Extended Formulations and Information Theory

Sebastian Pokutta

Georgia Institute of Technology
ISyE / ARC

September 2014, MIP 2014

Disclaimer: Citations and References...

Extended formulations

Given a polytope $P \subseteq \mathbb{R}^n$, what is the best way of expressing P by means of linear inequalities?

Given a polytope $P \subseteq \mathbb{R}^n$, what is the best way of expressing P by means of linear inequalities?

We want the study the expressive power of linear and semidefinite programs.

Given a polytope $P \subseteq \mathbb{R}^n$, what is the best way of expressing P by means of linear inequalities?

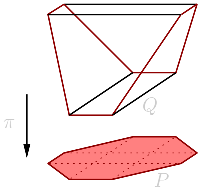
We want the study the expressive power of linear and semidefinite programs.

$\rightsquigarrow$ alternative measure of complexity independent of P vs. NP.

Definition (extension)

P, Q polytopes

Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

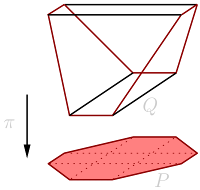
$\text{size}(Q) :=$ #facets of Q

$\text{xc}(P) := \min\{\text{size}(Q) \mid Q \text{ extension of } P\}$

Definition (extension)

P, Q polytopes

Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

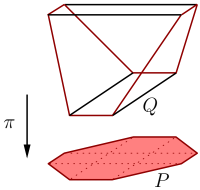
$\mathrm{size}(Q) :=$ #facets of Q

$\mathrm{xc}(P) := \min\{\mathrm{size}(Q) \mid Q$ extension of $P\}$

Definition (extension)

P, Q polytopes

Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

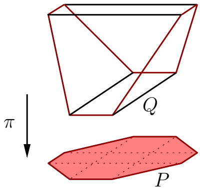
$\text{size}(Q) :=$ #facets of Q

$\text{xc}(P) := \min\{\text{size}(Q) \mid Q \text{ extension of } P\}$

Definition (extension)

P, Q polytopes

Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

$\mathrm{size}(Q) := \#$facets of Q
$\mathrm{xc}(P) := \min\{\mathrm{size}(Q) \mid Q \text{ extension of } P\}$

Why do we care for extended formulations?

Why do we care for extended formulations?

$\rightsquigarrow$ Quantifier elimination backwards.

Compact Extended Formulations.

Example: spanning tree polytope of $K_n = (V_n, E_n)$

Formulation 1:

Vars: x_{uv} $(uv \in E_n)$

$\sum_{uv \in E[U]} x_{uv} \leqslant |U| - 1 \; \forall U \neq \varnothing$
$x \geqslant 0$

$\sum_{uv \in E_n} x_{uv} = n - 1$

size $\approx 2^n$

Formulation 2:

Vars: x_{uv} $(uv \in E_n)$
$y_{\overrightarrow{uv},w}$ $(uv \in E_n,\, w \neq u, v)$

$x \geqslant 0$
$y \geqslant 0$

$x_{uv} - y_{\overrightarrow{uv},w} - y_{\overrightarrow{vu},w} = 0 \; \forall u, v, w$
$x_{uv} + \sum_{w \neq u,v} y_{\overrightarrow{uw},v} = 1 \; \forall u, v$
$\sum_{uv \in E_n} x_{uv} = n - 1$

size $\approx n^3 \rightarrow$ compact

Is there an EF with even fewer inequalities?

Compact Extended Formulations.

Example: spanning tree polytope of $K_n = (V_n, E_n)$

Formulation 1:

Vars: $x_{uv}\ (uv \in E_n)$

$\sum_{uv \in E[U]} x_{uv} \leqslant |U| - 1\ \forall U \neq \varnothing$
$x \geqslant 0$

$\sum_{uv \in E_n} x_{uv} = n - 1$

size $\approx 2^n$

Formulation 2:

Vars: $x_{uv}\ (uv \in E_n)$
$y_{\overrightarrow{uv},w}\ (uv \in E_n,\ w \neq u, v)$

$x \geqslant 0$
$y \geqslant 0$

$x_{uv} - y_{\overrightarrow{uv},w} - y_{\overrightarrow{vu},w} = 0\ \forall u, v, w$
$x_{uv} + \sum_{w \neq u,v} y_{\overrightarrow{uw},v} = 1\ \forall u, v$
$\sum_{uv \in E_n} x_{uv} = n - 1$

size $\approx n^3 \rightarrow$ compact

Is there an EF with even fewer inequalities?

Some Examples.

Some known results (constructions & lower bounds):

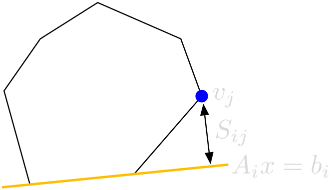
- $\mathrm{xc}($regular n-gon$) = \Theta(\log n)$ [Ben-Tal, Nemirovski'01]
- $\mathrm{xc}($generic n-gon$) = \Omega(\sqrt{n})$ [Fiorini, Rothvoss, Tiwary'11]
- $\mathrm{xc}(n$-permutahedron$) = \Theta(n \log n)$ [Goemans'09]
- $\mathrm{xc}($spanning tree polytope of $K_n) = O(n^3)$ [Kipp-Martin'87]
- $\mathrm{xc}($spanning tree polytope of planar graph $G) = \Theta(n)$ [Williams'01]
- $\mathrm{xc}($stable set polytope of perfect graph $G) = n^{O(\log n)}$ [Yannakakis'91]
- ...

Analyzing extended formulations...

Slack Matrices.

Let $A \in \mathbb{R}^{m \times d}$, $\quad b \in \mathbb{R}^m$, $\quad V = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^d \quad$ s.t.

$$P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \mathrm{conv}(V)$$



Definition (slack matrix)

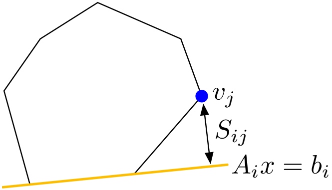
Slack matrix $S \in \mathbb{R}_+^{m \times n}$ of P (w.r.t. $Ax \leqslant b$ and V):

$$S_{ij} := b_i - A_i v_j$$

Slack Matrices.

Let $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$, $V = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^d$ s.t.

$$P = \{x \in \mathbb{R}^d \mid Ax \leqslant b\} = \mathrm{conv}(V)$$



Definition (slack matrix)

Slack matrix $S \in \mathbb{R}_{+}^{m \times n}$ of P (w.r.t. $Ax \leqslant b$ and V):

$$S_{ij} := b_i - A_i v_j$$

Nonnegative Factorizations and Factorization Theorem.

Definition

A rank-r nonnegative factorization of $S \in \mathbb{R}^{m \times n}$ is

$$S = TU \quad \text{where} \quad T \in \mathbb{R}_+^{m \times r} \quad \text{and} \quad U \in \mathbb{R}_+^{r \times n}$$

Definition (nonnegative rank of S)

$$\begin{aligned} \mathrm{rk}_+(S) &:= \min\{r \mid \exists \text{ rank-}r \text{ nonnegative factorization of } S\} \\ &= \min\{r \mid S \text{ is the sum of } r \text{ nonnegative rank-1 matrices}\} \end{aligned}$$

Theorem (factorization theorem [Yannakakis'91, FKPT'11])

For every slack matrix S of P:

$$\mathrm{xc}(P) = \mathrm{rk}_+(S)$$

Nonnegative Factorizations and Factorization Theorem.

Definition

A rank-r nonnegative factorization of $S \in \mathbb{R}^{m \times n}$ is

$$S = TU \quad \text{where} \quad T \in \mathbb{R}_{+}^{m \times r} \quad \text{and} \quad U \in \mathbb{R}_{+}^{r \times n}$$

Definition (nonnegative rank of S)

$$\begin{aligned} \mathrm{rk}_{+}(S) &:= \min\{r \mid \exists \text{ rank-}r \text{ nonnegative factorization of } S\} \\ &= \min\{r \mid S \text{ is the sum of } r \text{ nonnegative rank-1 matrices}\} \end{aligned}$$

Theorem (factorization theorem [Yannakakis'91, FKPT'11])

For every slack matrix S of P:

$$\mathrm{xc}(P) = \mathrm{rk}_{+}(S)$$

Nonnegative Factorizations and Factorization Theorem.

Definition

A rank-r nonnegative factorization of $S \in \mathbb{R}^{m \times n}$ is

$$S = TU \quad \text{where} \quad T \in \mathbb{R}_{+}^{m \times r} \quad \text{and} \quad U \in \mathbb{R}_{+}^{r \times n}$$

Definition (nonnegative rank of S)

$$\begin{aligned} \mathrm{rk}_{+}(S) &:= \min\{r \mid \exists \text{ rank-}r \text{ nonnegative factorization of } S\} \\ &= \min\{r \mid S \text{ is the sum of } r \text{ nonnegative rank-1 matrices}\} \end{aligned}$$

Theorem (factorization theorem [Yannakakis'91, FKPT'11])

For every slack matrix S of P:

$$\mathrm{xc}(P) = \mathrm{rk}_{+}(S)$$

Main goal: bound the nonnegative rank!

