# Applied Vector Optimization: Hunt for New Entropy Inequalities

Laszlo Csirmaz

Central European University, Budapest

Glasgow, July 14, 2015

# Outline

# What is entropy?

**Entropy** is a measure of information content.

Originating from physics, Claude Shannon made it the central notion in **information theory** in late 1940's.

$\vec{X}$ is a collection of discrete random variables taking $k$ possible configurations with probability

$$p_1, p_2, \ldots, p_k \geq 0, \qquad \text{where } p_1 + \cdots + p_k = 1.$$

The **entropy** of the collection $\vec{X}$ in *bits* is defined as

$$\boldsymbol{H}(\vec{X}) \stackrel{\text{def}}{=} \sum_{i=1}^{k} -p_i \log_2 p_i .$$

This is just right: random coin flipping has entropy 1 bit.

# The entropy vector

$\vec{X}$ is a collection of $n$ jointly distributed random variables. For each subset $A$ of $\{1, 2, \ldots, n\}$, $\boldsymbol{H}(A)$ denotes the entropy of the *marginal distribution* $\langle x_i : i \in A \rangle$.

The **entropy vector** of $\vec{X}$ is the $2^n - 1$ dimensional vector $\langle \boldsymbol{H}(A) \rangle$ indexed by the non-empty subsets $A$.

# The entropy vector

$\vec{X}$ is a collection of $n$ jointly distributed random variables. For each subset $A$ of $\{1, 2, \ldots, n\}$, $\boldsymbol{H}(A)$ denotes the entropy of the *marginal distribution* $\langle x_i : i \in A \rangle$.

The **entropy vector** of $\vec{X}$ is the $2^n - 1$ dimensional vector $\langle \boldsymbol{H}(A) \rangle$ indexed by the non-empty subsets $A$.

## Analogy

Entropy behaves like an asset.

# Supporting facts for the analogy

1. Larger group has more entropy:

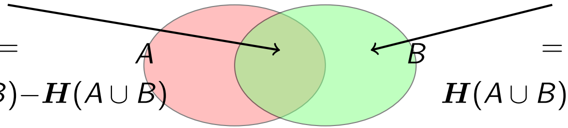
$$\text{if } A \subseteq B \text{ then } 0 \le \boldsymbol{H}(A) \le \boldsymbol{H}(B).$$

2. Independent information adds up:

$$\text{if } A \text{ and } B \text{ are independent, then } \boldsymbol{H}(A \cup B) = \boldsymbol{H}(A) + \boldsymbol{H}(B).$$

3. One can identify "private" and "joint" information:

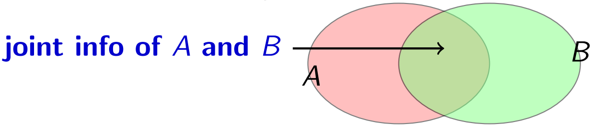
**joint info** $= \boldsymbol{H}(A)+\boldsymbol{H}(B)-\boldsymbol{H}(A \cup B)$



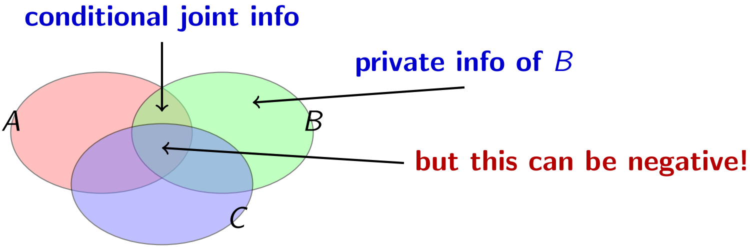

**private info** $= \boldsymbol{H}(A \cup B)-\boldsymbol{H}(A)$

# But the analogy breaks down ...

1. One cannot always "extract" the joint knowledge of $A$ and $B$ (no further random variable can be added which would act as their joint information)




2. With three subsets $A$, $B$, and $C$,




3. And many-many more subtle problems ...

# Outline

# Tools – Shannon (1949) and Zhang-Yeung (1998)

1. *Shannon inequalities*: 1) The private info is non-negative: if $A \subseteq B$ then $\boldsymbol{H}(A) \le \boldsymbol{H}(B)$.

