Entropy Vectors and Network Information Theory

Babak Hassibi and Sormeh Shadbakht
Department of Electrical Engineering
California Institute of Technology, Pasadena, CA 91125

Nortel Networks Seminar Series
University of Waterloo, December 3, 2007
Outline

• Introduction

• Network Information Theory
  – convex optimization problem over entropy vectors
  – duality and cut-set bounds
  – wired networks

• Characterization of $\Gamma_n^*$
  – Shannon and non-Shannon inequalities
  – quasi-uniform distributions
  – inner bound using lattice-based distributions
  – groups, Gaussians, and hyperdeterminants

• Conclusion
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, information theory has had virtually no impact on the design of most of the networks currently in use, such as the Internet.

- though there are some exceptions (Aloha)
- and things are changing (wireless networks, network coding)

In this talk, we will highlight some of the difficulties in developing network information theory. (And perhaps suggest an approach.)
**Network Information Theory**

**Multi-User Information Theory** is the study of the limits of information transfer between many users.

- single-user information theory is well understood (Shannon 1948)

\[
C = \max_{p_X(\cdot)} \{ H(y) - H(y|x) \}
\]

- multi-user information theory, however, is not: computing the capacity of even a three-node network is open

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

$$R = \text{cl} \left\{ R_i, i = 1, \ldots, m \mid R_i < \frac{1}{T} \left( H(X_i^T) - H(X_i^T | S_i^T) \right) \right\} \quad \text{as } T \to \infty$$
Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

$$\lim_{T \to \infty} \sup \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left( H(X_i^T) - H(X_i^T | S_i^T) \right)$$

This problem is notoriously 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!
Consider \( n \) discrete random variables with alphabet-size \( N \). For any set \( S \subseteq \{1, \ldots, n\} \), we have the normalized entropy \( h_S = \frac{1}{\log N} H(X_i, i \in S) \). The \( 2^n - 1 \) dimensional vector obtained from these entropies, is called an entropy vector.

Conversely, any \( 2^n - 1 \) dimensional vector which can be regarded as the entropy vector of some collection of \( n \) random variables, for some value of \( N \), is called entropic.

The space of entropic vectors is denoted by \( \Gamma_n^* \).

We have focused on normalized entropy, since it is what some sup in

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

and since it makes the the space \( \Gamma_n^* \) compact (a finite region), \( h_S \leq |S| \).
Convexity of $\Gamma^*_n$

We should not that, for any fixed $N$, the set of entropy vectors is highly non-convex.

However, $\Gamma^*_n$ is convex (precisely because $N$ is not fixed).

- One simple proof uses time-sharing: Suppose $h_x \in \Gamma^*_n$, corresponding to random variables $X_1, \ldots, X_n$ with alphabet-size $N_x$ and $h_y \in \Gamma^*_n$ corresponding to random variables $Y_1, \ldots, Y_n$ with alphabet-size $N_y$. Make $n_x$ independent copies of the first set and $n_y$ independent copies of the second so that together the new concatenated random variables have alphabet-size $N_x^{n_x} N_y^{n_y}$. The resulting entropy vector is clearly

$$\frac{n_x \log N_x}{n_x \log N_x + n_y \log N_y} h_x + \frac{n_y \log N_y}{n_x \log N_x + n_y \log N_y} h_y,$$

which, since $n_x$ and $n_y$ are arbitrary, implies convexity.
An Alternative Proof

If we form the convex combination of two distributions of fixed alphabet-size $N$:

$$p_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = p_\theta p_{X_1,\ldots,X_n}(z_1,\ldots,z_n) + (1-p_\theta)p_{Y_1,\ldots,Y_n}(z_1,\ldots,z_n),$$

it is certainly not true that

$$h_z = p_\theta h_x + (1-p_\theta)h_y.$$ 

However, it is true in the limit!
Thus, make $T$ independent copies of each random variable and consider the distribution

$$p_\theta \prod_{t=1}^{T} p_{X_t, \ldots, X_n}(z_{1t}, \ldots, z_{nt}) + (1 - p_\theta) \prod_{t=1}^{T} p_{Y_t, \ldots, Y_n}(z_{1t}, \ldots, z_{nt})$$

Now for any $S \subseteq \{1, \ldots, n\}$, we have

$$\underbrace{H(Z_S^T | \theta)}_{p_\theta H(X_S^T) + (1 - p_\theta) H(Y_S^T)} \leq H(Z_S^T) \leq \underbrace{H(Z_S^T, \theta)}_{= H(Z_S^T | \theta) + H(\theta)}$$

Normalizing by $\log N^T$ yields

$$p_\theta h_x + (1 - p_\theta) h_y \leq h_z \leq p_\theta h_x + (1 - p_\theta) h_y + \frac{p_\theta \log p_\theta - (1 - p_\theta) \log(1 - p_\theta)}{T \log N}$$

which shows convexity as $T \to \infty$.

