Parallel optimization on the Entropic Cone

Benoît Legat Raphaël M. Jungers

Université catholique de Louvain (UCL)

Shannon's Entropy

Univariate
Discrete random variable X taking values in $\mathcal{X}$.

Shannon's entropy:

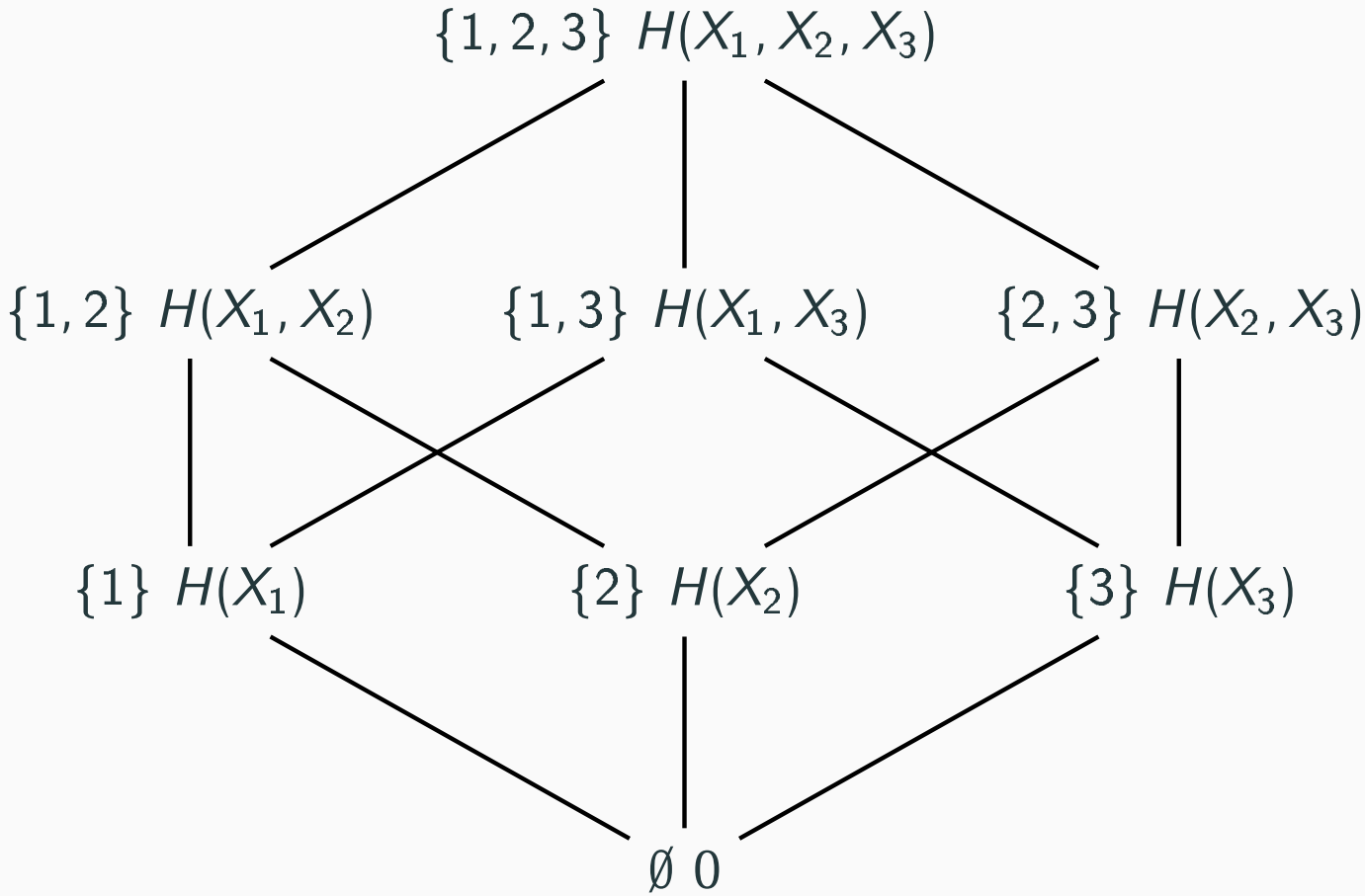
$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log \frac{1}{\Pr[X = x]}$$

Multivariate
Random variables $X = (X_1, \ldots, X_n)$.

