Entropy Vectors and Network Information Theory

Sormeh Shadbakht and Babak Hassibi

Department of Electrical Engineering
Caltech

Lee Center Workshop
May 25, 2007

Outline

Introduction

There has been a growing interest in information transmission over networks, spurred by

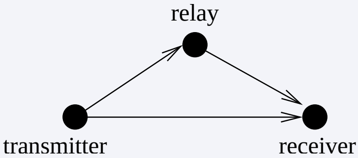
- wired networks, such as the Internet
- emerging wireless networks (sensor nets, ad hoc networks, etc.)

Historically, information theory has played a central role in the development of point-to-point communication systems. However, it has not been deployed enough in designing most of the networks currently in use. Although things are changing there are some difficulties in developing *network information theory* ; some of which we will highlight in this talk. (And perhaps suggest an approach.)

Network Information Theory

The study of wireless networks is related to **Multi-User Information Theory**, i.e., the study of the limits of information transfer among many users. In contrast to the Single-user case, many multi-user problems are not fully solved:

- Finding the capacity of even a three-node network is open;

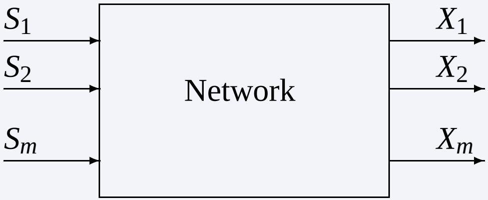


- situation reminiscent of mechanics where three-body problems have not been solved;

Therefore much of the focus has been on asymptotic results for *large* networks (analogous to statistical mechanics).

A Network Problem

Consider the following acyclic discrete memory-less network and assume that each source needs to transmit to its corresponding destination at rate R_i, $i = 1, 2, \ldots, m$:



It is shown that (cf. Ahlswede) the *rate region* for reliable communication is

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

Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

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

This problem is extremely difficult, since

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

Ergo: No one does it this way! However it motivates the next definition.

Entropy Vectors

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*.

- The space of non-normalized entropic vectors is denoted by Γ_n^*.
- We denote the region of normalized entropy vectors by Ω_n^*.

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 the entropy region finite, $h_\alpha \leq |\alpha|$, and the proof of the convexity of $\bar{\Omega}_n^*$ trivial.

Back to the network Problem

But what does all this say about our network problem?
Well we can now write it 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, which are of two kinds:

1. topological constraints which guarantee that outgoing variables of a node are functions of the ingoing variables and therefore enforce some linear relations between the corresponding entropies,
2. channel constraints which impose some linear constraint on the joint probability distributions.

Convex Formulation of the Network Problem

Theorem

Let $\Omega^*_{n,C}$ denote the space of entropic vectors that are constrained by the discrete memoryless channels in the network. Then the closure of this set, i.e., $\overline{\Omega}^*_{n,C}$, is convex.

Since the linear channel constraints have no effect on the proof of convexity of $\overline{\Omega}^*_n$. Our network problem thus becomes:

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

where $Ah = 0$ represents the topological constraints.

- Therefore we have circumvented both the *infinite-letter characterization* problem, and the *non-convexity*.

Of course, we still need to characterize the channel-constrained entropy space, $\Omega^*_{n,C}$ so we may not have really made the problem easier.

Ω_n^* (Γ_n^*)

- In some cases of interest such as the wired networks it turns out that we can only focus on Ω_n^*.

Clearly, the study of Ω_n^* (or Γ_n^*) is central to network problems. Although it has been the object of intense study in some quarters–mostly motivated by source coding–(Han, Fujishije, Yeung, Zhang, Chan, Romashchenko et al, Zeger et al) it has not gained as much attention as it perhaps should have.