A simple lower bound:
(arguably) the mother of all lower bounds)

$$S = TU \qquad \text{rank-}r \text{ nonnegative factorization}$$

$$= \sum_{k=1}^{r} T^k U_k \qquad \text{sum of } r \text{ nonnegative rank-1 matrices}$$

$$\Longrightarrow \operatorname{supp}(S) = \bigcup_{k=1}^{r} \operatorname{supp}(T^k U_k)$$

$$= \bigcup_{k=1}^{r} \operatorname{supp}(T^k) \times \operatorname{supp}(U_k) \qquad \text{union of } r \text{ rectangles}$$

$$\begin{aligned} S &= TU && \text{rank-}r\text{ nonnegative factorization} \\ &= \sum_{k=1}^{r} T^k U_k && \text{sum of } r \text{ nonnegative rank-1 matrices} \\ \Longrightarrow \operatorname{supp}(S) &= \bigcup_{k=1}^{r} \operatorname{supp}(T^k U_k) \\ &= \bigcup_{k=1}^{r} \operatorname{supp}(T^k) \times \operatorname{supp}(U_k) && \text{union of } r \text{ rectangles} \end{aligned}$$

Definition (rectangle covering number)

$\mathrm{rc}(S) :=$ min # rectangles whose union is $\operatorname{supp}(S)$

Observation [Yannakakis'91]

$$\mathrm{rk}_+(S) \geqslant \mathrm{rc}(S)$$

$$S = TU \qquad \text{rank-}r \text{ nonnegative factorization}$$

$$= \sum_{k=1}^{r} T^k U_k \qquad \text{sum of } r \text{ nonnegative rank-1 matrices}$$

$$\Longrightarrow \mathrm{supp}(S) = \bigcup_{k=1}^{r} \mathrm{supp}(T^k U_k)$$

$$= \bigcup_{k=1}^{r} \mathrm{supp}(T^k) \times \mathrm{supp}(U_k) \qquad \text{union of } r \text{ rectangles}$$

Definition (rectangle covering number)

$\mathrm{rc}(S) :=$ min # rectangles whose union is $\mathrm{supp}(S)$

Observation [Yannakakis'91]

$$\mathrm{rk}_+(S) \geqslant \mathrm{rc}(S)$$

$$S = TU \qquad \text{rank-}r \text{ nonnegative factorization}$$

$$= \sum_{k=1}^{r} T^k U_k \qquad \text{sum of } r \text{ nonnegative rank-1 matrices}$$

$$\implies \operatorname{supp}(S) = \bigcup_{k=1}^{r} \operatorname{supp}(T^k U_k)$$

$$= \bigcup_{k=1}^{r} \operatorname{supp}(T^k) \times \operatorname{supp}(U_k) \qquad \text{union of } r \text{ rectangles}$$

Definition (rectangle covering number)

$\operatorname{rc}(S) :=$ min $\#$ rectangles whose union is $\operatorname{supp}(S)$

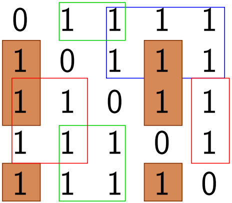
Observation [Yannakakis'91]

$$\operatorname{rk}_+(S) \geqslant \operatorname{rc}(S)$$

$$
\begin{aligned}
S &= TU && \text{rank-}r\text{ nonnegative factorization} \\
&= \sum_{k=1}^{r} T^k U_k && \text{sum of } r \text{ nonnegative rank-1 matrices} \\
\implies \operatorname{supp}(S) &= \bigcup_{k=1}^{r} \operatorname{supp}(T^k U_k) \\
&= \bigcup_{k=1}^{r} \operatorname{supp}(T^k) \times \operatorname{supp}(U_k) && \text{union of } r \text{ rectangles}
\end{aligned}
$$

0	1	1	1	1
1	0	1	1	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0

$$
\begin{aligned}
S = TU & \qquad \text{rank-}r \text{ nonnegative factorization} \\
= \sum_{k=1}^{r} T^k U_k & \qquad \text{sum of } r \text{ nonnegative rank-1 matrices} \\
\implies \operatorname{supp}(S) = \bigcup_{k=1}^{r} \operatorname{supp}(T^k U_k) & \\
= \bigcup_{k=1}^{r} \operatorname{supp}(T^k) \times \operatorname{supp}(U_k) & \qquad \text{union of } r \text{ rectangles}
\end{aligned}
$$



Definition (Fooling Set)

Let S be a nonnegative matrix. Then a fooling set F is a set of indices so that

1. $M(a, b) > 0$ for all $(a, b) \in F$.
2. for all $(a_1, b_1), (a_2, b_2) \in F$ distinct, either $M(a_1, b_2) = 0$ or $M(a_2, b_1) = 0$.

Definition (Fooling Set)

Let S be a nonnegative matrix. Then a fooling set F is a set of indices so that

1. $M(a, b) > 0$ for all $(a, b) \in F$.
2. for all $(a_1, b_1), (a_2, b_2) \in F$ distinct, either $M(a_1, b_2) = 0$ or $M(a_2, b_1) = 0$.

Lemma

If F is a fooling set for M of size k, then $\mathrm{rk}_+(M) \geq k$

Proof sketch.

No two elements of F can be in the same rank-1 matrix. □

Effectiveness of Fooling Set method is limited [Fiorini, Kaibel, Pashkovich, Theis'11]:

$$|F| = O(\text{rank}(M)^2)$$

Definition (Fooling Set)

Let S be a nonnegative matrix. Then a fooling set F is a set of indices so that

1. $M(a, b) > 0$ for all $(a, b) \in F$.
2. for all $(a_1, b_1), (a_2, b_2) \in F$ distinct, either $M(a_1, b_2) = 0$ or $M(a_2, b_1) = 0$.

Lemma (Fiorini, Kaibel, Pashkovich, Theis'11)

$P := [0, 1]^n$ *has a fooling set of size* $2n$.

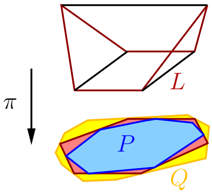
How about approximations?

Often we are interested in approximate LP formulations.

- $P \subseteq Q \subseteq \mathbb{R}^d$ with P polytope, Q polyhedron
- $L \subseteq \mathbb{R}^e$ polytope

Definition (extension of a pair)

L is an extension of (P, Q) if $\exists$ linear π with $P \subseteq \pi(L) \subseteq Q$



Definition (EF of a pair)

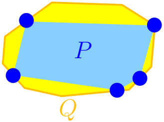
$Ex + Fy = g,\ y \geqslant \mathbf{0}$ is an extended formulation of (P, Q) if

$$x \in P \Longrightarrow \exists y : Ex + Fy = g, y \geqslant \mathbf{0} \Longrightarrow x \in Q$$

Factorization Theorem for Pairs.

Let $V = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^d$ s.t. $P = \mathrm{conv}(V)$

Let $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ s.t. $Q = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$



Definition (Slack matrix of pair)

Slack matrix $S = S^{P,Q} \in \mathbb{R}_+^{m \times n}$ of (P, Q) (w.r.t. $Ax \leqslant b$ and V):

$$S_{ij}^{P,Q} := b_i - A_i v_j$$

Definition (Extension complexity of a pair)

$\mathrm{xc}(P, Q) = \min\{\text{size}(L) \mid L \text{ is an extension of } (P, Q)\}$

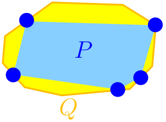
Theorem (Factorization theorem for pairs)

For every slack matrix $S^{P,Q}$ of (P, Q): $\quad \mathrm{xc}(P, Q) = \mathrm{rk}_+(S^{P,Q})$

Factorization Theorem for Pairs.

Let $V = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^d$ s.t. $P = \mathrm{conv}(V)$

Let $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ s.t. $Q = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$



Definition (Slack matrix of pair)

Slack matrix $S = S^{P,Q} \in \mathbb{R}_+^{m \times n}$ of (P, Q) (w.r.t. $Ax \leqslant b$ and V):

$$S_{ij}^{P,Q} := b_i - A_i v_j$$

Definition (Extension complexity of a pair)

$\mathrm{xc}(P, Q) = \min\{\text{size}(L) \mid L \text{ is an extension of } (P, Q)\}$

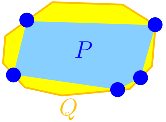
Theorem (Factorization theorem for pairs)

For every slack matrix $S^{P,Q}$ of (P, Q): $\quad \mathrm{xc}(P, Q) = \mathrm{rk}_+(S^{P,Q})$

Factorization Theorem for Pairs.