Joint Shannon's entropy: Given $J \subseteq [n] = \{1, \ldots, n\}$,

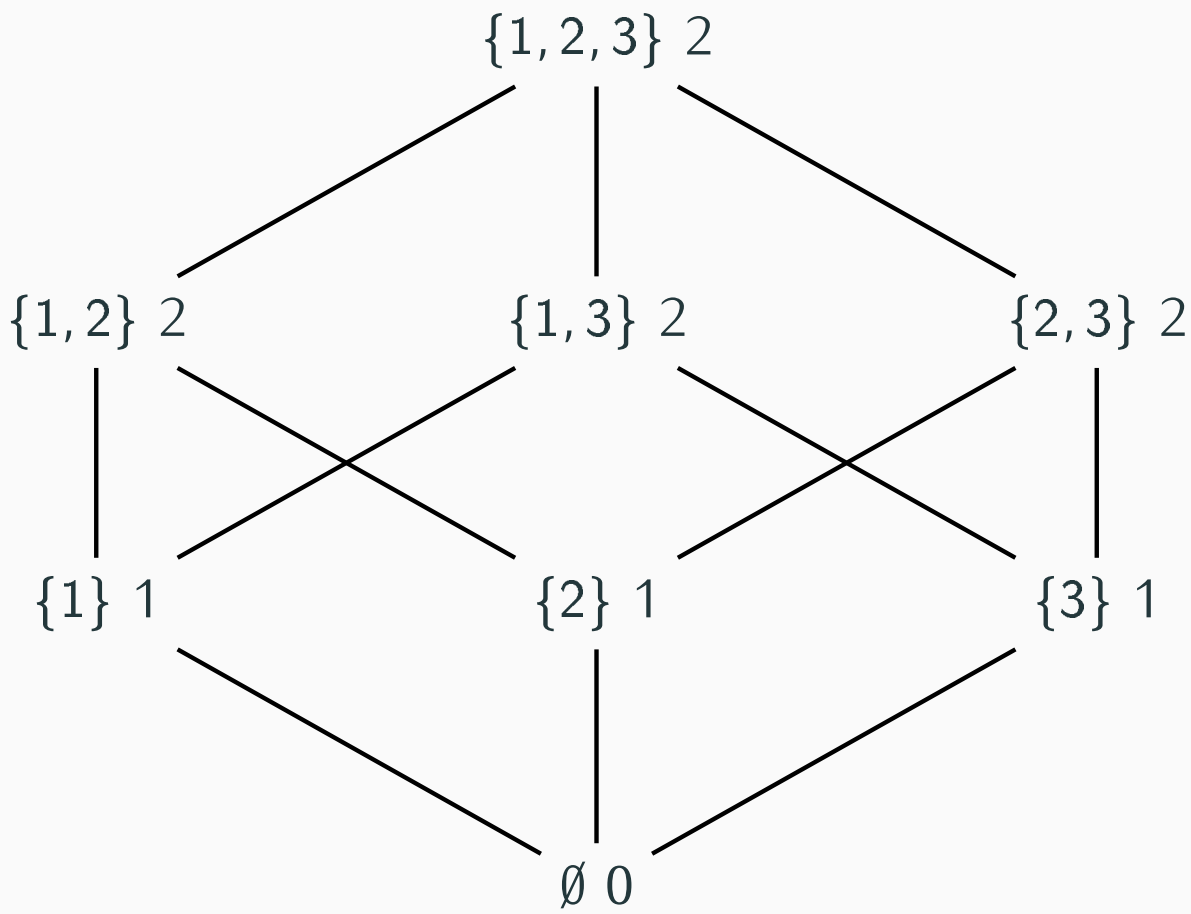
$$h_J \triangleq H(X_J) = \sum_{x \in \mathcal{X}_J} \Pr[X_J = x] \log \frac{1}{\Pr[X_J = x]}.$$

Entropy Vector



Example: One Time Pad

$$\text{plaintext} \oplus \text{key} = \text{ciphertext}.$$



Shannon's inequality

Normalization:

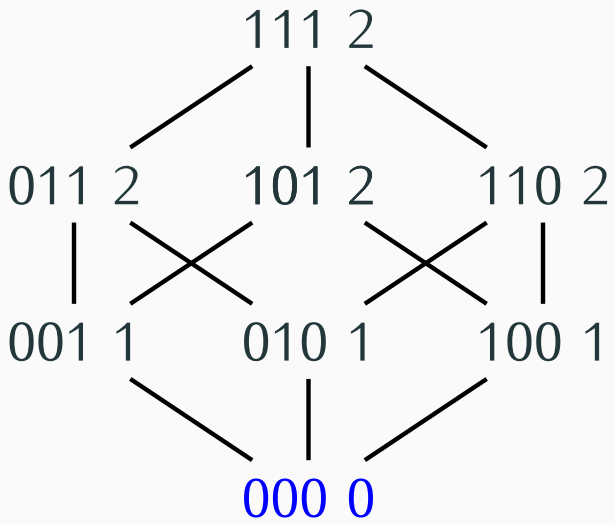
$$h_\emptyset = 0$$

Nondecreasing: If $I \subseteq J$ then

$$h_I \leq h_J$$

Submodular: For any J, K,

$$h_J + h_K \geq h_{J \cup K} + h_{J \cap K}$$



blue $= 0$

Shannon's inequality

Normalization:

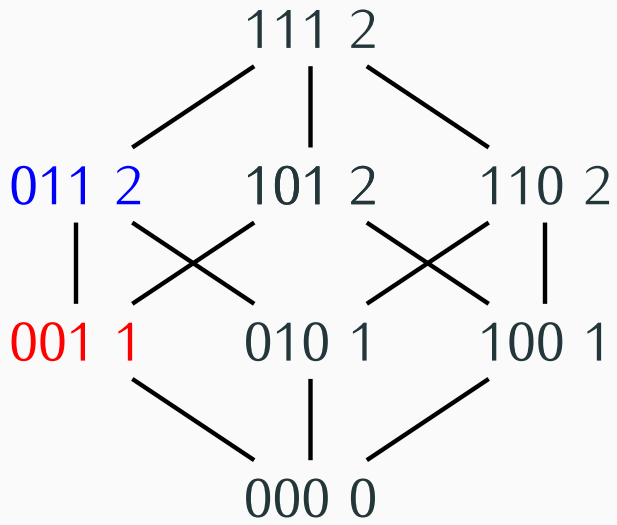
$$h_\emptyset = 0$$

Nondecreasing: If $I \subseteq J$ then

$$h_I \leq h_J$$

Submodular: For any J, K,

$$h_J + h_K \geq h_{J \cup K} + h_{J \cap K}$$



blue $\geq$ red

Shannon's inequality

Normalization:

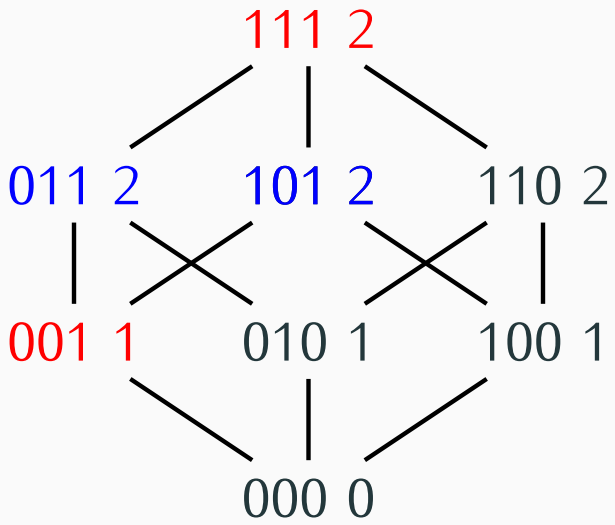
$$h_\emptyset = 0$$

Nondecreasing: If $I \subseteq J$ then

$$h_I \leq h_J$$

Submodular: For any J, K,

$$h_J + h_K \geq h_{J \cup K} + h_{J \cap K}$$



blue $\geq$ red

Non-Shannon inequalities

Entropic cone $\mathcal{H}_n$

Set of vectors h that are the entropy vector of n jointly distributed random variables.

[Zhang, Yeung 1997]: $\operatorname{cl}\mathcal{H}_n$ is a cone.

Shannon cone/polymatroid cone $\mathcal{P}_n$

Set of vectors h that satisfies all shannon's inequalities.