- The work of Han, Fujishige, Zhang and Yeung, has resulted in the complete characterization of Ω_n^* (Γ_n^*) for $n = 2,3$ and their relation to polymatroids and submodular functions. In particular, entropy satisfies the following properties:
  1. $h_\emptyset = 0$
  2. For $\alpha \subseteq \beta$: $h_\alpha \leq h_\beta$
  3. For any α, β: $h_{\alpha\cup\beta} + h_{\alpha\cap\beta} \leq h_\alpha + h_\beta$

Sunmodular functions and Shannon Inequalities

- The last inequality is called the *submodularity property*.
- They are referred to as the basic inequalities of Shannon information measures and follow from

$$I(X_1;X_2|X_3) = h(X_1,X_3) + h(X_2,X_3) - h(X_3) - h(X_1,X_2,X_3) \geq 0.$$

- Any inequality obtained as positive linear combinations of these is referred to as *Shannon inequality*.
- The space of all vectors of $2^n - 1$ dimensions whose components satisfy all such Shannon inequalities is denoted by Ω_n (Γ_n). It has been shown that

$$\Omega_2^* \ (\Gamma_2^*) = \Omega_2 \ (\Gamma_2) \quad \text{and} \quad \bar{\Omega}_3^* \ (\bar{\Gamma}_3^*) = \Omega_3 \ (\Gamma_3)$$

Non-Shannon Inequalities

For $n \geq 4$, recently several *non-Shannon*-type information inequalities have been discovered (Zhang and Yeung, Romashchenko et al, Zeger et al). Here is the original one:

$$\begin{aligned} I(X_3;X_4) \quad &\leq \quad I(X_3;X_4|X_1) + I(X_3;X_4|X_2) + \frac{1}{2}I(X_1;X_2) + \\ &\qquad \frac{1}{4}I(X_1;X_3,X_4) + \frac{1}{4}I(X_2;X_3,X_4). \end{aligned}$$

These inequalities demonstrate that Ω_4^* (Γ_4^*) is strictly smaller than Ω_4 (Γ_4):

$$\Omega_4^* \ (\Gamma_4^*) \subset \Omega_4 \ (\Gamma_4).$$

An Attempt at Characterizing Ω_n^*

One way of characterizing Ω_n^* is through minimizing all linear functionals of normalized entropy:

$$\min_{h \in \Omega_n^*} \sum_{\alpha \subseteq \mathscr{N}} a_\alpha h_\alpha,$$

for any $a \in \mathscr{R}^{2^n - 1}$, where $\mathscr{N} = \{1, \ldots, n\}$. If we fix the alphabet size to N and attempt to optimize over the unknown joint distribution $p_{X_{\mathscr{N}}}(x_{\mathscr{N}})$ then the KKT conditions necessitate that

$$\sum_{\alpha \subseteq \mathscr{N}} a_\alpha \log \frac{1}{p_{X_\alpha}(x_\alpha)} = c \quad \text{if} \;\; p_{X_{\mathscr{N}}}(x_{\mathscr{N}}) \neq 0,$$

for some constant c.

Quasi-Uniform Distributions

Thus, rather than searching over all possible distributions $p_{X_{\mathcal{N}}}(x_{\mathcal{N}})$, we need only search over those distributions that satisfy

$$\sum_{\alpha \subseteq \mathcal{N}} a_\alpha \log \frac{1}{p_{X_\alpha}(x_\alpha)} = c \quad \text{if} \quad p_{X_{\mathcal{N}}}(x_{\mathcal{N}}) \neq 0,$$

The above can have many solutions. A suggestive solution, that does not depend on a is the following. For any $\alpha \subseteq \mathcal{N}$:

$$p_{X_\alpha}(x_\alpha) = c_\alpha \quad \text{or} \quad 0$$

for some constant c_α, independent of $x_\alpha \in \{1, \ldots, N\}^{|\alpha|}$. These distributions take on zero or a constant value for all possible marginals, $p_{X_\alpha}(\cdot)$. We refer to them as *quasi-uniform*. Computing their entropy is straightforward:

$$h_\alpha = \frac{\log 1/c_\alpha}{\log N}.$$

Let Λ_n denote the space of entropy vectors generated by quasi-uniform distributions. Then the remarkable result of Chan and Yeung (2002) is that

Theorem (Quasi-Uniform Distribution)

$\overline{con}(\Lambda_n) = \bar{\Gamma}_n^*$, *i.e., the convex closure of* Λ_n *is the closure of* Γ_n^*.