Let $V = \{v_1, \ldots, v_n\} \subseteq \mathbb{R}^d$ s.t. $P = \text{conv}(V)$

Let $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ s.t. $Q = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$



Definition (Slack matrix of pair)

Slack matrix $S = S^{P,Q} \in \mathbb{R}_+^{m \times n}$ of (P, Q) (w.r.t. $Ax \leqslant b$ and V):

$$S_{ij}^{P,Q} := b_i - A_i v_j$$

Definition (Extension complexity of a pair)

$\text{xc}(P, Q) = \min\{\text{size}(L) \mid L \text{ is an extension of } (P, Q)\}$

Theorem (Factorization theorem for pairs)

For every slack matrix $S^{P,Q}$ of (P, Q): $\quad \text{xc}(P, Q) = \text{rk}_+(S^{P,Q})$

Linear encoding $(\mathcal{L}, \mathcal{O}) \rightsquigarrow$ pair of nested polyhedra $P \subseteq Q$:

- $P := \operatorname{conv}(\{x \in \{0,1\}^d \mid x \in \mathcal{L}\})$
- $Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathcal{O} \cap \mathbb{R}^d : w^\intercal x \leqslant \max\{w^\intercal y \mid y \in P\}\}$

Definition (ρ-approximate extended formulation, $\rho \geqslant 1$)

$Ex + Fy = g$, $y \geqslant \mathbf{0}$ is a ρ-approximate EF w.r.t. $(\mathcal{L}, \mathcal{O})$ if

1. $\forall w \in \mathbb{R}^d$:
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \geqslant \max\{w^\intercal x \mid x \in P\}$$
2. $\forall w \in \mathcal{O} \cap \mathbb{R}^d$:
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \leqslant \rho \max\{w^\intercal x \mid x \in P\}$$

Geometrically: $P \subseteq \{x \mid \exists y : Ex + Fy = g,\ y \geqslant \mathbf{0}\} \subseteq \rho Q$

Linear encoding $(\mathcal{L}, \mathcal{O}) \rightsquigarrow$ pair of nested polyhedra $P \subseteq Q$:

- $P := \operatorname{conv}(\{x \in \{0,1\}^d \mid x \in \mathcal{L}\})$
- $Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathcal{O} \cap \mathbb{R}^d : w^\intercal x \leqslant \max\{w^\intercal y \mid y \in P\}\}$

Definition (ρ-approximate extended formulation, $\rho \geqslant 1$)

$Ex + Fy = g,\ y \geqslant \mathbf{0}$ is a ρ-approximate EF w.r.t. $(\mathcal{L}, \mathcal{O})$ if

1. $\forall w \in \mathbb{R}^d$:
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \geqslant \max\{w^\intercal x \mid x \in P\}$$
2. $\forall w \in \mathcal{O} \cap \mathbb{R}^d$:
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \leqslant \rho \max\{w^\intercal x \mid x \in P\}$$

Geometrically: $P \subseteq \{x \mid \exists y : Ex + Fy = g,\ y \geqslant \mathbf{0}\} \subseteq \rho Q$

Linear encoding $(\mathcal{L}, \mathcal{O}) \rightsquigarrow$ pair of nested polyhedra $P \subseteq Q$:

- $P := \operatorname{conv}(\{x \in \{0,1\}^d \mid x \in \mathcal{L}\})$
- $Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathcal{O} \cap \mathbb{R}^d : w^\intercal x \leqslant \max\{w^\intercal y \mid y \in P\}\}$

Definition (ρ-approximate extended formulation, $\rho \geqslant 1$)

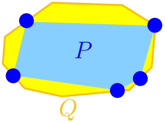
$Ex + Fy = g,\ y \geqslant \mathbf{0}$ is a ρ-approximate EF w.r.t. $(\mathcal{L}, \mathcal{O})$ if

1. $\forall w \in \mathbb{R}^d$:
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \geqslant \max\{w^\intercal x \mid x \in P\}$$
2. $\forall w \in \mathcal{O} \cap \mathbb{R}^d$:
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \leqslant \rho \max\{w^\intercal x \mid x \in P\}$$

Geometrically: $P \subseteq \{x \mid \exists y : Ex + Fy = g,\ y \geqslant \mathbf{0}\} \subseteq \rho Q$

Sizes of Approximate Extended Formulations.

- $\mathcal{L} \rightsquigarrow P = \mathrm{conv}(V)$
- $\mathcal{O} \rightsquigarrow Q = \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$



Observation:

1. $\rho Q = \{x \in \mathbb{R}^d \mid Ax \leqslant \rho b\}$
2. $S_{ij}^{P,\rho Q} = \rho b_i - A_i v_j = S_{ij}^{P,Q} + (\rho - 1) b_i$

Corollary

Minimum size of a ρ-approximate EF $= \mathrm{rk}_+(S^{P,\rho Q})$

A link to communication complexity

Deterministic Communication Protocols.

A Basic Model in Communication Complexity

$f : A \times B \to \{0, 1\}$ Boolean function ($\equiv$ binary matrix)

Two players:

- Alice knows $a \in A$
- Bob knows $b \in B$

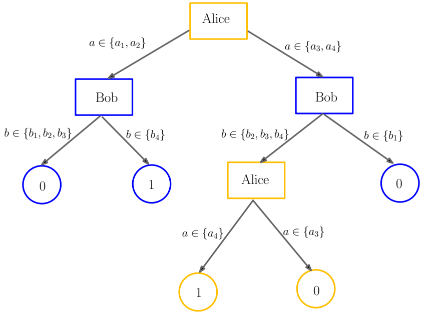
want to **compute** $f(a, b)$ **by exchanging bits**

Goal: Minimize complexity := #bits exchanged

Deterministic Communication Protocols.

Example

	b_1	b_2	b_3	b_4
a_1	0	0	0	1
a_2	0	0	0	1
a_3	0	0	0	0
a_4	0	1	1	1



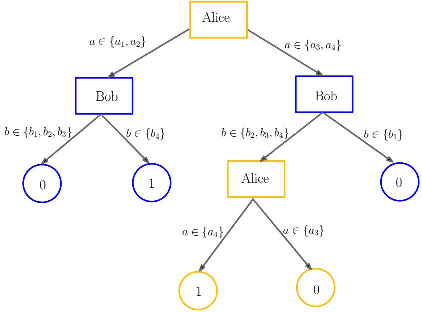
Observation

$\exists$ complexity c protocol for computing $f \implies \mathrm{rk}_+(f) \leqslant 2^c$

Deterministic Communication Protocols.

Example

	b_1	b_2	b_3	b_4
a_1	0	0	0	1
a_2	0	0	0	1
a_3	0	0	0	0
a_4	0	1	1	1



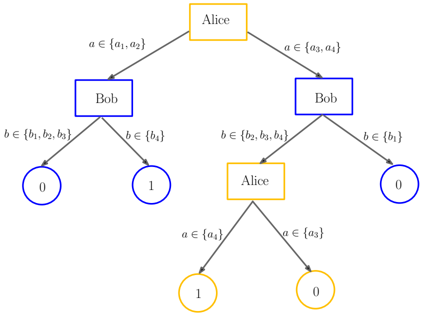
Observation

$\exists$ complexity c protocol for computing $f \implies \mathrm{rk}_+(f) \leqslant 2^c$

Deterministic Communication Protocols.

Example

	b_1	b_2	b_3	b_4
a_1	0	0	0	1
a_2	0	0	0	1
a_3	0	0	0	0
a_4	0	1	1	1



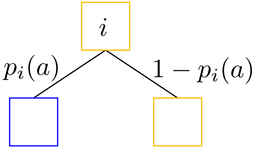
Observation

$\exists$ complexity c protocol for computing $f \implies \mathrm{rk}_+(f) \leqslant 2^c$

Computation in Expectation.

The main differences:

- Alice and Bob can use (private) random bits to make choices



- $f : A \times B \to \mathbb{R}_+$, Alice and Bob can output any value $\in \mathbb{R}_+$

Theorem ([Faenza, Fiorini, Grappe, Tiwary'11],[Zhang'12])

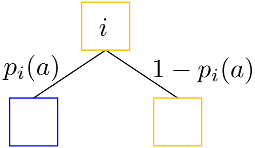
If $c = c(f)$ is the minimum complexity of a randomized communication protocol with nonnegative outputs computing f in expectation, then

$$\mathrm{rk}_+(f) = \Theta(2^c)$$

Computation in Expectation.

The main differences:

- Alice and Bob can use (private) random bits to make choices



- $f : A \times B \to \mathbb{R}_+$, Alice and Bob can output any value $\in \mathbb{R}_+$

Theorem ([Faenza, Fiorini, Grappe, Tiwary'11],[Zhang'12])

If $c = c(f)$ is the minimum complexity of a randomized communication protocol with nonnegative outputs computing f in expectation, then

$$\mathrm{rk}_+(f) = \Theta(2^c)$$

A threefold characterization.

Three ways to look at EFs:

1. A linear system $Ex + Fy = g$, $y \geqslant \mathbf{0}$ with $y \in \mathbb{R}^r$
2. A rank-r nonnegative factorization $S = TU$ of slack matrix S
3. A $\log r$-complexity randomized protocol with nonnegative outputs computing S in expectation

A threefold characterization.