   2) Conditional joint info is non-negative: for any three subsets

   $$\boldsymbol{H}(A \cup C) + \boldsymbol{H}(B \cup C) - \boldsymbol{H}(A \cup B \cup C) - \boldsymbol{H}(C) \ge 0.$$

   The minimal set has $n + n(n-1)2^{n-3}$ inequalities.

# Tools – Shannon (1949) and Zhang-Yeung (1998)

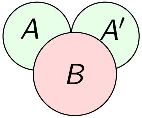
1. *Shannon inequalities*: 1) The private info is non-negative: if $A \subseteq B$ then $\boldsymbol{H}(A) \le \boldsymbol{H}(B)$.

   2) Conditional joint info is non-negative: for any three subsets

   $$\boldsymbol{H}(A \cup C) + \boldsymbol{H}(B \cup C) - \boldsymbol{H}(A \cup B \cup C) - \boldsymbol{H}(C) \ge 0.$$

   The minimal set has $n + n(n-1)2^{n-3}$ inequalities.

2. Creating *an independent copy* $A'$ of $A$ over $B$:

   a) $(A, B)$ and $(A', B)$ are identically distributed;

   b) $A$ and $A'$ are *independent* given $B$, that is

   $$\boldsymbol{H}(A \cup B) + \boldsymbol{H}(A' \cup B) - \boldsymbol{H}(A \cup A' \cup B) - \boldsymbol{H}(B) = 0.$$

# Tools – Shannon (1949) and Zhang-Yeung (1998)

1. *Shannon inequalities*: 1) The private info is non-negative: if $A \subseteq B$ then $\boldsymbol{H}(A) \le \boldsymbol{H}(B)$.

   2) Conditional joint info is non-negative: for any three subsets

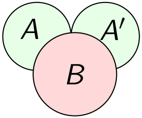
   $$\boldsymbol{H}(A \cup C) + \boldsymbol{H}(B \cup C) - \boldsymbol{H}(A \cup B \cup C) - \boldsymbol{H}(C) \ge 0.$$

   The minimal set has $n + n(n-1)2^{n-3}$ inequalities.

2. Creating *an independent copy* $A'$ of $A$ over $B$:

   a) $(A, B)$ and $(A', B)$ are identically distributed;

   b) $A$ and $A'$ are *independent* given $B$, that is




   $$\boldsymbol{H}(A \cup B) + \boldsymbol{H}(A' \cup B) - \boldsymbol{H}(A \cup A' \cup B) - \boldsymbol{H}(B) = 0.$$

3. Iterate step 2 by splitting the variable set again and again.

# An example

- Start with variables $a, b, c, d$, their entropy vector is $\mathbf{x} \in \mathbb{R}^{15}$.
- $(a', b')$ is a copy of $(a, b)$ over $(c, d)$. The entropy vector of $a, b, c, d, a', b'$ is $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{63}$.
- The $246 \times 64$ matrix $M$ gives all Shannon inequalities for the six variables $a, b, c, d, a', b'$:

$$M \cdot (\mathbf{x}, \mathbf{y}) \geq 0. \tag{1}$$

- The $13 \times 63$ matrix $A$ describes that $abcd$ and $a'b'cd$ are identical; and $ab$ and $a'b'$ are independent over $cd$:

$$A \cdot (\mathbf{x}, \mathbf{y}) = 0. \tag{2}$$

This comes from $\boldsymbol{H}(a) = \boldsymbol{H}(a')$, $\boldsymbol{H}(ac) = \boldsymbol{H}(a'c)$, ..., and

$$\boldsymbol{H}(abcd) + \boldsymbol{H}(a'b'cd) - \boldsymbol{H}(aba'b'cd) - \boldsymbol{H}(cd) = 0.$$

- Consider the linear constraints in (1) and (2). Do they have any consequence on $\mathbf{x}$ beyond the Shannon inequalities?