**Theorem 1** The closure of the set of entropic vectors, $\bar{\Gamma}_n^*$, is convex.
But what does all this say about our network problem?

Well we can now write it as

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

subject to $h \in \Gamma_n^*$ (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
2. channel constraints
Assume the signals $X_{i_1}, \ldots, X_{i_k}$ arrive at a non-source node and the signals $X_{j_1}, \ldots, X_{j_l}$ are transmitted. This can be represented as the following linear constraints on the entropy vector:

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

At source nodes, if $S_i$ and $S_j$ are independent,

$$h(S_i, S_j) - h(S_i) - h(S_j) = 0,$$

and if $S_i = S_j$, $h(S_i, S_j) = h(S_i) = h(S_j)$. 

$h$ represents entropy.
Channel constraints do not translate directly to entropies. What they do is constrain the joint distribution of all random variables in the network (which then determines the admissible entropy vectors):

\[ p(X_i, X_j) = p(X_j | X_i) p(X_i), \]

or, equivalently,

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

which is a linear constraint on the joint distribution.
Convex Formulation of the Network Problem

Theorem 2 Let $\Gamma^*_{n,C}$ 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., $\overline{\Gamma}^*_{n,C}$, is convex.

Since the linear channel constraints have no effect on the two proofs of convexity that have been given. Our network problem thus becomes:

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

where $Ah = 0$ represents the topological constraints.

Thus, by going to the space of entropy vectors, we have circumvented both the infinite-letter characterization problem, as well as the non-convexity.

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

However, the machinery of convex optimization yields some interesting consequences.

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

Now any choice of \( \lambda \) yields an upper bound:

\[
\max_{h \in \Gamma^*_{n,C}, Ah=0} \alpha^T h \leq \max_{h \in \Gamma^*_{n,C}} \left( \alpha + A^T \lambda \right)^T h
\]

Consider any partition of the nodes of the network into two sets, one containing the sources and the other containing the destinations, and set the entries of \( \lambda \) that correspond to nodes whose edges do not cross the cut to zero. Optimizing over the remaining components of \( \lambda \) becomes the point-to-point problem of communicating between these two sets of nodes. These are the well known cut-set bounds.
In the current framework, solving network information theory problems requires characterizing the set $\Gamma^*_{n,C}$. This seems formidable (to say the least). However, for wired networks things simplify considerably.

In a wired network, each link represents a (discrete memoryless) channel. Furthermore, the signals transmitted on outgoing edges of a node $(X_i, X_j)$ can be different, while the signals impinging on a node $(X_k, X_l)$ are received without interference.
Clearly, if the signal $X_i$ is the input to a certain channel on link $i$ and $Y_i$ is the corresponding output, the channel imposes the following constraints

$$h(X_i) + h(Y_i) - h(X_i, Y_i) \leq C_i,$$

where $C_i$ is the capacity of the link. We can therefore define a random variable $Z_i$, such that

$$h(Z_i) = h(X_i) + h(Y_i) - h(X_i, Y_i) \leq C_i.$$

It turns out that, in the wired case, all the quantities of interest can be expressed in terms of the $Z_i$. In fact, for the $Z_i$:

$$\Gamma^*_{n, C} = \Gamma_n^* \cap \{h | h_{1:n} \leq c\},$$

where $c$ is the vector of link capacities and $n$ is the number of links in the network.
Thus, the network problem becomes:

$$\max_{h \in \Gamma_n^*, h_1:n \leq c, Ah=0} \alpha^T h.$$ 

This is significant for two reasons:

- For wired networks, the channels affect the rate region only through their capacities.
- For wired networks, determining the rate region requires determining only $\Gamma_n^*$.

This result has also been noted by R. Koetter, R. Yeung and others.