Three ways to look at EFs:

1. A linear system $Ex + Fy = g$, $y \geqslant \mathbf{0}$ with $y \in \mathbb{R}^r$
2. A rank-r nonnegative factorization $S = TU$ of slack matrix S
3. A $\log r$-complexity randomized protocol with nonnegative outputs computing S in expectation

A threefold characterization.

Three ways to look at EFs:

1. A linear system $Ex + Fy = g$, $y \geqslant \mathbf{0}$ with $y \in \mathbb{R}^r$
2. A rank-r nonnegative factorization $S = TU$ of slack matrix S
3. A $\log r$-complexity randomized protocol with nonnegative outputs computing S in expectation

In summary: bound the nonnegative rank!
(both for approximate or exact linear EFs)

Common methods for construction of EFs.

1. Balas' union (union of polyhedra)
2. Reflection relations
3. Dualization
4. Extended formulations from dynamic programs
(we consider those to be part of 3)

Common methods for construction of EFs.

1. Balas' union (union of polyhedra)
2. Reflection relations
3. Dualization
4. Extended formulations from dynamic programs
   (we consider those to be part of 3)

Common methods for construction of EFs.

1. Balas' union (union of polyhedra)
2. Reflection relations
3. Dualization
4. Extended formulations from dynamic programs (we consider those to be part of 3)

Common methods for construction of EFs.

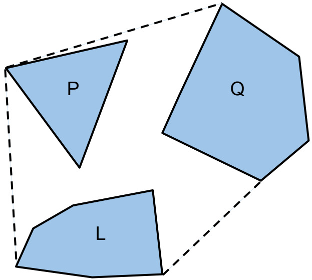
1. Balas' union (union of polyhedra)
2. Reflection relations
3. Dualization
4. Extended formulations from dynamic programs (we consider those to be part of 3)

Balas' union of polyhedra.

Idea: Express the union of polytopes as a polytopes.

[Balas 1985]

Approximately: $\mathrm{xc}(\mathrm{conv}(\bigcup_i P_i)) \leq \sum_i \mathrm{xc}(P_i)$.

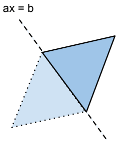


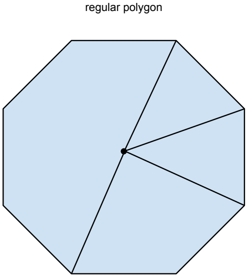
Used for approximate EF of the knapsack problem. [Bienstock 2008]

Reflection relations.

Idea: reflect (one side of) a polytope at a hyperplane.

[Kaibel, Pashkovich 2010]





Construction of regular n-gon with $O(\log n)$ many inequalities.

[Ben-Tal, Nemirovski 1999]

Dualization.

Idea: insert separation LP into the primal via dualization.

[Martin, 1991]

Spanning tree polytope of complete graph K_n with $\Theta(n^3)$ inequalities (example from the beginning).

How about semidefinite EFs?
Essentially the same theory applies...

Semidefinite Extended Formulations.

Definition (PSD matrix)

A matrix $U \in \mathbb{R}^{r \times r}$ is PSD if U is symmetric and

$$x^\intercal U x \geq 0 \quad \forall x \in \mathbb{R}^r.$$

Let $\mathbb{S}^r_+$ denote the set of $r \times r$ PSD matrices.

Definition (Spectral Decomposition)

U is $r \times r$ PSD iff U admits a *spectral decomposition*

$$U = \sum_{i=1}^{r} \lambda_i u_i u_i^\intercal,$$

$\lambda_1, \dots, \lambda_r \geq 0$, $u_1, \dots, u_r$ an orthonormal basis.

Semidefinite Extended Formulations.

Definition (PSD matrix)

A matrix $U \in \mathbb{R}^{r \times r}$ is PSD if U is symmetric and

$$x^\mathsf{T} U x \geq 0 \quad \forall x \in \mathbb{R}^r.$$

Let $\mathbb{S}^r_+$ denote the set of $r \times r$ PSD matrices.

Definition (Spectral Decomposition)

U is $r \times r$ PSD iff U admits a *spectral decomposition*

$$U = \sum_{i=1}^{r} \lambda_i u_i u_i^\mathsf{T},$$

$\lambda_1, \ldots, \lambda_r \geq 0$, $u_1, \ldots, u_r$ an orthonormal basis.

Definition (Operator norm)

For a matrix $T \in \mathbb{R}^{r \times r}$ the operator norm of T is

$$\|T\|_{\mathrm{op}} = \max_{\|x\|_2 = 1} \|Tx\|_2$$

For a PSD matrix $U \in \mathbb{R}^{r \times r}$

$$\|U\|_{\mathrm{op}} = \max_{\|x\|_2 = 1} x^{\mathsf{T}} U x = \text{ largest eigenvalue of } U.$$

Definition (Trace)

For a matrix $T \in \mathbb{R}^{r \times r}$, we define $\mathrm{Tr}\left[T\right] = \sum_{i=1}^{r} T_{ii}$.

Remark (Trace Inner Product)

For $A, B \in \mathbb{R}^{r \times r}$ symmetric, $\mathrm{Tr}\left[AB\right] = \sum_{i,j \in [r]} A_{ij} B_{ij}$.

Fact

For PSD matrices $U, V \in \mathbb{S}_{+}^{r}$,

$$\mathrm{Tr}\left[UV\right] = \sum_{i,j \in [r]} \lambda_i \gamma_j \left\langle u_i, v_j \right\rangle^2 \geq 0 \ ,$$

where $U = \sum_{i=1}^{r} \lambda_i u_i u_i^{\intercal}$ and $V = \sum_{j=1}^{r} \gamma_j v_j v_j^{\intercal}$ are the respective spectral decompositions.

Definition (Trace)

For a matrix $T \in \mathbb{R}^{r \times r}$, we define $\mathrm{Tr}\left[T\right] = \sum_{i=1}^{r} T_{ii}$.

Remark (Trace Inner Product)

For $A, B \in \mathbb{R}^{r \times r}$ *symmetric,* $\mathrm{Tr}\left[AB\right] = \sum_{i,j \in [r]} A_{ij} B_{ij}$.

Fact

For PSD matrices $U, V \in \mathbb{S}^r_+$,

$$\mathrm{Tr}\left[UV\right] = \sum_{i,j \in [r]} \lambda_i \gamma_j \left\langle u_i, v_j \right\rangle^2 \geq 0 \ ,$$

where $U = \sum_{i=1}^{r} \lambda_i u_i u_i^\intercal$ *and* $V = \sum_{j=1}^{r} \gamma_j v_j v_j^\intercal$ *are the respective spectral decompositions.*

Definition (Trace)

For a matrix $T \in \mathbb{R}^{r \times r}$, we define $\mathrm{Tr}\left[T\right] = \sum_{i=1}^{r} T_{ii}$.

Remark (Trace Inner Product)

For $A, B \in \mathbb{R}^{r \times r}$ symmetric, $\mathrm{Tr}\left[AB\right] = \sum_{i,j \in [r]} A_{ij} B_{ij}$.

Fact

For PSD matrices $U, V \in \mathbb{S}^{r}_{+}$,

$$\mathrm{Tr}\left[UV\right] = \sum_{i,j \in [r]} \lambda_i \gamma_j \left\langle u_i, v_j \right\rangle^2 \geq 0 \ ,$$

where $U = \sum_{i=1}^{r} \lambda_i u_i u_i^{\mathsf{T}}$ and $V = \sum_{j=1}^{r} \gamma_j v_j v_j^{\mathsf{T}}$ are the respective spectral decompositions.

Definition (SDP Extension)

$P = \{x \in \mathbb{R}^n : Ax \leq b\}$ polytope with m facets. Then

$$Q = \{(z, Y) : C_i z + \mathrm{Tr}\,[D_i Y] = d_i, \forall\ i \in [l], Y \in \mathbb{S}^r_+, z \in \mathbb{R}^l\},$$

is an SDP extension of P of size r if $\exists\ \pi : \mathbb{R}^l \times \mathbb{S}^r_+ \to \mathbb{R}^n$ such that

$$P = \pi(Q).$$

Definition (SDP Extension Complexity)

$\mathrm{xc}_{\mathrm{sdp}}(P) :=$ minimum size of any SDP extension of P.

Definition (SDP Extension)

$P = \{x \in \mathbb{R}^n : Ax \leq b\}$ polytope with m facets. Then

$$Q = \{(z, Y) : C_i z + \mathrm{Tr}\,[D_i Y] = d_i, \forall\ i \in [l], Y \in \mathbb{S}^r_+, z \in \mathbb{R}^l\},$$

is an SDP extension of P of size r if $\exists\ \pi : \mathbb{R}^l \times \mathbb{S}^r_+ \to \mathbb{R}^n$ such that

$$P = \pi(Q).$$

Definition (SDP Extension Complexity)

$\mathrm{xc}_{\mathrm{sdp}}(P) :=$ minimum size of any SDP extension of P.