# YES! – the Zhang–Yeung inequality

**YES!**

$$3\boldsymbol{H}(ac) + 3\boldsymbol{H}(ad) + 3\boldsymbol{H}(cd) + \boldsymbol{H}(bc) + \boldsymbol{H}(bd) -$$
$$- \boldsymbol{H}(a) - 2\boldsymbol{H}(c) - 2\boldsymbol{H}(d) - \boldsymbol{H}(ab) - 4\boldsymbol{H}(acd) - \boldsymbol{H}(bcd) \geq 0$$

and 12 other similar inequalities (by permuting $a, b, c, d$).

# A recipe for getting new entropy inequalities

1. Start with four variables and entropy vector $\mathbf{x} \in \mathbb{R}^{15}$.
2. In several steps create a copy of a subset of the variables over the remaining ones.
   The entropy vector of the final set of variables is $(\mathbf{x}, \mathbf{y})$.
3. Collect the Shannon inequalities for the final set of variables:

$$M \cdot (\mathbf{x}, \mathbf{y}) \ge 0. \tag{1}$$

4. Collect the equations which describe that copied variables have identical distribution, and are conditionally independent:

$$A \cdot (\mathbf{x}, \mathbf{y}) = 0. \tag{2}$$

5. Project the convex polytope determined by the linear constrains (1) and (2) to the first 15 coordinate to get the polytope $\mathcal{Q}$.
6. Determine the facets of $\mathcal{Q}$ different from the (15 dimensional) Shannon inequalities: the equation of these facets give the new entropy inequalities.

# Outline

# Revisiting the recipe


1. Take the dual view: we need the *vertices* of the projection.
2. Get rid of the homogeneity by using a well-chosen cross-section (and reduce the problem to 14 dimensions).
3. Take a new "smart" coordinate system in $\mathbb{R}^{14}$ such that
   a) the cross-section is in the non-negative orthant of $\mathbb{R}^{14}$;
   b) if $\mathbf{x} \le \mathbf{x}'$, $\mathbf{x} \in \mathbb{R}^{14}$ is in the cross-section then so is $x'$.

**The vertices are just the extremal points of the Pareto front.**

# A **revised** recipe for getting new entropy inequalities

- From the description of the copy steps generate the matrix $M$ and the vector $\mathbf{b}$ such that

$$\mathcal{P} = \{(\mathbf{x},\mathbf{y}) : \mathbf{x} \ge 0, \mathbf{y} \ge 0, M\cdot(\mathbf{x},\mathbf{y}) = \mathbf{b}\,\}$$

  defines the set of *feasible solutions*.

- Use *vector optimization* to find all extremal solutions of the following **linear vector optimization problem:**

$$\text{solve } \min_{\mathbf{y}}\{\mathbf{x}\in\mathbb{R}^{14} : (\mathbf{x},\mathbf{y})\in\mathcal{P}\,\}.$$

- The extremal solutions yield the minimal independent set of new entropy inequalities which generate any other inequality derivable from the same set of copy steps.

The *objective space* can be reduced to $\mathbb{R}^{10}$: I can *prove* that in every extremal solution the last four coordinates must be zero.

# Outline

# The Vertex Separation Oracle

# The Vertex Separation Oracle for Vector Optimization

Let $\mathcal{Q}$ be the convex hull of all extremal solutions of the **linear vector optimization problem**

$$\text{solve } \min_{\mathbf{y}}\{\mathbf{x} \in \mathbb{R}^{10} : (\mathbf{x}, \mathbf{y}) \in \mathcal{P}\}.$$

$\mathcal{Q}$ can be reached by inquiring the

**Vertex Separation Oracle $\boldsymbol{VSO}$**

Q: (the equation of) a closed halfspace $H \subseteq \mathbb{R}^{10}$.
A: if $\mathcal{Q} \subseteq H$ then the answer is **inside**,
if $\mathcal{A} \not\subseteq H$, then the answer is a **vertex** of $\mathcal{Q}$ not in $H$.