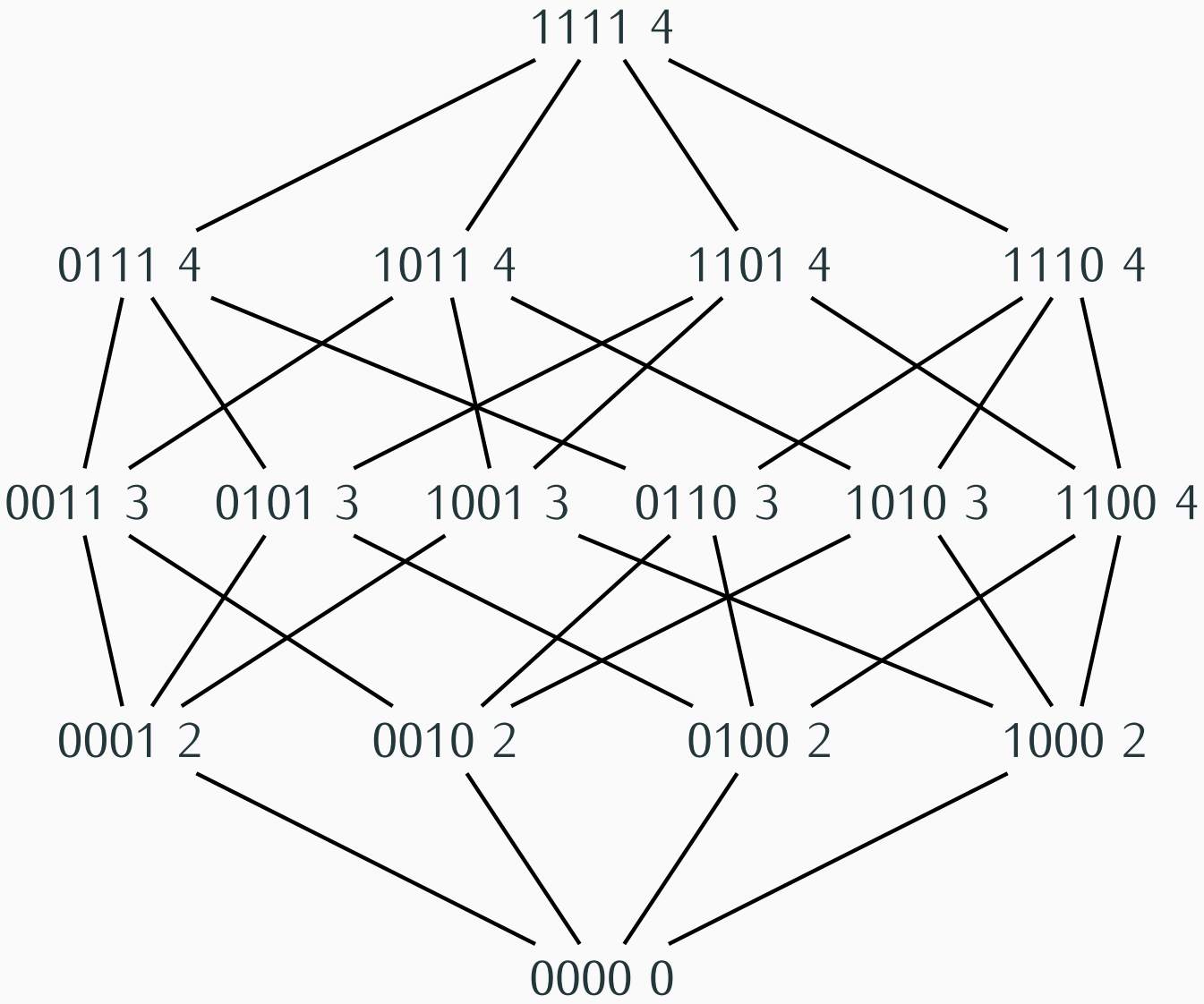
Non-Shannon inequalities

Inequalities valid for $\operatorname{cl}\mathcal{H}_n$ but not for $\mathcal{P}_n$.

Does $\operatorname{cl}\mathcal{H}_n = \mathcal{P}_n$ or are there non-Shannon inequalities ?

First non-Shannon inequalities [Zhang, Yeung 1998]

Non-entropic vector satisfying Shannon's inequalities:



How many non-Shannon inequalities ?

[Matúš 2007]
The cone $\operatorname{cl}\mathcal{H}_n$ is not polyhedral.

Polyhedral convex set: finite intersection of half-spaces.

$$\bigcap \{\, x \mid \langle a, x \rangle \leq \beta \,\}$$

[Mejia and Montoya 2015]
The cone $\operatorname{cl}\mathcal{H}_n$ is not semialgebraic.

Algebraic convex set: finite intersection of sets defined by a polynomial inequality

$$\bigcap \{\, x \mid p(x) \leq \beta \,\}$$

Semialgebraic convex set: projection of algebraic set.

Generating non-Shannon inequalities : Inner-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any J, K, $I = J \cap K$:

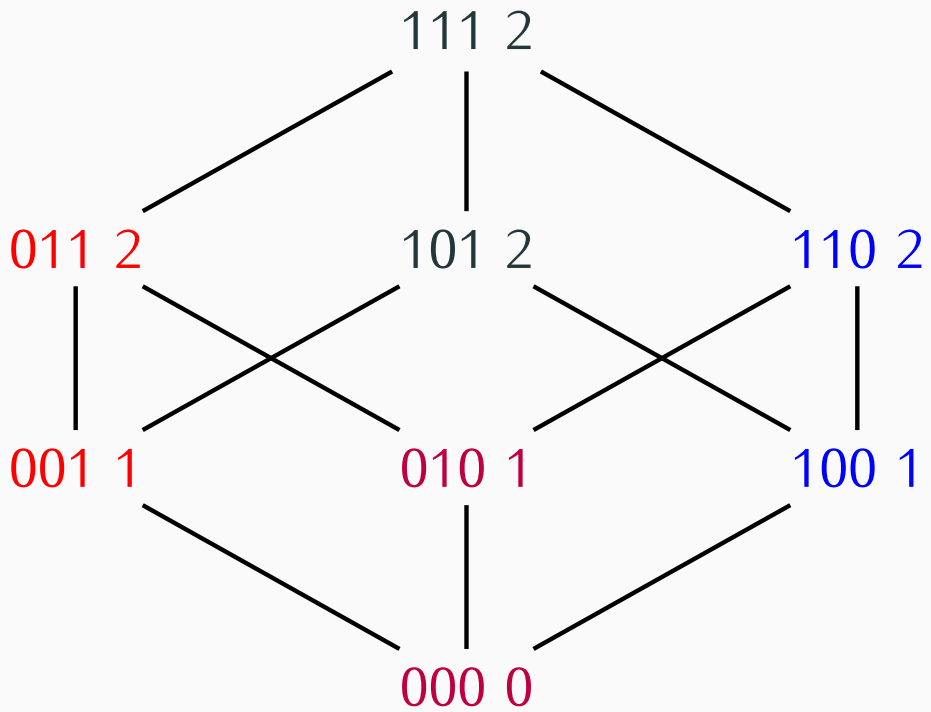
$$\frac{p_J(x_J)p_K(x_K)}{p_I(x_I)} = p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_I(x_I)$$

Generating non-Shannon inequalities : Inner-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any J, K, $I = J \cap K$:

$$\frac{p_J(x_J)p_K(x_K)}{p_I(x_I)} = p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_I(x_I)$$

Let's try it with $J = 011$ and $K = 110$ (so $I = 010$).