PSD Factorizations and SDP Extensions.

Definition (PSD factorization)

A rank-r PSD factorization of $S \in \mathbb{R}_+^{m \times N}$ is given by $U_i, V_j \in \mathbb{S}_+^r$ if for all $i \in [m], j \in [N]$ we have

$$S_{ij} = \mathrm{Tr}\left[U_i V_j\right].$$

Proposition (Extensions from Factorizations)

Let P be polytope and let S be slack matrix of P. Then

$$Q = \{(x, Y) : A_i x + \mathrm{Tr}\left[U_i Y\right] = b_i, \forall i \in [m], Y \in \mathbb{S}_+^r\}$$

is a SDP extension of P of size r.

PSD Factorizations and SDP Extensions.

Definition (PSD factorization)

A rank-r PSD factorization of $S \in \mathbb{R}_+^{m \times N}$ is given by $U_i, V_j \in \mathbb{S}_+^r$ if for all $i \in [m], j \in [N]$ we have

$$S_{ij} = \mathrm{Tr}\left[U_i V_j\right].$$

Proposition (Extensions from Factorizations)

Let P be polytope and let S be slack matrix of P. Then

$$Q = \{(x, Y) : A_i x + \mathrm{Tr}\left[U_i Y\right] = b_i, \forall i \in [m], Y \in \mathbb{S}_+^r\}$$

is a SDP extension of P of size r.

PSD Factorizations and Extensions.

Definition (PSD Rank)

$$\mathrm{rk}_{\mathrm{psd}}(S) := \min\{r \mid \exists \text{ rank-}r \text{ PSD factorization of } S\}$$

[Gouveia, Thomas, Parillo '11]

Theorem (Factorization Theorem)

For every slack matrix S of P:

$$\mathrm{xc}_{\mathrm{sdp}}(P) = \mathrm{rk}_{\mathrm{psd}}(S)$$

Information Theory: the basics

Why Information Theory—more than a party trick?

1. Great whenever we want to model that something is 'learned'
   Prime examples: Minimax Theory in Statistics, Machine Learning, etc.
2. Heavily used in theoretical computer science
   Prime examples: Multiplicative Weight Updates, Communication Complexity, Data Structures, etc.
3. Measure *information* as a resource rather than *computation*

Upcoming survey: The Information-theoretic Method in Optimization

Why Information Theory—more than a party trick?

1. Great whenever we want to model that something is 'learned'
   Prime examples: Minimax Theory in Statistics, Machine Learning, etc.
2. Heavily used in theoretical computer science
   Prime examples: Multiplicative Weight Updates, Communication Complexity, Data Structures, etc.
3. Measure *information* as a resource rather than *computation*

Upcoming survey: The Information-theoretic Method in Optimization

Why Information Theory—more than a party trick?

1. Great whenever we want to model that something is 'learned'
   Prime examples: Minimax Theory in Statistics, Machine Learning, etc.
2. Heavily used in theoretical computer science
   Prime examples: Multiplicative Weight Updates, Communication Complexity, Data Structures, etc.
3. Measure *information* as a resource rather than *computation*

Upcoming survey: The Information-theoretic Method in Optimization

Why Information Theory—more than a party trick?

1. Great whenever we want to model that something is 'learned'
   Prime examples: Minimax Theory in Statistics, Machine Learning, etc.
2. Heavily used in theoretical computer science
   Prime examples: Multiplicative Weight Updates, Communication Complexity, Data Structures, etc.
3. Measure *information* as a resource rather than *computation*

Upcoming survey: The Information-theoretic Method in Optimization

The paradigm of information.

1. Computational unboundedness:
   We care for flow of information
2. Key is the reconstruction principle:
   A encodes $B \Rightarrow A$ contains all information about B
3. Very intuitive theory:
   Common sense reasoning can go a long way
4. In information space most operations are very natural and easy to perform

The paradigm of information.

1. Computational unboundedness:
   We care for flow of information
2. Key is the reconstruction principle:
   A encodes $B \Rightarrow A$ contains all information about B
3. Very intuitive theory:
   Common sense reasoning can go a long way
4. In information space most operations are very natural and easy to perform

The paradigm of information.

1. Computational unboundedness:
   We care for flow of information
2. Key is the reconstruction principle:
   A encodes $B \Rightarrow A$ contains all information about B
3. Very intuitive theory:
   Common sense reasoning can go a long way
4. In information space most operations are very natural and easy to perform

The paradigm of information.

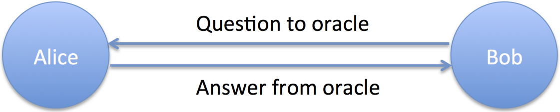
1. Computational unboundedness:
   We care for flow of information

2. Key is the reconstruction principle:
   A encodes $B \Rightarrow A$ contains all information about B

3. Very intuitive theory:
   Common sense reasoning can go a long way

4. In information space most operations are very natural and easy to perform

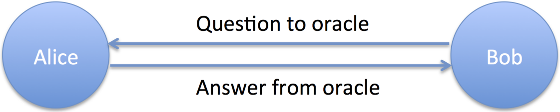
The paradigm of information.



For bigger picture some non-EF examples:

1. Blackbox optimization.
   Question: current point x
   Answer: $\nabla f(x)$ and $f(x)$
2. Separation oracle.
   Question: current point x
   Answer: separating hyperplane
3. Compressed sensing.
   Question: measurement vector x
   Answer: outcome of measurement Ax

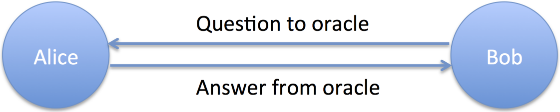
The paradigm of information.



For bigger picture some non-EF examples:

1. Blackbox optimization.
   Question: current point x
   Answer: $\nabla f(x)$ and $f(x)$
2. Separation oracle.
   Question: current point x
   Answer: separating hyperplane
3. Compressed sensing.
   Question: measurement vector x
   Answer: outcome of measurement Ax

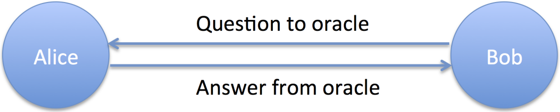
The paradigm of information.



For bigger picture some non-EF examples:

1. Blackbox optimization.
   Question: current point x
   Answer: $\nabla f(x)$ and $f(x)$
2. Separation oracle.
   Question: current point x
   Answer: separating hyperplane
3. Compressed sensing.
   Question: measurement vector x
   Answer: outcome of measurement Ax

The paradigm of information.



For bigger picture some non-EF examples:

1. Blackbox optimization.
   Question: current point x
   Answer: $\nabla f(x)$ and $f(x)$
2. Separation oracle.
   Question: current point x
   Answer: separating hyperplane
3. Compressed sensing.
   Question: measurement vector x
   Answer: outcome of measurement Ax

Notation and Notions.

Notation:

1. Random variables: $\mathbf{A}$
2. Events: $\mathcal{E}$
3. Conditionals (combination of RVs and Events): $\mathfrak{C}$
4. We write $a \in \mathbf{A}$ for $a \in \mathrm{Range}(\mathbf{A})$

Notions:

1. Often we identify an RV $\mathbf{\Pi}$ with its distribution

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}\,(\mathbf{A})| < \infty$. Then the *entropy of $\mathbf{A}$*:

$$\mathbb{H}\,[\mathbf{A}] \coloneqq - \sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}\,[\mathbf{A} = a] \log \mathbb{P}\,[\mathbf{A} = a]\,.$$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of* $\mathbf{A}$:

$$\mathbb{H}[\mathbf{A}] \coloneqq - \sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of $\mathbf{A}$*:

$$\mathbb{H}[\mathbf{A}] \coloneqq - \sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Interpretation:

1. Meta interpretation: 'information/randomness' in $\mathbf{A}$
2. Expected encoding length
3. Expected number of bits in optimal coding:

$$\mathbb{H}[\mathbf{A}] \leq L(C, \mathbf{A}) \leq \mathbb{H}[\mathbf{A}] + 1$$

4. Extraction of random bits: use biased coin with entropy h to generate h unbiased bits per flip

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of* $\mathbf{A}$:

$$\mathbb{H}[\mathbf{A}] := -\sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Rules:

1. Conditional entropy: $\mathbb{H}[\mathbf{A} \mid \mathbf{B}] = \mathbb{E}_{b \sim \mathbf{B}}[\mathbb{H}[\mathbf{A} \mid \mathbf{B} = b]]$.
2. Bounds: $0 \leq \mathbb{H}[\mathbf{A}] \leq \log|\mathrm{Range}(\mathbf{A})|$ ($=$ iff uniform)
3. Monotonicity: $\mathbb{H}[\mathbf{A}] \geq \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$.
4. Independence: $\mathbf{A} \perp \mathbf{B}$ if and only if $\mathbb{H}[\mathbf{A}] = \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$
5. Chain rule: $\mathbb{H}[\mathbf{A}, \mathbf{B}] = \mathbb{H}[\mathbf{A}] + \mathbb{H}[\mathbf{B} \mid \mathbf{A}]$.
6. Subadditivity: $\mathbb{H}[(\mathbf{A}_1, \ldots, \mathbf{A}_n)] \leq \sum_{i \in [n]} \mathbb{H}[\mathbf{A}_i]$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of* $\mathbf{A}$:

$$\mathbb{H}[\mathbf{A}] \coloneqq - \sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Rules:

1. Conditional entropy: $\mathbb{H}[\mathbf{A} \mid \mathbf{B}] = \mathbb{E}_{b \sim \mathbf{B}}[\mathbb{H}[\mathbf{A} \mid \mathbf{B} = b]]$.
2. Bounds: $0 \leq \mathbb{H}[\mathbf{A}] \leq \log |\mathrm{Range}(\mathbf{A})|$ ($=$ iff uniform)
3. Monotonicity: $\mathbb{H}[\mathbf{A}] \geq \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$.
4. Independence: $\mathbf{A} \perp \mathbf{B}$ if and only if $\mathbb{H}[\mathbf{A}] = \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$
5. Chain rule: $\mathbb{H}[\mathbf{A}, \mathbf{B}] = \mathbb{H}[\mathbf{A}] + \mathbb{H}[\mathbf{B} \mid \mathbf{A}]$.
6. Subadditivity: $\mathbb{H}[(\mathbf{A}_1, \ldots, \mathbf{A}_n)] \leq \sum_{i \in [n]} \mathbb{H}[\mathbf{A}_i]$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of* $\mathbf{A}$:

$$\mathbb{H}[\mathbf{A}] \coloneqq -\sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Rules:

1. Conditional entropy: $\mathbb{H}[\mathbf{A} \mid \mathbf{B}] = \mathbb{E}_{b \sim \mathbf{B}}[\mathbb{H}[\mathbf{A} \mid \mathbf{B} = b]]$.
2. Bounds: $0 \leq \mathbb{H}[\mathbf{A}] \leq \log|\mathrm{Range}(\mathbf{A})|$ ($=$ iff uniform)
3. Monotonicity: $\mathbb{H}[\mathbf{A}] \geq \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$.
4. Independence: $\mathbf{A} \perp \mathbf{B}$ if and only if $\mathbb{H}[\mathbf{A}] = \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$
5. Chain rule: $\mathbb{H}[\mathbf{A}, \mathbf{B}] = \mathbb{H}[\mathbf{A}] + \mathbb{H}[\mathbf{B} \mid \mathbf{A}]$.
6. Subadditivity: $\mathbb{H}[(\mathbf{A}_1, \ldots, \mathbf{A}_n)] \leq \sum_{i \in [n]} \mathbb{H}[\mathbf{A}_i]$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of $\mathbf{A}$*:

$$\mathbb{H}[\mathbf{A}] := -\sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Rules:

1. Conditional entropy: $\mathbb{H}[\mathbf{A} \mid \mathbf{B}] = \mathbb{E}_{b \sim \mathbf{B}}[\mathbb{H}[\mathbf{A} \mid \mathbf{B} = b]]$.
2. Bounds: $0 \leq \mathbb{H}[\mathbf{A}] \leq \log|\mathrm{Range}(\mathbf{A})|$ ($=$ iff uniform)
3. Monotonicity: $\mathbb{H}[\mathbf{A}] \geq \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$.
4. Independence: $\mathbf{A} \perp \mathbf{B}$ if and only if $\mathbb{H}[\mathbf{A}] = \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$
5. Chain rule: $\mathbb{H}[\mathbf{A}, \mathbf{B}] = \mathbb{H}[\mathbf{A}] + \mathbb{H}[\mathbf{B} \mid \mathbf{A}]$.
6. Subadditivity: $\mathbb{H}[(\mathbf{A}_1, \ldots, \mathbf{A}_n)] \leq \sum_{i \in [n]} \mathbb{H}[\mathbf{A}_i]$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of* $\mathbf{A}$:

$$\mathbb{H}[\mathbf{A}] \coloneqq - \sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Rules:

1. Conditional entropy: $\mathbb{H}[\mathbf{A} \mid \mathbf{B}] = \mathbb{E}_{b \sim \mathbf{B}}[\mathbb{H}[\mathbf{A} \mid \mathbf{B} = b]]$.
2. Bounds: $0 \leq \mathbb{H}[\mathbf{A}] \leq \log|\mathrm{Range}(\mathbf{A})|$ ($=$ iff uniform)
3. Monotonicity: $\mathbb{H}[\mathbf{A}] \geq \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$.
4. Independence: $\mathbf{A} \perp \mathbf{B}$ if and only if $\mathbb{H}[\mathbf{A}] = \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$
5. Chain rule: $\mathbb{H}[\mathbf{A}, \mathbf{B}] = \mathbb{H}[\mathbf{A}] + \mathbb{H}[\mathbf{B} \mid \mathbf{A}]$.
6. Subadditivity: $\mathbb{H}[(\mathbf{A}_1, \ldots, \mathbf{A}_n)] \leq \sum_{i \in [n]} \mathbb{H}[\mathbf{A}_i]$

Entropy (of a random variable).

$\mathbf{A}$ discrete RV with $|\mathrm{Range}(\mathbf{A})| < \infty$. Then the *entropy of* $\mathbf{A}$:

$$\mathbb{H}[\mathbf{A}] := -\sum_{a \in \mathrm{Range}(\mathbf{A})} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \mathbb{P}[\mathbf{A} = a]}_{\text{encoding length}}.$$

Rules:

1. Conditional entropy: $\mathbb{H}[\mathbf{A} \mid \mathbf{B}] = \mathbb{E}_{b \sim \mathbf{B}}[\mathbb{H}[\mathbf{A} \mid \mathbf{B} = b]]$.
2. Bounds: $0 \leq \mathbb{H}[\mathbf{A}] \leq \log|\mathrm{Range}(\mathbf{A})|$ ($=$ iff uniform)
3. Monotonicity: $\mathbb{H}[\mathbf{A}] \geq \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$.
4. Independence: $\mathbf{A} \perp \mathbf{B}$ if and only if $\mathbb{H}[\mathbf{A}] = \mathbb{H}[\mathbf{A} \mid \mathbf{B}]$
5. Chain rule: $\mathbb{H}[\mathbf{A}, \mathbf{B}] = \mathbb{H}[\mathbf{A}] + \mathbb{H}[\mathbf{B} \mid \mathbf{A}]$.
6. Subadditivity: $\mathbb{H}[(\mathbf{A}_1, \ldots, \mathbf{A}_n)] \leq \sum_{i \in [n]} \mathbb{H}[\mathbf{A}_i]$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathbb{I}[\mathbf{A}; \mathbf{B}] := \mathbb{H}[\mathbf{A}] - \mathbb{H}[\mathbf{A} \mid \mathbf{B}].$$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of* $\mathbf{A}$ *and* $\mathbf{B}$:

$$\mathbb{I}[\mathbf{A};\mathbf{B}] \coloneqq \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}}.$$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathbb{I}[\mathbf{A};\mathbf{B}] := \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}} .$$

Interpretation:

1. Meta interpretation: The amount of information leaked about $\mathbf{A}$ by observing $\mathbf{B}$.
2. From single RV (as in entropy) to interaction of RVs.
3. Models information gained from observation.

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of* $\mathbf{A}$ *and* $\mathbf{B}$:

$$\mathbb{I}[\mathbf{A};\mathbf{B}] := \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}}.$$

Rules:

1. Conditional mutual information ($\mathfrak{C}$ conditional): $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] := \mathbb{H}[\mathbf{A} \mid \mathfrak{C}] - \mathbb{H}[\mathbf{A} \mid \mathfrak{C}, \mathbf{B}]$
2. Bounds: $0 \leq \mathbb{I}[\mathbf{A};\mathbf{B}] \leq \mathbb{H}[\mathbf{A}]$.
3. Symmetry: $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] = \mathbb{I}[\mathbf{B};\mathbf{A} \mid \mathfrak{C}]$.
4. Chain rule: $\mathbb{I}[\mathbf{A}_1, \mathbf{A}_2;\mathbf{B}] = \mathbb{I}[\mathbf{A}_1;\mathbf{B}] + \mathbb{I}[\mathbf{A}_2;\mathbf{B} \mid \mathbf{A}_1]$.
5. Direct sum: If $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are independent. Then

$$\mathbb{I}[\mathbf{A}_1, \ldots, \mathbf{A}_n;\mathbf{B}] \geq \sum_{i \in [n]} \mathbb{I}[\mathbf{A}_i;\mathbf{B}].$$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathbb{I}[\mathbf{A};\mathbf{B}] := \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}}.$$

Rules:

1. Conditional mutual information ($\mathfrak{C}$ conditional): $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] := \mathbb{H}[\mathbf{A} \mid \mathfrak{C}] - \mathbb{H}[\mathbf{A} \mid \mathfrak{C}, \mathbf{B}]$
2. Bounds: $0 \leq \mathbb{I}[\mathbf{A};\mathbf{B}] \leq \mathbb{H}[\mathbf{A}]$.
3. Symmetry: $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] = \mathbb{I}[\mathbf{B};\mathbf{A} \mid \mathfrak{C}]$.
4. Chain rule: $\mathbb{I}[\mathbf{A}_1, \mathbf{A}_2;\mathbf{B}] = \mathbb{I}[\mathbf{A}_1;\mathbf{B}] + \mathbb{I}[\mathbf{A}_2;\mathbf{B} \mid \mathbf{A}_1]$.
5. Direct sum: If $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are independent. Then

$$\mathbb{I}[\mathbf{A}_1, \ldots, \mathbf{A}_n;\mathbf{B}] \geq \sum_{i \in [n]} \mathbb{I}[\mathbf{A}_i;\mathbf{B}].$$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathbb{I}[\mathbf{A}; \mathbf{B}] \coloneqq \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}}.$$

Rules:

1. Conditional mutual information ($\mathfrak{C}$ conditional): $\mathbb{I}[\mathbf{A}; \mathbf{B} \mid \mathfrak{C}] \coloneqq \mathbb{H}[\mathbf{A} \mid \mathfrak{C}] - \mathbb{H}[\mathbf{A} \mid \mathfrak{C}, \mathbf{B}]$
2. Bounds: $0 \leq \mathbb{I}[\mathbf{A}; \mathbf{B}] \leq \mathbb{H}[\mathbf{A}]$.
3. Symmetry: $\mathbb{I}[\mathbf{A}; \mathbf{B} \mid \mathfrak{C}] = \mathbb{I}[\mathbf{B}; \mathbf{A} \mid \mathfrak{C}]$.
4. Chain rule: $\mathbb{I}[\mathbf{A}_1, \mathbf{A}_2; \mathbf{B}] = \mathbb{I}[\mathbf{A}_1; \mathbf{B}] + \mathbb{I}[\mathbf{A}_2; \mathbf{B} \mid \mathbf{A}_1]$.
5. Direct sum: If $\mathbf{A}_1, \dots, \mathbf{A}_n$ are independent. Then
$$\mathbb{I}[\mathbf{A}_1, \dots, \mathbf{A}_n; \mathbf{B}] \geq \sum_{i \in [n]} \mathbb{I}[\mathbf{A}_i; \mathbf{B}].$$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of* $\mathbf{A}$ *and* $\mathbf{B}$:

$$\mathbb{I}[\mathbf{A};\mathbf{B}] := \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}}.$$

Rules:

1. Conditional mutual information ($\mathfrak{C}$ conditional): $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] := \mathbb{H}[\mathbf{A} \mid \mathfrak{C}] - \mathbb{H}[\mathbf{A} \mid \mathfrak{C}, \mathbf{B}]$
2. Bounds: $0 \leq \mathbb{I}[\mathbf{A};\mathbf{B}] \leq \mathbb{H}[\mathbf{A}]$.
3. Symmetry: $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] = \mathbb{I}[\mathbf{B};\mathbf{A} \mid \mathfrak{C}]$.
4. Chain rule: $\mathbb{I}[\mathbf{A}_1, \mathbf{A}_2;\mathbf{B}] = \mathbb{I}[\mathbf{A}_1;\mathbf{B}] + \mathbb{I}[\mathbf{A}_2;\mathbf{B} \mid \mathbf{A}_1]$.
5. Direct sum: If $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are independent. Then

$$\mathbb{I}[\mathbf{A}_1, \ldots, \mathbf{A}_n;\mathbf{B}] \geq \sum_{i \in [n]} \mathbb{I}[\mathbf{A}_i;\mathbf{B}].$$

Mutual information (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *mutual information of* $\mathbf{A}$ *and* $\mathbf{B}$:

$$\mathbb{I}[\mathbf{A};\mathbf{B}] := \underbrace{\mathbb{H}[\mathbf{A}]}_{\text{initial uncertainty}} - \underbrace{\mathbb{H}[\mathbf{A} \mid \mathbf{B}]}_{\text{residual uncertainty}} .$$

Rules:

1. Conditional mutual information ($\mathfrak{C}$ conditional): $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] := \mathbb{H}[\mathbf{A} \mid \mathfrak{C}] - \mathbb{H}[\mathbf{A} \mid \mathfrak{C}, \mathbf{B}]$
2. Bounds: $0 \leq \mathbb{I}[\mathbf{A};\mathbf{B}] \leq \mathbb{H}[\mathbf{A}]$.
3. Symmetry: $\mathbb{I}[\mathbf{A};\mathbf{B} \mid \mathfrak{C}] = \mathbb{I}[\mathbf{B};\mathbf{A} \mid \mathfrak{C}]$.
4. Chain rule: $\mathbb{I}[\mathbf{A}_1, \mathbf{A}_2;\mathbf{B}] = \mathbb{I}[\mathbf{A}_1;\mathbf{B}] + \mathbb{I}[\mathbf{A}_2;\mathbf{B} \mid \mathbf{A}_1]$.
5. Direct sum: If $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are independent. Then

$$\mathbb{I}[\mathbf{A}_1, \ldots, \mathbf{A}_n;\mathbf{B}] \geq \sum_{i \in [n]} \mathbb{I}[\mathbf{A}_i;\mathbf{B}] .$$

A first example: sorting by comparison.

Let $\mathbb{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbb{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}\left[\mathbf{F}\right] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}\left[\mathbf{F}\right] = \mathbb{I}\left[\mathbf{F}; \mathbf{\Pi}\right] \leq \mathbb{H}\left[\mathbf{\Pi}\right] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}\left[\mathbf{\Pi}_i\right]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

A first example: sorting by comparison.

Let $\mathbf{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbf{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}\left[\mathbf{F}\right] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}\left[\mathbf{F}\right] = \mathbb{I}\left[\mathbf{F}; \mathbf{\Pi}\right] \leq \mathbb{H}\left[\mathbf{\Pi}\right] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}\left[\mathbf{\Pi}_i\right]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

A first example: sorting by comparison.

Let $\mathbf{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbf{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}[\mathbf{F}] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}[\mathbf{F}] = \mathbb{I}[\mathbf{F}; \mathbf{\Pi}] \leq \mathbb{H}[\mathbf{\Pi}] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}[\mathbf{\Pi}_i]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

A first example: sorting by comparison.

Let $\mathbf{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbf{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}[\mathbf{F}] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}[\mathbf{F}] = \mathbb{I}[\mathbf{F}; \mathbf{\Pi}] \leq \mathbb{H}[\mathbf{\Pi}] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}[\mathbf{\Pi}_i]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

A first example: sorting by comparison.

Let $\mathbf{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbf{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}[\mathbf{F}] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}[\mathbf{F}] = \mathbb{I}[\mathbf{F}; \mathbf{\Pi}] \leq \mathbb{H}[\mathbf{\Pi}] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}[\mathbf{\Pi}_i]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

A first example: sorting by comparison.

Let $\mathbf{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbf{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}[\mathbf{F}] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}[\mathbf{F}] = \mathbb{I}[\mathbf{F}; \mathbf{\Pi}] \leq \mathbb{H}[\mathbf{\Pi}] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}[\mathbf{\Pi}_i]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

A first example: sorting by comparison.

Let $\mathbf{F}$ be a permutation of $1, \ldots, n$ chosen uniformly random.

Task: Sort $\mathbf{F}$ using only comparisons of the form $f_i < f_j$?

Then: $\mathbb{H}[\mathbf{F}] = \log n! = \Theta(n \log n)$.

Let $\mathbf{\Pi} = (\mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_\ell) \in \{0,1\}^\ell$ transcript of answers
(query independent of instance conditioned on what learned so far)

Reconstruction principle (conservation of information):

$$\Theta(n \log n) = \mathbb{H}[\mathbf{F}] = \mathbb{I}[\mathbf{F}; \mathbf{\Pi}] \leq \mathbb{H}[\mathbf{\Pi}] \leq \sum_{i \leq \ell} \underbrace{\mathbb{H}[\mathbf{\Pi}_i]}_{\leq 1} \leq \ell.$$

(the algorithm's queries are an encoding for the instances)

$\Rightarrow \ell = \Omega(n \log n)$ required comparisons.

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of* $\mathbf{A}$ *and* $\mathbf{B}$:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) \coloneqq \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}}.$$

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) := \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}}.$$

Interpretation:

1. Meta interpretation: How many bits do we pay extra for encoding with $\mathbf{A}$ with a code for $\mathbf{B}$.
2. While not as nice as entropy and mutual information, it is the *Ur*-quantity
3. Models distance of distribution (non-symmetric).

