = 0$ *represents the topological constraints.*

- Circumvented both the *infinite-letter characterization* problem, as well as the *non-convexity*
- *Characterization of* $\Omega^*_{n,C}$ *is of fundamental importance in network information theory*

Some Consequences: Duality and Cutset bounds

Convex optimization machinery, yields interesting consequences, e.g. using the duality argument we obtain:

$$\max_{h\in\Omega^*_{n,C}, Ah=0} \alpha^T h = \max_{h\in\Omega^*_{n,C}} \min_{\lambda} \left(\alpha^T h + \lambda^T A h\right) = \min_{\lambda} \max_{h\in\Omega^*_{n,C}} (\alpha + A^T\lambda)^T h$$

Now any choice of λ yields an upper bound:

$$\max_{h\in\Omega^*_{n,C}, Ah=0} \alpha^T h \leq \max_{h\in\Omega^*_{n,C}} (\alpha + A^T\lambda)^T h$$

- Partition the nodes of the network into two sets, one containing sources and the other the destinations,
- If i is a node whose edges do not cross the cut, set $\lambda_i = 0$
- Optimizing over the remaining entries of λ boils down to the point-to-point problem

Wired vs. Wireless

- Characterizing $\Omega^*_{n,C}$ is formidable in general
- For wired networks things simplify,

Theorem

The network problem for wired networks can be cast as a convex optimization over the unconstrained *entropy region,*

$$\max_{h \in \Omega^*_n, h_{1:n} \le c, Ah = 0} \alpha^T h.$$

- For wireless networks, interaction of signals such as interference should be considered
  $\Longrightarrow$ Studying bounds on the entropy of sums of independent random variables

Entropy Region of $x,\ y$ and $z = x + y$

Let $x,\ y \in \mathcal{R}^m$ be independent vector-valued continuous random variables, $z = x + y$ and h the corresponding *normalized* entropy vector

Define,

$$\Psi_3 : \text{the entropy region of } x,\ y \text{ and } z$$

$$\Upsilon = \{h | h_{xy} = h_{xz} = h_{yz} = h_x + h_y,\ h_{xyz} = -\infty\},$$

$$\Xi = \{h | h_z \geq \frac{1}{2}\log(e^{2h_x} + e^{2h_y})\}.$$

Note that the set Ξ contains the points which satisfy the so-called *Entropy Power Inequality*

Theorem (Entropy region of $x,\ y$ and $z = x + y$)

$$\Psi_3 = \Gamma_3 \cap \Upsilon \cap \Xi.$$

Proof:

- from assumptions and the classical EPI, $\Psi_3 \subseteq \Gamma_3 \cap \Upsilon \cap \Xi$
- lower bound for EPI is achieved by Gaussian variables with proportional covariance matrices
- for fixed h_x and h_y, h_z can grow unbounded,

$$x : \mathcal{N}(0, \epsilon I_m + \sigma_x^2 U_x U_x^T)$$
$$y : \mathcal{N}(0, \epsilon I_m + \sigma_y^2 U_y U_y^T)$$

where U_x, U_y are $m \times m/2$ unitary matrices s.t,

$$U_x^T U_x = U_y^T U_y = I_{m/2} \text{ and } U_x^T U_y = 0$$

Calculating normalized entropy of Gaussians give,

$$h_x = \frac{1}{2}\log 2\pi e + \frac{1}{4}\log \epsilon(\epsilon + \sigma_x^2),$$
$$h_y = \frac{1}{2}\log 2\pi e + \frac{1}{4}\log \epsilon(\epsilon + \sigma_y^2).$$

since,

$$z : \mathcal{N}(0, 2\epsilon I_m + \sigma_x^2 U_x U_x^T + \sigma_y^2 U_y U_y^T),$$

entropy of z can be computed in the same fashion

$$h_z = \frac{1}{2}\log 2\pi e + \frac{1}{4}\log(2\epsilon + \sigma_x^2)(2\epsilon + \sigma_y^2).$$

Let,

$$\sigma_x^2 = \frac{e^{4c_x}}{(2\pi e)^2 \epsilon} - \epsilon \quad , \quad \sigma_y^2 = \frac{e^{4c_y}}{(2\pi e)^2 \epsilon} - \epsilon$$

for some finite positive c_x and c_y and let $\epsilon \to 0$, therefore we obtain,

$$h_x = c_x, \quad h_y = c_y, \quad h_z \to \infty.$$

- noting that Ξ is a convex set, completes the proof.

Conclusions

- Large class of network information theory problems can be cast as convex optimization problems over the convex set of *channel-constrained entropies*
- Cut-set bounds follow from convex optimization duality
- Characterizing Γ_n^* is of fundamental importance; open for $n \geq 4$
- Identifying the entropy region of sums of independent variables is important for wireless networks
- This region was obtained for 3 independent vector-valued continuously distributed random variables, $x,\ y$ and $z = x + y$; Entropy power inequality turns out to be the main constraining inequality in this case