In other words, considering quasi-uniform distributions is sufficient for characterizing Γ_n^*. These are the distributions we will henceforth focus on.

Lattice Structure

Determining all quasi-uniform distributions appears to be a hopelessly complicated combinatorial problem. Since we are looking for a construction that can be generalized to any n, it seems reasonable to impose some structure. Consider the lattice

$$x = Mz,$$

where $x \in \mathscr{R}^n$ are points in the lattice, $M \in \mathscr{R}^{n\times n}$ is the so-called lattice-generating matrix, and $z \in \mathscr{Z}^n$ is an integer vector. To guarantee that x have integer entries, we will impose that M have integer entries. We will call the resulting lattice $\mathscr{L}(M)$.

Lattice-generated distributions

Definition (Lattice-Generated Distribution)

A probability distribution over n random variables with alphabet size N each, will be called *lattice-generated*, if for some lattice $\mathscr{L}(M)$, we have

$$p_{X_{\mathcal{N}}}(x_{\mathcal{N}}) = \begin{cases} c & \text{if } x_{\mathcal{N}} \in \{0,\dots,N-1\}^n \cap \mathscr{L}(M), \\ 0 & \text{otherwise.} \end{cases}$$

But when a lattice-generated distribution is quasi-uniform?

Lemma (Lattice-Generated Quasi-Uniform Distributions)

A lattice-generated distribution is quasi-uniform if the lattice has a period that divides N.
The latter is true if, and only if, the matrix $M^{-1}N$ has integer entries.

Lemma (Entropy Extraction)

Consider a lattice-generated distribution with period dividing N. Consider any collection of random variables X_α and partition the rows of the lattice-generating matrix M accordingly:

$$M = \begin{pmatrix} M_\alpha \\ M_{\alpha^c} \end{pmatrix},$$

where M_α is $|\alpha|$-by-n. Then the normalized entropy of X_α is:

$$h_\alpha = |\alpha| - \frac{\log\left(\text{gcd}\left(\text{all } |\alpha|\text{-by-}|\alpha| \text{ minors of } M_\alpha\right)\right)}{\log N}.$$

In our case, this calculation simplifies to:

$$h_\alpha = |\alpha| - \frac{\log\left(\min_\tau (\Pi_{i\in\{\text{row indices of } M_\alpha\}} N^{\gamma_{i,\tau(i)}})\right)}{\log N}$$

where τ is a permutation of column indices of M_α.

An Inner Region for Ω_n^*

Let Δ_n denote the space of entropy vectors obtained from lattice-generated quasi-uniform distributions.

Theorem (An Inner Region for Entropic Vectors)

$$\overline{\textit{con}}(\Delta_n) \subseteq \bar{\Omega}_n^*$$

where $\overline{\textit{con}}(\cdot)$ represents the convex closure. This region is furthermore a polytope.

Two Random Variables

It turns out that wlog we may take M lower triangular:

$$M = \begin{pmatrix} M_{12} & 0 \\ M_{21} & M_{22} \end{pmatrix}.$$

In fact, for any integers M_{ij} it is always possible to find a large enough integer N and positive rational numbers γ_{ij} such that $M_{ij} = N^{\gamma_{ij}}$. Furthermore, for large enough N it follows that $\gcd(N^{\gamma_{ij}}, N^{\gamma_{kl}}) = N^{\min(\gamma_{ij},\gamma_{kl})}$. We will therefore take

$$M = \begin{pmatrix} N^{\gamma_{11}} & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} \end{pmatrix},$$

from which it follows:

$$h_1 = 1 - \gamma_{11} \quad , \quad h_2 = 1 - \min(\gamma_{21}, \gamma_{22}) \quad , \quad h_{12} = 2 - \gamma_{11} - \gamma_{22}$$

