

# 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:

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

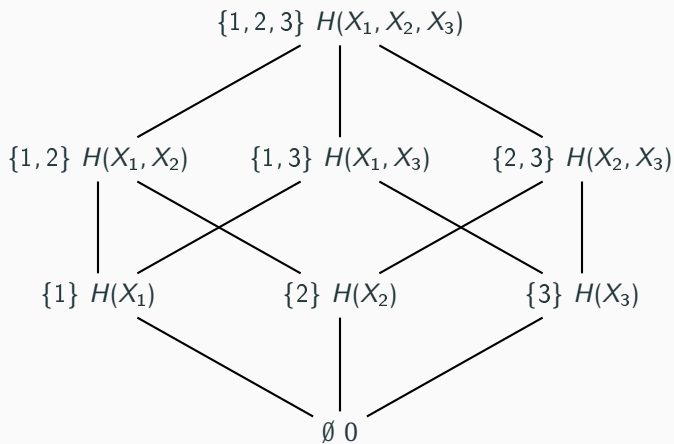
## Multivariate

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

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

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

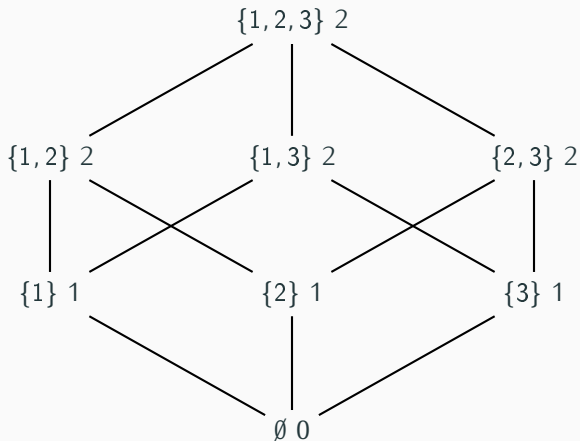
# Entropy Vector



## Example: One Time Pad

plaintext  $\oplus$  key = ciphertext.

$X_1$  = plaintext,      $X_2$  = key,      $X_3$  = ciphertext.



# Shannon's inequality

Normalization:

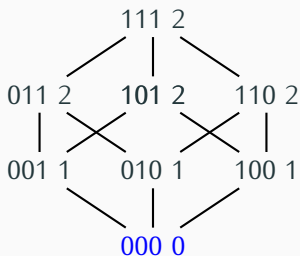
$$h_{\emptyset} = 0$$

Nondecreasing: If  $I \subseteq J$  then

$$h_I \leq h_J$$

Submodular: For any  $J, K$ ,

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



$$\text{blue} = 0$$

# Shannon's inequality

*Normalization:*

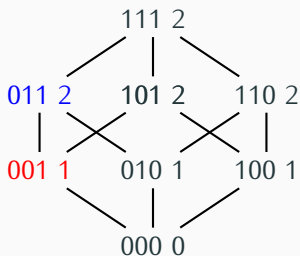
$$h_{\emptyset} = 0$$

**Nondecreasing:** If  $I \subseteq J$  then

$$h_I \leq h_J$$

*Submodular:* For any  $J, K$ ,

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



$$\text{blue} \geq \text{red}$$

# Shannon's inequality

*Normalization:*

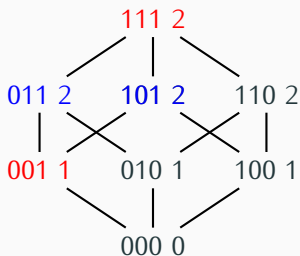
$$h_{\emptyset} = 0$$

*Nondecreasing:* If  $I \subseteq J$  then

$$h_I \leq h_J$$

**Submodular:** For any  $J, K$ ,

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



$$\text{blue} \geq \text{red}$$

# 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]:  $\text{cl } \mathcal{H}_n$  is a cone.

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

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

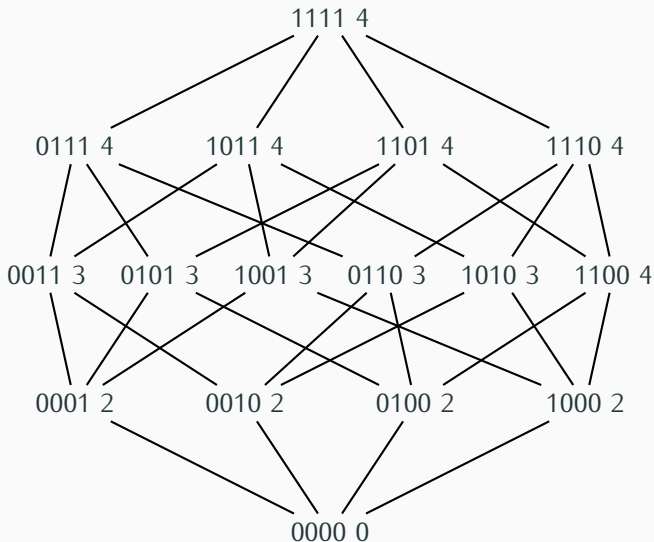
## Non-Shannon inequalities

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

Does  $\text{cl } \mathcal{H}_n = \mathcal{P}_n$  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  $\text{cl } \mathcal{H}_n$  is not polyhedral.

Polyhedral convex set: finite intersection of half-spaces.

$$\bigcap_{i=1}^m \{x \mid \langle a_i, x \rangle \leq \beta_i\}$$

[Mejia and Montoya 2015]

The cone  $\text{cl } \mathcal{H}_n$  is not semialgebraic.

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

$$\bigcup_{i=1}^m \bigcap_{j=1}^{J_i} \{x \mid p_j(x) \leq \beta_j\}$$

## 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  $J, K, I = J \cap K$ :

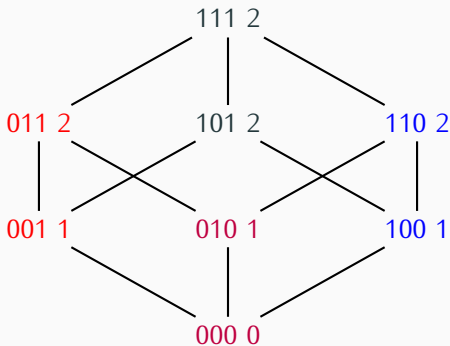
$$\frac{p_J(x_J)p_K(x_K)}{p_I(x_I)} = p_{J|I}(x_J|x_I)p_{K|I}(x_K|x_I)p_I(x_I)$$

## 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  $I = 010$ ).

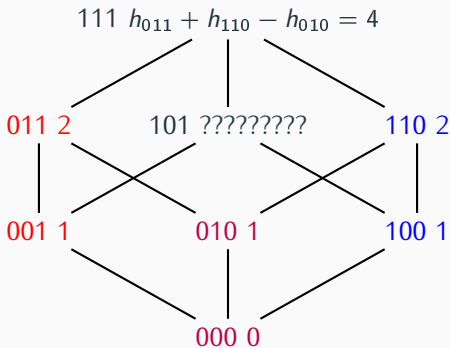


## Generating non-Shannon inequalities : Inner-Adhesivity

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  $I \subseteq J$ ,

$$\frac{p(x)p_J(x_K)}{p_I(x_I)} = p_{[n] \cup I}(x|x_I)p_{J \setminus I}(x_K|x_I)p_I(x_I),$$

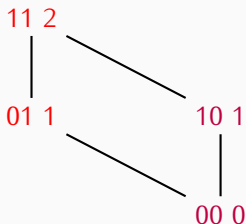
where  $n' = n + |J \setminus I|$  and  $K = ([n'] \setminus [n]) \cup I$ .