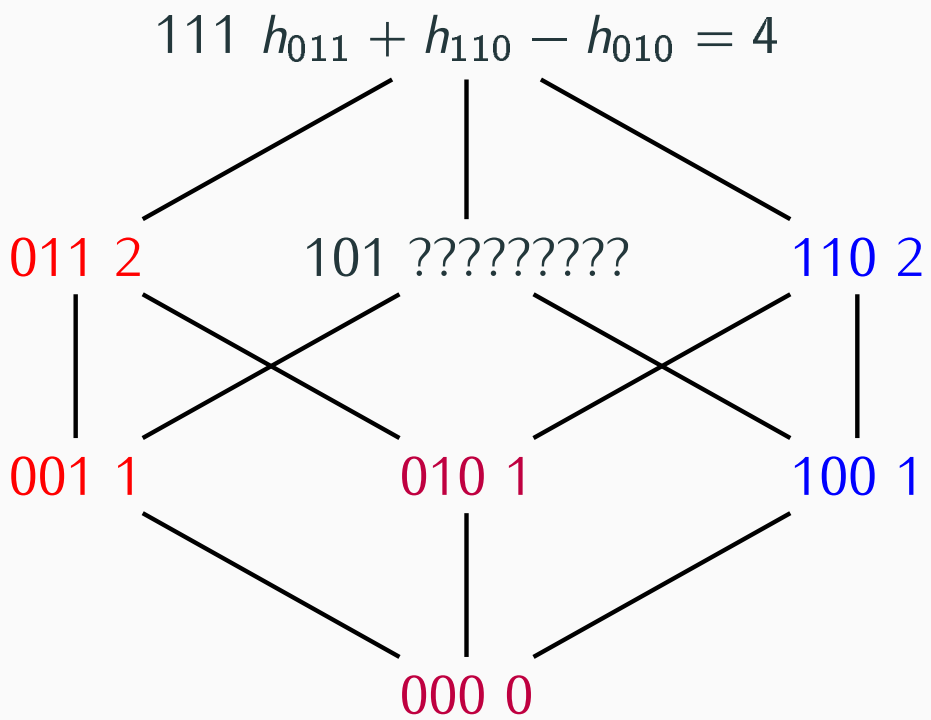


Generating non-Shannon inequalities : Inner-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any J, K, $I = J \cap K$:

$$\frac{p_J(x_J)p_K(x_K)}{p_I(x_I)} = p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_I(x_I)$$

Let's try it with $J = 011$ and $K = 110$ (so $I = 010$).



Generating non-Shannon inequalities : Self-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any $I \subseteq J$,

$$\frac{p_{J'}(x_{J'})p_J(x_K)}{p_I(x_I)} = p_{J'|I}(x_{J'}|x_I)p_{J|I}(x_K|x_I)p_I(x_I),$$

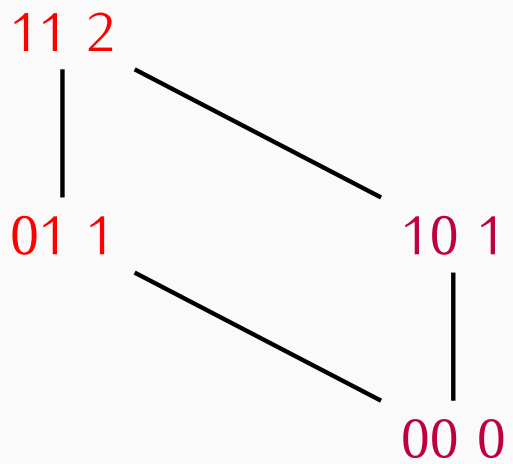
where $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus [n]) \cup I$.

Generating non-Shannon inequalities : Self-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any $I \subseteq J$,

$$\frac{p_{J'}(x_{J'})p_J(x_K)}{p_I(x_I)} = p_{J'|I}(x_{J'}|x_I)p_{J|I}(x_K|x_I)p_I(x_I),$$

where $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus [n]) \cup I$. Let's try it with $J = 11$ and $I = 10$ (so $n' = 3$, $J' = 011$, $K = 110$).

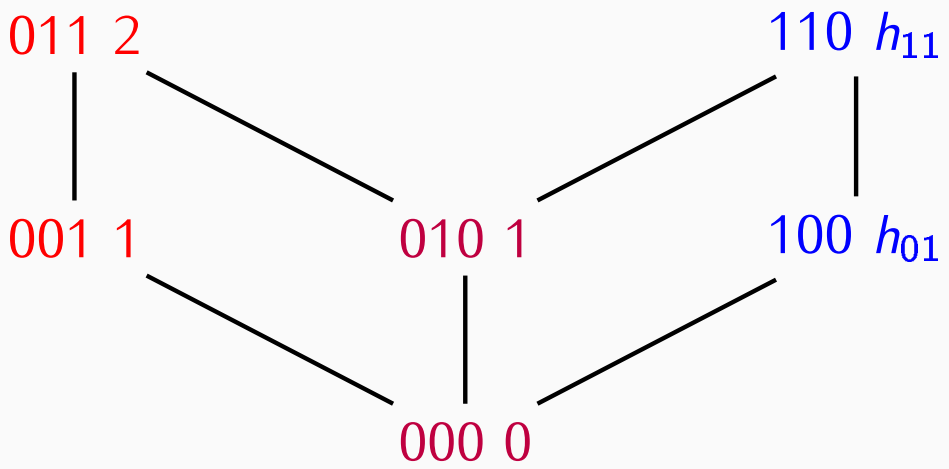


Generating non-Shannon inequalities : Self-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any $I \subseteq J$,

$$\frac{p_{J'}(x_{J'})p_J(x_K)}{p_I(x_I)} = p_{J'|I}(x_{J'}|x_I)p_{J|I}(x_K|x_I)p_I(x_I),$$

where $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus [n]) \cup I$. Let's try it with $J = 11$ and $I = 10$ (so $n' = 3$, $J' = 011$, $K = 110$).