The $\boldsymbol{VSO}$ can be implemented by returning the lexicographically minimal solution in $\mathbf{x}$ of the **scalar LP**

$$\text{solve } \min_{\mathbf{x},\mathbf{y}}\{\mathbf{h} \cdot \mathbf{x} : (\mathbf{x}, \mathbf{y}) \in \mathcal{P}\},$$

where the halfspace has equation $\mathbf{h} \cdot \mathbf{x} \geq c$.

# Inner approximation using double description

## Double Description method for vertex enumeration with $\boldsymbol{VSO}$

To enumerate the vertices of $\mathcal{Q}$ generate the *approximating sequence* $\mathcal{Q}_1 \subseteq \mathcal{Q}_2 \subseteq \cdots \subseteq \mathcal{Q}$ maintaining in each step

1. all vertices and facets of $\mathcal{Q}_j$,
2. for each *facet* of $\mathcal{Q}_j$ whether it is know to be a facet of $\mathcal{Q}$.

To get $\mathcal{Q}_{j+1}$ form $\mathcal{Q}_j$ pick a facet $f$ of $\mathcal{Q}_j$ which is not known to be a facet of $\mathcal{Q}$. Call the $\boldsymbol{VSO}$ with the half-space $f \geq 0$.

1. If the answer is **inside**, mark $f$ as a facet of $\mathcal{Q}$, and continue.
2. Otherwise let $\mathcal{Q}_{j+1}$ be the convex hull of $\mathcal{Q}_j$ and the vertex returned.

Stop when all facets of $\mathcal{Q}_j$ are facets of $\mathcal{Q}$: **you are done!**

$\mathcal{Q}$ has "ideal" vertices along the positive direction of the coordinate axes – these vertices plus any internal point can serve as the initial $\mathcal{Q}_1$ simplex approximation.

# Outline

# It works!

The algorithm was used successfully for combinatorial optimization problems with **ten** objectives. In the representative results $n$ is the dimension of $\mathbf{y}$, and $m$ is the number of rows in the matrix $M$:

| $m$ × $n$ | Vertices | Facets | Running time |
|---|---|---|---|
| 4055×370 | 19 | 58 | 1:10:10 |
| 4009×370 | 40 | 103 | 3:24:37 |
| 3891×358 | 30 | 102 | 3:34:31 |
| 3963×362 | 167 | 235 | 9:20:19 |
| 4007×370 | 318 | 356 | 13:20:08 |
| 4007×370 | 318 | 356 | 14:34:42 |
| 4007×370 | 297 | 648 | 22:02:39 |
| 3913×362 | 779 | 1269 | 37:15:33 |
| 3987×362 | 4510 | 7966 | 427:43:30 |
| 3893×362 | 10387 | 13397 | 716:36:32 |

Altogether over 400 new entropy inequalities were obtained.

# Conclusions

- Vector optimization approach was used successfully solving ten dimensional vector optimization problems with about 400 dimensional problem space and 4000 constraints.
- A new optimization paradigm has been identified: the objective is defined indirectly by **separation oracle**: When inquired whether can an existing solution be improved in a certain direction, it answers either **no**, or gives an **extremal solution** improving along the given direction. A dual version of Benson's algorithm is proposed solving an optimization problem given by a Vertex Separation Oracle.
- All problems are highly degenerate which required special attention on numerical stability.
- More background work is required for obtaining entropy inequalities involving **five** random variables. The objective space has 27 dimensions, which is beyond the reach of the present technique.

# Thank you for your attention