We shall therefore for the remainder of this talk focus on $\Gamma_n^*$.
Clearly, the study of $\Gamma^*_n$ is central to network problems. Although it has been the object of intense study in some quarters—mostly motivated by source coding—(Han, Fujishige, 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 $\Gamma^*_n$ 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 $\alpha, \beta$: $h_{\alpha \cup \beta} + h_{\alpha \cap \beta} \leq h_\alpha + h_\beta$
Submodular Functions and Shannon Inequalities

- The last inequality is called the \textit{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 inequalities obtained as positive linear combinations of these are referred to as \textit{Shannon inequalities}.
- The space of all vectors of $2^n - 1$ dimensions whose components satisfy all such Shannon inequalities is denoted by $\Gamma_n$ (it is essentially the space of all matroidal vectors). It has been shown that

\[ \Gamma_2^* = \Gamma_2 \quad \text{and} \quad \Gamma_3^* = \Gamma_3 \]
For $n \geq 4$, recently several \textit{non-Shannon}-type information inequalities have been discovered (Zhang and Yeung, Romashchenko et al, Zeger et al, Matus). Here is the original one:

$$I(X_3; X_4) \leq I(X_3; X_4|X_1) + I(X_3; X_4|X_2) + \frac{1}{2} I(X_1; X_2) + \frac{1}{4} I(X_1; X_3, X_4) + \frac{1}{4} I(X_2; X_3, X_4).$$

These inequalities demonstrate that $\Gamma^*_4$ is strictly smaller than $\Gamma_4$:

$$\Gamma^*_4 \subset \Gamma_4.$$
An Attempt at Characterizing $\Gamma^*_n$

One way of characterizing $\Gamma^*_n$ is through minimizing all linear functionals of normalized entropy:

$$\min_{h \in \Gamma^*_n} \sum_{\alpha \subseteq \mathcal{N}} a_\alpha h_\alpha,$$

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

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

for some constant $c$. 

Quasi-Uniform Distributions

Thus, rather than searching over all possible distributions \( p_{X_N}(x_N) \), we need only search over those distributions that satisfy

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

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

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

for some constant \( c_\alpha \), independent of the point \( x_\alpha \in \{1, \ldots, N\}^{|\alpha|} \).

These are distributions that 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 $\Lambda_n$ denote the space of entropy vectors generated by quasi-uniform distributions. Then the remarkable result of Chan and Yeung (2002) is that

**Theorem 3 (Quasi-Uniform Distribution)** \( \bar{\Lambda}_n = \bar{\Gamma}_n^*, \) i.e., the closure of $\Lambda_n$ is the closure of $\Gamma_n^*$.

In other words, considering quasi-uniform distributions is sufficient for characterizing $\Gamma_n^*$. These are the distributions we will henceforth focus on.
Some quasi-uniform distributions for $n = 2$ and $n = 3$. 
Determining all quasi-uniform distributions appears to be a hopelessly complicated combinatorial problem. We only know the solution for $N = 2$.

**Theorem 4** Let the $n$ random variables $X_1, \ldots, X_n$ of alphabet size $N = 2$ be quasi-uniform. Then, by a suitable choice of the origin, if

$$p_{X_1,\ldots,X_n}(x'_1, \ldots, x'_n) = p_{X_1,\ldots,X_n}(x''_1, \ldots, x''_n) \neq 0,$$

then

$$p_{X_1,\ldots,X_n}(x'_1 \oplus x''_1, \ldots, x'_n \oplus x''_n) \neq 0.$$

In other words, the support of the distribution forms a vector space.
Distributions from Lattices

Taking cue from this last result, consider the lattice

\[ x = Mz, \]

where \( x \in \mathcal{R}^n \) are points in the lattice, \( M \in \mathcal{R}^{n \times n} \) is the so-called lattice-generating matrix, and \( z \in \mathcal{Z}^n \) 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 \( \mathcal{L}(M) \).

**Definition 1 (Lattice-Generated Distribution)** A probability distribution over \( n \) random variables with alphabet size \( N \) each, will be called lattice-generated, if for some lattice \( \mathcal{L}(M) \), we have

\[ p_{X_N}(x_N) = c, \]

a constant, whenever \( x_N \in \{0, \ldots, N - 1\}^n \cap \mathcal{L}(M) \) and

\[ p_{X_N}(x_N) = 0, \]

otherwise.
Quasi-Uniform Lattice-Generated Distributions

Of course, we need to ask when a lattice-generated distribution is quasi-uniform.