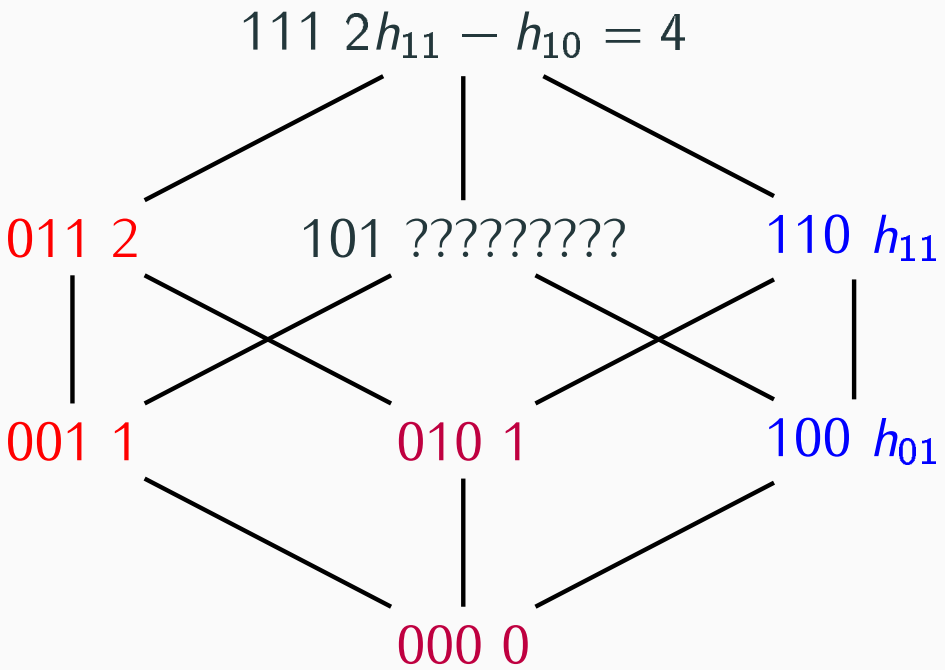


Generating non-Shannon inequalities : Self-Adhesivity

Given n jointly distributed random variables of probability mass function p, we can generate the following new probability mass function for any $I \subseteq J$,

$$\frac{p_{J'}(x_{J'})p_J(x_K)}{p_I(x_I)} = p_{J'|I}(x_{J'}|x_I)p_{J|I}(x_K|x_I)p_I(x_I),$$

where $n' = n + |J \setminus I|$, $J' = [n]$ and $K = ([n'] \setminus [n]) \cup I$. Let's try it with $J = 11$ and $I = 10$ (so $n' = 3$, $J' = 011$, $K = 110$).



Structure of the adhesivity invariance

If $h \in \mathcal{H}_n$ then there exists $g \in \mathcal{H}_{n'}$ such that $\forall I \subseteq L \subseteq J' \cup K$

$$g_L = h_{L \cap J'} + h_{L \cap K} - h_I$$

If $I \nsubseteq L$, g_L is *free* !

Structure of the adhesivity invariance

If $h \in \mathcal{H}_n$ then there exists $g \in \mathcal{H}_{n'}$ such that $\forall I \subseteq L \subseteq J' \cup K$

$$g_L = h_{L \cap J'} + h_{L \cap K} - h_I$$

If $I \nsubseteq L$, g_L is *free* !

To generate non-Shannon inequalities, we use the fact that these g_L *exists* and g is also entropic, so $g \in \mathcal{P}_{n'}$ and we can enforce adhesivity invariance on g too.

Structure of the adhesivity invariance

If $h \in \mathcal{H}_n$ then there exists $g \in \mathcal{H}_{n'}$ such that $\forall I \subseteq L \subseteq J' \cup K$

$$g_L = h_{L \cap J'} + h_{L \cap K} - h_I$$

If $I \nsubseteq L$, g_L is *free* !

To generate non-Shannon inequalities, we use the fact that these g_L *exists* and g is also entropic, so $g \in \mathcal{P}_{n'}$ and we can enforce adhesivity invariance on g too.

The adhesivity invariance is linear: Each adhesivity can be modeled by

$$\exists g \in \mathcal{H}_{n'} \text{ such that } Wg = Th$$

for some matrix W, T.

Stochastic Programming

Markov chain of linear programs. Linear program at state u:

$$\begin{aligned} Q(x, u) = \text{minimize } & c^T y + \mathcal{Q}_u(y) \\ \text{s.t. } & W_u y = h_u - T_u x, \\ & x \geq 0. \end{aligned}$$

where $\mathcal{Q}_u(y)$ is the sum of $Q(x, v)$ for each state v accessible from u weighted by the probability to go from state u to state v.

Stochastic Programming

Markov chain of linear programs. Linear program at state u:

$$\begin{aligned} Q(x, u) = \text{minimize } & c^T y + \mathcal{Q}_u(y) \\ \text{s.t. } & W_u y = h_u - T_u x, \\ & x \geq 0. \end{aligned}$$

where $\mathcal{Q}_u(y)$ is the sum of $Q(x, v)$ for each state v accessible from u weighted by the probability to go from state u to state v.

If it is infeasible, $Q(x, u) = +\infty$.