We still need to check the condition that M generates a quasi-uniform distribution. This requires that

$$NM^{-1} = \begin{pmatrix} N^{1-\gamma_{11}} & 0 \\ -N^{1-\gamma_{11}-\gamma_{22}+\gamma_{21}} & N^{1-\gamma_{22}} \end{pmatrix}$$

have integer entries, which means that

$$\gamma_{11} \leq 1 \quad , \quad \gamma_{22} \leq 1 \quad , \quad \gamma_{11}+\gamma_{22} \leq 1+\gamma_{21}.$$

It is straightforward to show that the convex hull is

$$\begin{cases} h_1 = 1-\gamma_{11} \quad , \quad h_2 = 1-\gamma_{21} \quad , \quad h_{12} = 2-\gamma_{11}-\gamma_{22} \\ 0 \leq \gamma_{11} \leq 1 \, , \, 0 \leq \gamma_{22} \leq 1 \, , \, 0 \leq \gamma_{21} \leq \gamma_{22} \, , \, \gamma_{11}+\gamma_{22} \leq 1+\gamma_{21} \end{cases}$$

which can be shown to be Ω_2.

Three Random Variables

$$M = \begin{pmatrix} N^{\gamma_{11}} & 0 & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} \end{pmatrix}.$$

Insisting on quasi-uniformity by forcing the elements of $M^{-1}N$ to be integers:

$$0 \leq \gamma_{ij} \leq 1$$
$$\gamma_{ii} + \gamma_{jj} - \gamma_{ij} \leq 1 \quad i \geq j$$
$$\gamma_{11} + \gamma_{22} + \gamma_{33} - \gamma_{21} - \gamma_{32} \leq 1$$

Extracting the entropies

$$
\begin{aligned}
h_1 &= 1-\gamma_{11} \\
h_2 &= 1-\min(\gamma_{21},\gamma_{22}) \\
h_3 &= 1-\min(\gamma_{31},\gamma_{32},\gamma_{33}) \\
h_{12} &= 2-\gamma_{11}-\gamma_{22} \\
h_{13} &= 2-\gamma_{11}-\min(\gamma_{32},\gamma_{33}) \\
h_{23} &= 2-\min(\gamma_{21}+\min(\gamma_{32},\gamma_{33}),\gamma_{22}+\min(\gamma_{31}+\gamma_{33})) \\
h_{123} &= 3-\gamma_{11}-\gamma_{22}-\gamma_{33}
\end{aligned}
$$

$$
M = \begin{pmatrix} N^{\gamma_{11}} & 0 & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} \end{pmatrix}.
$$

Theorem

$$\overline{con}(\Delta_3) = \bar{\Omega}_3^*$$

Proof: Clearly, $\overline{\text{con}}\Delta_3 \subseteq \bar{\Omega}_3^*$ since all vectors in $\overline{\text{con}}\Delta_3$ are entropic. Conversely, consider the region defined by

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_{12} \\ h_{23} \\ h_{31} \\ h_{123} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \\ k_5 \\ k_6 \\ k_7 \\ k_8 \end{pmatrix}, \quad k_i \geq 0.$$

Each column in the matrix can be obtained by a lattice-generated distribution. Therefore the region must be a subset of $\overline{\text{con}}\Delta_3$.