# 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  $I \subseteq J$ ,

$$\frac{p(x)p_J(x_K)}{p_I(x_I)} = p_{[n] \cup I}(x|x_I)p_{J \setminus I}(x_K|x_I)p_I(x_I),$$

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

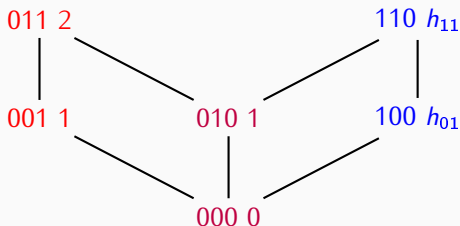


# Generating non-Shannon inequalities : Self-Adhesivity

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

$$\frac{p(x)p_J(x_K)}{p_I(x_I)} = p_{[n] \setminus I}(x| x_I)p_{J \setminus I}(x_K|x_I)p_I(x_I),$$

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

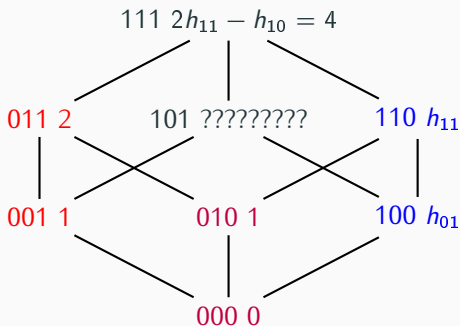


# Generating non-Shannon inequalities : Self-Adhesivity

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

$$\frac{p(x)p_J(x_K)}{p_I(x_I)} = p_{[n] \setminus I}(x| x_I)p_{J \setminus I}(x_K|x_I)p_I(x_I),$$

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



## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \quad \Rightarrow \quad g_L = h_L$$

## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \quad \Rightarrow \quad g_L = h_L$$

If  $L \subseteq K$ ,

$$q_L(x) = p_K(x) \quad \Rightarrow \quad g_L = h_L$$

## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \quad \Rightarrow \quad g_L = h_L$$

If  $L \subseteq K$ ,

$$q_L(x) = p_K(x) \quad \Rightarrow \quad g_L = h_L$$

If  $I \subseteq L \subseteq J \cup K$ ,

## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \Rightarrow g_L = h_L$$

If  $L \subseteq K$ ,

$$q_L(x) = p_K(x) \Rightarrow g_L = h_L$$

If  $I \subseteq L \subseteq J \cup K$ ,

$$q(x) = p_{L \cap J|I}(x_{L \cap J}|x_I)p_{L \cap K|I}(x_{L \cap K}|x_I)p_I(x_I) \Rightarrow g_L = h_{L \cap J} + h_{L \cap K} - h_I$$

## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \Rightarrow g_L = h_L$$

If  $L \subseteq K$ ,

$$q_L(x) = p_K(x) \Rightarrow g_L = h_L$$

If  $I \subseteq L \subseteq J \cup K$ ,

$$q(x) = p_{L \cap J|I}(x_{L \cap J}|x_I)p_{L \cap K|I}(x_{L \cap K}|x_I)p_I(x_I) \Rightarrow g_L = h_{L \cap J} + h_{L \cap K} - h_I$$

Otherwise,

## Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \Rightarrow g_L = h_L$$

If  $L \subseteq K$ ,

$$q_L(x) = p_K(x) \Rightarrow g_L = h_L$$

If  $I \subseteq L \subseteq J \cup K$ ,

$$q(x) = p_{L \cap J|I}(x_{L \cap J}|x_I)p_{L \cap K|I}(x_{L \cap K}|x_I)p_I(x_I) \Rightarrow g_L = h_{L \cap J} + h_{L \cap K} - h_I$$

Otherwise,  $g_L$  is free...

# Structure of the adhesivity invariance

Given  $n$  jointly distributed random variables of probability mass function  $p$ , consider  $J, K, I = J \cap K$  and

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

If  $L \subseteq J$ ,

$$q_L(x) = p_L(x) \Rightarrow g_L = h_L$$

If  $L \subseteq K$ ,

$$q_L(x) = p_K(x) \Rightarrow g_L = h_L$$

If  $I \subseteq L \subseteq J \cup K$ ,

$$q(x) = p_{L \cap J|I}(x_{L \cap J}|x_I)p_{L \cap K|I}(x_{L \cap K}|x_I)p_I(x_I) \Rightarrow g_L = h_{L \cap J} + h_{L \cap K} - h_I$$

Otherwise,  $g_L$  is free... But  $g \in \mathcal{H}_{|J \cup K|}$ !

# Structure of the adhesivity invariance

## Linearity

Each adhesivity can be modeled by

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

for some matrix  $W, T$ .