Stochastic Program for the Entropic Cone

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h + \mathcal{Q}_0(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}. \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathsf{a})) = \text{minimize } & \mathcal{Q}_{(n,\mathsf{a})}(g) \\ \text{s.t. } & g \in \mathsf{a}(h) \\ & g \in \mathcal{P}_{n'}. \end{aligned}$$

Stochastic Program for the Entropic Cone

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h + \mathcal{Q}_0(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}. \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathsf{a})) = \text{minimize } & \mathcal{Q}_{(n,\mathsf{a})}(g) \\ \text{s.t. } & g \in \mathsf{a}(h) \\ & g \in \mathcal{P}_{n'}. \end{aligned}$$

This state is accessible for all linear programs of dimension $2^n - 1$ so $\mathcal{Q}_{(n,\mathsf{a})}$ only depends on n'.

Stochastic Program for the Entropic Cone

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h + \mathcal{Q}_{n_0}(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}. \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathsf{a})) = \text{minimize } & \mathcal{Q}_{n'}(g) \\ \text{s.t. } & g \in \mathsf{a}(h) \\ & g \in \mathcal{P}_{n'}. \end{aligned}$$

This state is accessible for all linear programs of dimension $2^n - 1$ so $\mathcal{Q}_{(n,\mathsf{a})}$ only depends on n'.

Stochastic Program for the Entropic Cone

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h + \mathcal{Q}_{n_0}(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}. \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathsf{a})) = \text{minimize } & \mathcal{Q}_{n'}(g) \\ \text{s.t. } & g \in \mathsf{a}(h) \\ & g \in \mathcal{P}_{n'}. \end{aligned}$$

This state is accessible for all linear programs of dimension $2^n - 1$ so $\mathcal{Q}_{(n,\mathsf{a})}$ only depends on n'.

The function $\mathcal{Q}_n$ is either 0 or $+\infty$.

Stochastic Program for the Entropic Cone

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h \\ &\text{s.t. } h \in \mathcal{P}_{n_0} \\ &\quad h \in \operatorname{dom}(\mathcal{Q}_{n_0}). \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathsf{a})) = \text{minimize } &0 \\ \text{s.t. } &g \in \mathsf{a}(h) \\ &g \in \mathcal{P}_{n'} \\ &g \in \operatorname{dom}(\mathcal{Q}_{n'}). \end{aligned}$$

This state is accessible for all linear programs of dimension $2^n - 1$ so $\mathcal{Q}_{(n,\mathsf{a})}$ only depends on n'.

The function $\mathcal{Q}_n$ is either 0 or $+\infty$.

Polyhedral approximation of $\mathrm{dom}(\mathcal{Q}_n)$

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h \\ &\text{s.t. } h \in \mathcal{P}_{n_0} \\ &\quad h \in \mathrm{dom}(\mathcal{Q}_{n_0}). \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathrm{a})) = \text{minimize } &0 \\ \text{s.t. } &g \in \mathrm{a}(h) \\ &g \in \mathcal{P}_{n'} \\ &g \in \mathrm{dom}(\mathcal{Q}_{n'}). \end{aligned}$$

The domain of $\mathcal{Q}_n$ is approximated by the polyhedron

$$\{\, x \mid D_n x \leq d_n \,\}.$$

Polyhedral approximation of $\mathrm{dom}(\mathcal{Q}_n)$

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h \\ &\quad \text{s.t. } h \in \mathcal{P}_{n_0} \\ &\qquad D_{n_0} h \leq d_{n_0}. \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h, (n, \mathsf{a})) = \text{minimize } &0 \\ &\text{s.t. } g \in \mathsf{a}(h) \\ &\quad g \in \mathcal{P}_{n'} \\ &\quad D_{n'} g \leq d_{n'}. \end{aligned}$$

The domain of $\mathcal{Q}_n$ is approximated by the polyhedron

$$\{ x \mid D_n x \leq d_n \}.$$

Polyhedral approximation of $\mathrm{dom}(\mathcal{Q}_n)$

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h \\ &\quad \text{s.t. } h \in \mathcal{P}_{n_0} \\ &\quad\quad D_{n_0} h \leq d_{n_0}. \end{aligned}$$

Linear Program for each adhesivity a:

$$\begin{aligned} Q(h,(n,\mathsf{a})) = \text{minimize } &0 \\ &\text{s.t. } g \in \mathsf{a}(h) \\ &\quad g \in \mathcal{P}_{n'} \\ &\quad D_{n'} g \leq d_{n'}. \end{aligned}$$

The domain of $\mathcal{Q}_n$ is approximated by the polyhedron

$$\{\, x \mid D_n x \leq d_n \,\}.$$

Program at state (n, a) is infeasible $\rightarrow$ unbouded ray of the dual $\rightarrow$ new cut for D_n, d_n.

Stochastic Dual Dynamic Programming

for $k = 1, 2, \ldots, K$ **do**

Pick a set P of paths of length m starting at the initial state.

Solve the optimization program

$$\begin{aligned} &\text{minimize } c(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0} \cap \mathbb{F} \cap \text{dom}(\mathcal{Q}_0), \end{aligned}$$

for $i = 1, 2, \ldots, m$ **do**

for all $p \in P$ **do**

Solve $Q(h, (n_{p,i}, \mathrm{a}_{p,i}))$ where h is the optimal solution of the previous state in the path.

if the program is infeasible **then**

Add a feasibility cut for $\text{dom}(Q(\cdot, (n_{p,i}, \mathrm{a}_{p,i})))$

Remove p from P

end if

end for

end for

end for

Ingleton score

$$\mathbb{I}^* \triangleq \inf_{0 \neq h \in \mathcal{H}_4} \mathbb{I}_{ij}(h)$$

where $\mathbb{I}_{ij}(h) = \langle \square_{ij}, h \rangle / h_{[n]}$ and $\square_{ij}$ is the Ingleton dual entropy vector.