Write the above matrix equation as

$$h = \begin{pmatrix} A & a \end{pmatrix} \begin{pmatrix} k \\ k_8 \end{pmatrix} = Ak + ak_8,$$

so

$$A^{-1}h - A^{-1}ak_8 = k \geq 0.$$

Computing A^{-1} and $A^{-1}ak_8$, yields

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_{12} \\ h_{23} \\ h_{31} \\ h_{123} \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} k_8$$

The point is to show that for any entropic vector h one can find a non-negative k_8 such that the above inequalities are satisfied. The first three are clearly satisfied. The next three are satisfied provided

$$k_8 \leq \min_{i,j,k}(-h_i + h_{ij} + h_{ki} - h_{ijk}) \geq 0,$$

and the last inequality if

$$k_8 \geq -\sum_i h_i + \sum_{i,j} h_{ij} - h_{ijk} = -h_1 - h_2 - h_3 + h_{12} + h_{23} + h_{31} - h_{123}$$

It is straightforward to show that the upper bound on k_8 exceeds the lower bound and so the region for k_8 is non-empty:

$$-h_1 + h_{12} + h_{31} - h_{123} \geq -h_1 - h_2 - h_3 + h_{12} + h_{23} + h_{31} - h_{123}$$

□

Four Random Variables

$$M = \begin{pmatrix} N^{\gamma_{11}} & 0 & 0 & 0 \\ N^{\gamma_{21}} & N^{\gamma_{22}} & 0 & 0 \\ N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}} & 0 \\ N^{\gamma_{41}} & N^{\gamma_{42}} & N^{\gamma_{43}} & N^{\gamma_{44}} \end{pmatrix}.$$

MN^{-1} having integer entries yields

$$\begin{aligned} &0 \le \gamma_{ii} \le 1, \\ &\gamma_{ii} + \gamma_{jj} - \gamma_{ij} \le 1, \quad i > j \\ &\gamma_{ii} + \gamma_{jj} + \gamma_{kk} - \gamma_{ij} - \gamma_{jk} \le 1, \quad i > j > k \\ &\gamma_{11} + \gamma_{22} + \gamma_{33} + \gamma_{44} - \gamma_{21} - \gamma_{32} - \gamma_{43} \le 1, \end{aligned} \tag{1}$$

Some of the entropies are as follows:

$$
\begin{aligned}
h_1 &= 1-\gamma_{11}\\
h_2 &= 1-\min(\gamma_{21},\gamma_{22})\\
h_3 &= 1-\min(\gamma_{31},\gamma_{32},\gamma_{33})\\
h_4 &= 1-\min(\gamma_{41},\gamma_{42},\gamma_{43},\gamma_{44})\\
h_{12} &= 2-\gamma_{11}-\gamma_{22}\\
h_{13} &= 2-\gamma_{11}-\min(\gamma_{32},\gamma_{33})\\
h_{14} &= 2-\gamma_{11}-\min(\gamma_{42},\gamma_{43},\gamma_{44})\\
h_{23} &= 2-\min(\gamma_{21}+\min(\gamma_{32},\gamma_{33}),\gamma_{22}+\min(\gamma_{31}+\gamma_{33}))\\
h_{24} &= 2-\min(\gamma_{21}+\min(\gamma_{42},\gamma_{43},\gamma_{44}),\gamma_{22}+\min(\gamma_{41},\gamma_{43},\gamma_{44}))\\
h_{34} &= 2-\min(\gamma_{31}+\min(\gamma_{42},\gamma_{43},\gamma_{44}),\gamma_{32}+\min(\gamma_{41},\gamma_{43},\gamma_{44}),\gamma_{33}+\\
&\qquad \min(\gamma_{41},\gamma_{42},\gamma_{44}))\\
h_{123} &= 3-\gamma_{11}-\gamma_{22}-\gamma_{33}\\
h_{124} &= 3-\gamma_{11}-\gamma_{22}-\min(\gamma_{43},\gamma_{44})
\end{aligned}
$$

etc.

Conclusions:

- Showed that a large class of network information theory problems can be cast as convex optimization problems over the convex set of *channel-constrained entropy vectors*.
- Thus, the problem is to characterize the space of entropy vectors.
- Characterizing Ω_n^* for $n \geq 4$ is a fundamental open problem related to non-Shannon inequalities.
- Constructed an inner bound on Ω_n^* using lattice-based distributions which is at least tight for $n = 2, 3$. Furthermore the inner bound is a polytope allowing for efficient use via linear programming.

Thank you!