**Lemma 1 (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.
The resulting entropies can now be constructed.

**Lemma 2 (Entropy Extraction)** Consider a lattice-generated distribution with period dividing $N$. Consider any collection of random variables $X_S$ and partition the rows of the lattice-generating matrix $M$ accordingly:

$$M = \begin{bmatrix} M_S \\ M_{S^c} \end{bmatrix},$$

where $M_S$ is $|S|$-by-$n$. Then the normalized entropy of $X_S$ is given by

$$h_\alpha = |S| - \frac{\log(\gcd(\text{all } |S|\text{-by-}|S| \text{ minors of } M_S)))}{\log N}.$$
Let $\Delta_n$ denote the space of entropy vectors obtained from lattice-generated quasi-uniform distributions.

**Theorem 5 (An Inner Region for Entropic Vectors)**

$$ \text{con}(\Delta_n) \subseteq \Gamma_n^* $$

where $\text{con}(\cdot)$ represents the convex closure. This region is furthermore a polytope.
It turns out that wlog we may take $M$ lower triangular:

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

In fact, for any integers $M_{ij}$ it is always possible to find a large enough integer $N$ and positive rational numbers $\gamma_{ij}$ 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

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

from which it follows:

$$h_1 = 1 - \gamma_{11} , \quad h_2 = 1 - \min(\gamma_{21}, \gamma_{22}) , \quad h_{12} = 2 - \gamma_{11} - \gamma_{22}$$
We still need to check the condition that $M$ generates a quasi-uniform distribution. This requires that

\[
NM^{-1} = \begin{bmatrix}
N^{1-\gamma_{11}} & 0 \\
-N^{1-\gamma_{11} - \gamma_{22} + \gamma_{21}} & N^{1-\gamma_{22}}
\end{bmatrix}
\]

have integer entries, which means that

\[
\gamma_{11} \leq 1 \ , \ \gamma_{22} \leq 1 \ , \ \gamma_{11} + \gamma_{22} \leq 1 + \gamma_{21}.
\]

It is straightforward to show that the convex hull is

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

which can be shown to be $\Gamma_2$. 
Three Random Variables

\[ M = \begin{bmatrix}
N^{\gamma_{11}} & 0 & 0 \\
N^{\gamma_{21}} & N^{\gamma_{22}} & 0 \\
N^{\gamma_{31}} & N^{\gamma_{32}} & N^{\gamma_{33}}
\end{bmatrix}. \]

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

\[ 0 \leq \gamma_{ij} \leq 1 \]
\[ \gamma_{ii} + \gamma_{jj} - \gamma_{ij} \leq 1 \]
\[ \gamma_{11} + \gamma_{22} + \gamma_{33} - \gamma_{21} - \gamma_{32} \leq 1 \]
Extracting the entropies

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

\[
M = \begin{bmatrix}
    N\gamma_{11} & 0 & 0 \\
    N\gamma_{21} & N\gamma_{22} & 0 \\
    N\gamma_{31} & N\gamma_{32} & N\gamma_{33}
\end{bmatrix}.
\]
Theorem 6

\[ \text{con} \Delta_3 = \Gamma_3^* \]

Proof: Clearly, \( \text{con} \Delta_3 \subseteq \Gamma_3^* \) since all vectors in \( \text{con} \Delta_3 \) are entropic. To prove the other direction consider the region defined by

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

Each column in the matrix can be obtained by a lattice-generated distribution. Therefore the region must be a subset of \( \text{con} \Delta_3 \).
Write the above matrix equation as

\[ h = \begin{bmatrix} A & a \end{bmatrix} \begin{bmatrix} k \\ \kappa_8 \end{bmatrix} = Ak + ak_8, \]

so

\[ A^{-1}h - A^{-1}a = k \geq 0. \]

Computing \( A^{-1} \) and \( A^{-1}ak_8 \), yields

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 1 & 0 & 1 & -1 & 1 \\
0 & -1 & 0 & 1 & 1 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & 1 & 1 & -1 & 0 \\
1 & 1 & 1 & -1 & -1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
h_1 \\
h_2 \\
h_3 \\
h_{12} \\
h_{23} \\
h_{31} \\
h_{123}
\end{bmatrix}
\geq
\begin{bmatrix}
0 \\
0 \\
0 \\
1 \\
1 \\
1 \\
-1
\end{bmatrix}
\]
The point is to show that for any entropic vector $h$ one can find a non-negative $k_8$ such that 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