[Dougherty et al. 2011]

$$-0.15789 \leq \mathbb{I}^*.$$

[Matúš and Csirmaz 2013]

$$\mathbb{I}^* \leq -0.09243.$$

Attacking the Ingleton score

Only self-adhesivity, $n \leq 7$ and $m \leq 2$

$$-0.15854 \leq \mathbb{I}^*.$$

Attacking the Ingleton score

Only self-adhesivity, $n \leq 7$ and $m \leq 2$

$$-0.15854 \leq \mathbb{I}^*.$$

Only self-adhesivity, $n \leq 8$ and $m \leq 3$

$$-0.15789 \leq \mathbb{I}^*.$$

Attacking the Ingleton score

Only self-adhesivity, $n \leq 7$ and $m \leq 2$

$$-0.15854 \leq \mathbb{I}^*.$$

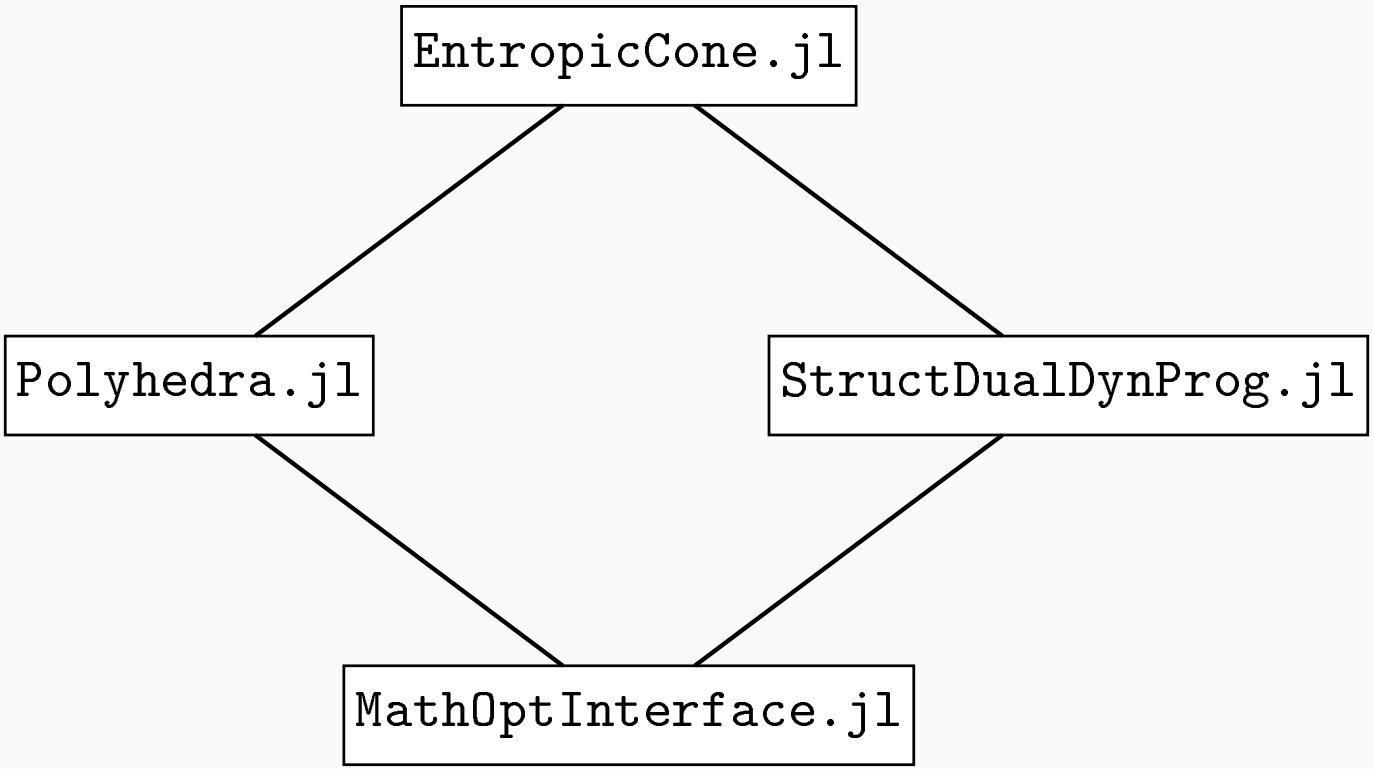
Only self-adhesivity, $n \leq 8$ and $m \leq 3$

$$-0.15789 \leq \mathbb{I}^*.$$

Only self-adhesivity, $n \leq 9$ and $m \leq 4$
Stay tuned.

`EntropicCone.jl` toolbox

`EntropicCone.jl` : Optimization on and Approximation of the Entropic Cone.



Conclusion

- Parallelizable optimization algorithm on the Entropic Cone.
- Possible to add nonlinear constraints and use nonlinear objective.
- Current best lower bound on $\mathbb{I}^*$ obtained in under a minute.