## Symmetry

if  $J', K', I'$  is a permutation of  $J, K, I$ , there is a permutation matrix  $P$  such that  $W', T' = WP, TP$ .

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

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

where  $Q_u(y)$  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  $v$ .

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

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

where  $Q_u(y)$  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:

$$\begin{aligned} &\text{minimize } c^T h + Q_0(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}. \end{aligned}$$

Linear Program for each adhesivity  $a$ :

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

# Stochastic Program for the Entropic Cone

Linear Program at initial state:

$$\begin{aligned} &\text{minimize } c^T h + Q_0(h) \\ &\text{s.t. } h \in \mathcal{P}_{n_0}. \end{aligned}$$

Linear Program for each adhesivity  $a$ :

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

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

# Stochastic Program for the Entropic Cone

Linear Program at initial state:

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

Linear Program for each adhesivity  $a$ :

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

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

# Stochastic Program for the Entropic Cone

Linear Program at initial state:

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

Linear Program for each adhesivity  $a$ :

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

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

The function  $Q_n$  is either 0 or  $+\infty$ .

# Stochastic Program for the Entropic Cone

Linear Program at initial state:

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

Linear Program for each adhesivity  $a$ :

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

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

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

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

Linear Program at initial state:

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

Linear Program for each adhesivity  $a$ :

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

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

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

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

Linear Program at initial state:

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

Linear Program for each adhesivity  $a$ :

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

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

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

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

Linear Program at initial state:

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

Linear Program for each adhesivity  $a$ :

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

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  $\rightarrow$  unbounded ray of the dual  $\rightarrow$  new cut for  $D_n, d_n$ .

# Stochastic Dual Dynamic Programming

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

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

Solve the optimization program

minimize  $c(h)$

s.t.  $h \in \mathcal{P}_{n_0} \cap \mathbb{F} \cap \text{dom}(\mathcal{Q}_0),$

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

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

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

**if** the program is infeasible **then**

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

Remove  $p$  from  $P$

**end if**

**end for**

**end for**

**end for**

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

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

## Challenges

Dimension of problems is  $2^n - 1$ :

- Memory efficient representation (symmetry helps a lot)
- Efficient cut pruning algorithm (symmetry generates a lot of constraints)

# Attacking the Ingleton score

## Challenges

Dimension of problems is  $2^n - 1$ :

- Memory efficient representation (symmetry helps a lot)
- Efficient cut pruning algorithm (symmetry generates a lot of constraints)

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

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

# Attacking the Ingleton score

## Challenges

Dimension of problems is  $2^n - 1$ :

- Memory efficient representation (symmetry helps a lot)
- Efficient cut pruning algorithm (symmetry generates a lot of constraints)

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

## Challenges

Dimension of problems is  $2^n - 1$ :

- Memory efficient representation (symmetry helps a lot)
- Efficient cut pruning algorithm (symmetry generates a lot of constraints)

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

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

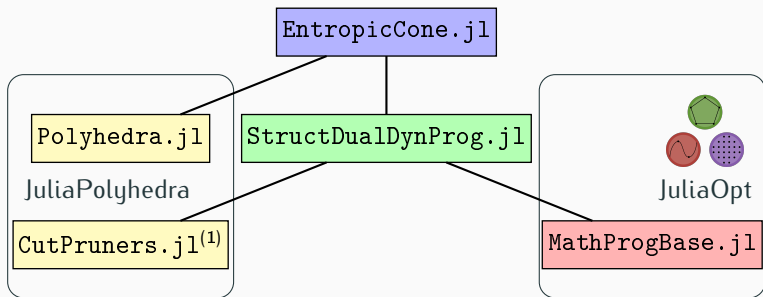
**Only self-adhesivity,  $n \leq 8$  and  $m \leq 3$**

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

**Only self-adhesivity,  $n \leq 9$  and  $m \leq 4$**

*Stay tuned.*

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



<sup>(1)</sup> Joint work with François Pacaud of ENPC

# 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.

## Challenges ahead

- Modifying problem in-place when adding constraints
- Parallelization
- Being combine each algorithm parts from all open source SDDP toolboxes under the same interface
- Testing and benchmarking of different methods
- Beating the best known lower bound