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

\( MN^{-1} \) having integer entries yields

\[
\begin{align*}
\gamma_{ii} & \geq 0 \\
\gamma_{ii} + \gamma_{jj} - \gamma_{ij} & \leq 1 \\
\gamma_{ii} + \gamma_{jj} + \gamma_{kk} - \gamma_{ij} - \gamma_{jk} & \leq 1 \\
\gamma_{11} + \gamma_{22} + \gamma_{33} + \gamma_{44} - \gamma_{21} - \gamma_{32} - \gamma_{43} & \leq 1
\end{align*}
\]
\[ h_1 = 1 - \gamma_{11} \]
\[ h_2 = 1 - \min(\gamma_{21}, \gamma_{22}) \]
\[ h_3 = 1 - \min(\gamma_{31}, \gamma_{32}, \gamma_{33}) \]
\[ h_4 = 1 - \min(\gamma_{41}, \gamma_{42}, \gamma_{43}, \gamma_{44}) \]
\[ h_{12} = 2 - \gamma_{11} - \gamma_{22} \]
\[ h_{13} = 2 - \gamma_{11} - \min(\gamma_{32}, \gamma_{33}) \]
\[ h_{14} = 2 - \gamma_{11} - \min(\gamma_{42}, \gamma_{43}, \gamma_{44}) \]
\[ h_{23} = 2 - \min(\gamma_{21} + \min(\gamma_{32}, \gamma_{33}), \gamma_{22} + \min(\gamma_{31} + \gamma_{33})) \]
\[ h_{24} = 2 - \min(\gamma_{21} + \min(\gamma_{42}, \gamma_{43}, \gamma_{44}), \gamma_{22} + \min(\gamma_{41}, \gamma_{43}, \gamma_{44})) \]
\[ h_{34} = 2 - \min(\gamma_{31} + \min(\gamma_{42}, \gamma_{43}, \gamma_{44}), \gamma_{32} + \min(\gamma_{41}, \gamma_{43}, \gamma_{44}), \gamma_{33} + \min(\gamma_{41}, \gamma_{42}, \gamma_{44})) \]
\[ h_{123} = 3 - \gamma_{11} - \gamma_{22} - \gamma_{33} \]
\[ h_{124} = 3 - \gamma_{11} - \gamma_{22} - \min(\gamma_{43}, \gamma_{44}) \]

etc.
Remarks on the Construction

- Our construction yields a polytope inner bound which allows one to obtain achievability results for networks via linear programming.
- It is tight for $n = 2, 3$.
- For $n > 3$ it includes all scalar and vector linear network codes.
- But is it tight?

To answer this we need to look at the connection to group theory...
Entropy and Groups

Consider a finite group $G$ and let $G_1, \ldots, G_n$ be subgroups of $G$. For any $S \subseteq \{1, \ldots, n\}$ define

$$v_S = \log \frac{|G|}{| \cap_{\alpha \in S} G_{\alpha}|}.$$

We will call the resulting $2^n - 1$ dimensional vector derived from all such $S$ as group-derived.

**Theorem 7 (Chan and Yeung)** Let the space of all group-derived vectors be denoted by $\mathcal{G}_n$. Then

$$\tilde{\mathcal{G}}_n = \Gamma^*_n.$$
Lattice Construction and Groups

Recall that our construction was based on the period \( N \) lattice

\[ x = Mz, \]

If we take each lattice point in \( \{0, 1, \ldots, N - 1\}^n \) to be an element of \( G \) and the operation to be addition modulo \( N \), then \( G \) is an Abelian group.

Let \( G_i, i = 1, \ldots, n \) be the subgroup obtained by insisting that the \( i \)-th entry of \( x \) be zero. Then it is not too hard to see that:

\[ v_S = h_S, \quad \forall S \subseteq \{1, \ldots, n\}. \]

**Theorem 8 (Chan)** If \( G \) is an Abelian group, then the resulting entropy vectors satisfy the Ingleton bound

\[ h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl} \geq h_{ijk} + h_{ijl} + h_{kl} + h_i + h_j. \]

But it is known that entropy can violate the Ingleton bound (more on this in a moment) and so our construction cannot be tight!
Entropy Vectors for Continuous Random Variables

Let $X_i \in \mathcal{R}^m$, $i = 1, \ldots, n$ be vector-valued continuous random variables. The normalized entropy is now defined as

$$h_S = \frac{1}{m} H(X_i, i \in S),$$

and the space of normalized entropic vectors denoted by $\Omega_n^*$. 

**Theorem 9 (Chan)** Let

$$\sum_{\alpha \in \{1, \ldots, n\}} k_{\alpha} h_\alpha \geq 0,$$

be an inequality for continuous random variables. Then

$$\sum_{\alpha \in \{1, \ldots, n\}} k_{\alpha} h_\alpha + \sum_{i=1}^{n} r_i (h_{i,c} - h_{i,c}) \geq 0,$$

for any $r_i \geq 0$ is an inequality for discrete random variables. Conversely any inequality for discrete random variables must be of this form.
Example

This implies that it is sufficient (and perhaps simpler) to study continuous random variables.

For example, take $n = 2$. The only inequality in the continuous case is

$$h_1 + h_2 - h_{12} \geq 0.$$ 

Thus, the inequalities for the discrete case are

$$h_1 + h_2 - h_{12} + r_1(h_{12} - h_2) + r_2(h_{12} - h_1) \geq 0,$$

for any $r_1, r_2 \geq 0$. For example:

- $r_1 = 1, r_2 = 0 : h_1 \geq 0$
- $r_1 = 0, r_2 = 1 : h_2 \geq 0$
- $r_1 = 1, r_2 \to \infty : h_{12} \geq h_1$
- $r_1 \to \infty, r_2 = 1 : h_{12} \geq h_2$
The most obvious class of continuous random variables to consider are Gaussians.

In this case, we start with a $nm \times nm$ positive definite covariance matrix $R$. Let $R_S$ be the principal minor determined by the rows and columns in set $S$. Then we have

$$h_S = \frac{1}{m} \log \det R_S.$$ 

Thus, the study of entropy leads us to the study of determinant inequalities. This is a subject with a long history.
Determinantal Inequalities

- Hadamard Inequality

\[ \det R_{11} \det R_{22} \geq \det \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}. \]

- Koteljanskii Inequality

\[ \det R_\alpha \det R_\beta \geq \det R_{\alpha \cup \beta} \det R_{\alpha \cap \beta}. \]
Some Observations

- For $n = 2, 3$ it can be shown that Gaussians achieve the full entropy region.
- For $n = 4$ Gaussians can violate the Ingleton bound:

$$R = \begin{bmatrix}
1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} & 1 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 1 
\end{bmatrix}$$

Just check!
Cayley’s Hyperdeterminant

In general, any $n \times n$ matrix has $2^n - 1$ principal minors. Very recently, (Oct 2007) Holtz and Sturmfels have noted that these $2^n - 1$ values satisfy Cayley’s hyperdeterminant formula.

- To be precise, for $n = 3$, the principal minors $p_1, p_2, p_3, p_{12}, p_{23}, p_{31}, p_{123}$ satisfy

$$ (p_{123} - p_1 p_{23} - p_2 p_{31} - p_3 p_{12} + 2p_1 p_2 p_3)^2 = 4(p_1 p_2 - p_{12})(p_2 p_3 - p_{23})(p_3 p_1 - p_{31}). $$
This is equivalent to the condition that the following 6 nonlinear equations have nonzero solutions $x_0, x_1, y_0, y_1, z_0, z_1$:

\[
\begin{align*}
    x_0y_0 + p_1x_1y_0 + p_2x_0y_1 + p_{12}x_1y_1 &= 0 \\
    p_3x_0y_0 + p_{31}x_1y_0 + p_{23}x_0y_1 + p_{123}x_1y_1 &= 0 \\
    y_0z_0 + p_2y_1z_0 + p_3y_0z_1 + p_{23}y_1z_1 &= 0 \\
    p_1y_0z_0 + p_{12}y_1z_0 + p_{31}y_0z_1 + p_{123}y_1z_1 &= 0 \\
    z_0x_0 + p_3z_1x_0 + p_1z_0x_1 + p_{31}z_1x_1 &= 0 \\
    p_2z_0x_0 + p_{23}z_1x_0 + p_{12}z_0x_1 + p_{123}z_1x_1 &= 0
\end{align*}
\]

For $n > 3$ things get increasingly more complicated.
**Conclusion**

- 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.
- Showed how cut-set bounds follow from convex-optimization duality.
- Argued that for wired networks, the optimization is over the unconstrained $\Gamma_n^*$.
- Characterizing $\Gamma_n^*$ for $n \geq 4$ is a fundamental open problem.
- Constructed an inner bound on $\Gamma_n^*$ using lattice-based distributions. Inner bound is a polytope allowing for efficient use via linear programming. Tight for $n = 2, 3$, but not for $n > 4$, though it includes all scalar and vector linear network codes.
- Studied connections to matroids, groups, Gaussians and determinantal inequalities.