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of* $\mathbf{A}$ *and* $\mathbf{B}$:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) := \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}}.$$

Rules:

1. Nonnegativity: $0 \leq \mathrm{D}(\mathbf{A} \,\|\, \mathbf{B})$.
2. Entropy: $\mathbb{H}[\mathbf{A}] = \log|\mathrm{Range}(\mathbf{A})| - \mathrm{D}(\mathbf{A} \,\|\, \mathbf{U})$.
3. Unique minimizer: $\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) = 0$ if and only if $\mathbf{A} = \mathbf{B}$.
4. Chain rule: $\mathrm{D}(\mathbf{A}_1, \mathbf{A}_2 \,\|\, \mathbf{B}_1, \mathbf{B}_2) = \mathrm{D}(\mathbf{A}_1 \,\|\, \mathbf{B}_1) + \mathrm{D}(\mathbf{A}_2 \mid \mathbf{A}_1 \,\|\, \mathbf{B}_2 \mid \mathbf{B}_1)$.
5. Direct sum: Let $(\mathbf{A}_1, \mathbf{B}_1), \ldots (\mathbf{A}_n, \mathbf{B}_n)$ be mutually independent. Then

$$\mathrm{D}(\mathbf{A}_1, \ldots, \mathbf{A}_n \,\|\, \mathbf{B}_1, \ldots, \mathbf{B}_n) = \sum_{i \in [n]} \mathrm{D}(\mathbf{A}_i \,\|\, \mathbf{B}_i).$$

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) := \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}}.$$

Rules:

1. Nonnegativity: $0 \leq \mathrm{D}(\mathbf{A} \,\|\, \mathbf{B})$.
2. Entropy: $\mathbb{H}[\mathbf{A}] = \log |\mathrm{Range}(\mathbf{A})| - \mathrm{D}(\mathbf{A} \,\|\, \mathbf{U})$.
3. Unique minimizer: $\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) = 0$ if and only if $\mathbf{A} = \mathbf{B}$.
4. Chain rule:
   $\mathrm{D}(\mathbf{A}_1, \mathbf{A}_2 \,\|\, \mathbf{B}_1, \mathbf{B}_2) = \mathrm{D}(\mathbf{A}_1 \,\|\, \mathbf{B}_1) + \mathrm{D}(\mathbf{A}_2 \mid \mathbf{A}_1 \,\|\, \mathbf{B}_2 \mid \mathbf{B}_1)$.
5. Direct sum: Let $(\mathbf{A}_1, \mathbf{B}_1), \ldots (\mathbf{A}_n, \mathbf{B}_n)$ be mutually independent. Then

$$\mathrm{D}(\mathbf{A}_1, \ldots, \mathbf{A}_n \,\|\, \mathbf{B}_1, \ldots, \mathbf{B}_n) = \sum_{i \in [n]} \mathrm{D}(\mathbf{A}_i \,\|\, \mathbf{B}_i).$$

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) := \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \log \underbrace{\frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}}.$$

Rules:

1. Nonnegativity: $0 \leq \mathrm{D}(\mathbf{A} \,\|\, \mathbf{B})$.
2. Entropy: $\mathbb{H}[\mathbf{A}] = \log |\mathrm{Range}(\mathbf{A})| - \mathrm{D}(\mathbf{A} \,\|\, \mathbf{U})$.
3. Unique minimizer: $\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) = 0$ if and only if $\mathbf{A} = \mathbf{B}$.
4. Chain rule: $\mathrm{D}(\mathbf{A}_1, \mathbf{A}_2 \,\|\, \mathbf{B}_1, \mathbf{B}_2) = \mathrm{D}(\mathbf{A}_1 \,\|\, \mathbf{B}_1) + \mathrm{D}(\mathbf{A}_2 \mid \mathbf{A}_1 \,\|\, \mathbf{B}_2 \mid \mathbf{B}_1)$.
5. Direct sum: Let $(\mathbf{A}_1, \mathbf{B}_1), \ldots (\mathbf{A}_n, \mathbf{B}_n)$ be mutually independent. Then

$$\mathrm{D}(\mathbf{A}_1, \ldots, \mathbf{A}_n \,\|\, \mathbf{B}_1, \ldots, \mathbf{B}_n) = \sum_{i \in [n]} \mathrm{D}(\mathbf{A}_i \,\|\, \mathbf{B}_i).$$

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) := \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}} .$$

Rules:

1. Nonnegativity: $0 \leq \mathrm{D}(\mathbf{A} \,\|\, \mathbf{B})$.
2. Entropy: $\mathbb{H}[\mathbf{A}] = \log |\mathrm{Range}(\mathbf{A})| - \mathrm{D}(\mathbf{A} \,\|\, \mathbf{U})$.
3. Unique minimizer: $\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) = 0$ if and only if $\mathbf{A} = \mathbf{B}$.
4. Chain rule: $\mathrm{D}(\mathbf{A}_1, \mathbf{A}_2 \,\|\, \mathbf{B}_1, \mathbf{B}_2) = \mathrm{D}(\mathbf{A}_1 \,\|\, \mathbf{B}_1) + \mathrm{D}(\mathbf{A}_2 \mid \mathbf{A}_1 \,\|\, \mathbf{B}_2 \mid \mathbf{B}_1)$.
5. Direct sum: Let $(\mathbf{A}_1, \mathbf{B}_1), \dots (\mathbf{A}_n, \mathbf{B}_n)$ be mutually independent. Then

$$\mathrm{D}(\mathbf{A}_1, \dots, \mathbf{A}_n \,\|\, \mathbf{B}_1, \dots, \mathbf{B}_n) = \sum_{i \in [n]} \mathrm{D}(\mathbf{A}_i \,\|\, \mathbf{B}_i) .$$

Relative Entropy (of two random variables).

$\mathbf{A}, \mathbf{B}$ discrete RVs with $|\mathrm{Range}(\mathbf{A})|, |\mathrm{Range}(\mathbf{B})| < \infty$. Then the *relative entropy of $\mathbf{A}$ and $\mathbf{B}$*:

$$\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) := \sum_{a \in \mathbf{A}} \mathbb{P}[\mathbf{A} = a] \underbrace{\log \frac{\mathbb{P}[\mathbf{A} = a]}{\mathbb{P}[\mathbf{B} = a]}}_{\text{divergence in bits}}.$$

Rules:

1. Nonnegativity: $0 \leq \mathrm{D}(\mathbf{A} \,\|\, \mathbf{B})$.
2. Entropy: $\mathbb{H}[\mathbf{A}] = \log |\mathrm{Range}(\mathbf{A})| - \mathrm{D}(\mathbf{A} \,\|\, \mathbf{U})$.
3. Unique minimizer: $\mathrm{D}(\mathbf{A} \,\|\, \mathbf{B}) = 0$ if and only if $\mathbf{A} = \mathbf{B}$.
4. Chain rule: $\mathrm{D}(\mathbf{A}_1, \mathbf{A}_2 \,\|\, \mathbf{B}_1, \mathbf{B}_2) = \mathrm{D}(\mathbf{A}_1 \,\|\, \mathbf{B}_1) + \mathrm{D}(\mathbf{A}_2 \mid \mathbf{A}_1 \,\|\, \mathbf{B}_2 \mid \mathbf{B}_1)$.
5. Direct sum: Let $(\mathbf{A}_1, \mathbf{B}_1), \dots (\mathbf{A}_n, \mathbf{B}_n)$ be mutually independent. Then

$$\mathrm{D}(\mathbf{A}_1, \dots, \mathbf{A}_n \,\|\, \mathbf{B}_1, \dots, \mathbf{B}_n) = \sum_{i \in [n]} \mathrm{D}(\mathbf{A}_i \,\|\, \mathbf{B}_i).$$

The reconstruction principle on steroids: Fano's inequality.

The reconstruction principle is a special case of Fano's inequality:

Consider Markov chain $\underbrace{\mathbf{X}}_{\text{hidden RV}} \rightarrow \underbrace{\mathbf{Y}}_{\text{observation}} \rightarrow \underbrace{\hat{\mathbf{X}}}_{\text{guess of } \mathbf{X}}$.

Now, let $\mathbf{E}$ indicate whether $\hat{\mathbf{X}} = \mathbf{X}$.

The reconstruction principle on steroids: Fano's inequality.

The reconstruction principle is a special case of Fano's inequality:

Consider Markov chain $\underbrace{\mathbf{X}}_{\text{hidden RV}} \rightarrow \underbrace{\mathbf{Y}}_{\text{observation}} \rightarrow \underbrace{\hat{\mathbf{X}}}_{\text{guess of } \mathbf{X}}$.

Now, let $\mathbf{E}$ indicate whether $\hat{\mathbf{X}} = \mathbf{X}$.

$$\underbrace{\mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}\right]}_{\text{remaining uncertainty}} \leq \mathbb{H}\left[\mathbf{X}, \mathbf{E} \,\middle|\, \hat{\mathbf{X}}\right] \leq \mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}, \mathbf{E}\right] + \mathbb{H}\left[\mathbf{E}\right]$$

$$\leq \mathbb{P}\left[\mathbf{E} = 0\right] \mathbb{H}\left[\mathbf{X}\right] + \mathbb{H}\left[\mathbf{E}\right]$$

We obtain: $\frac{\mathbb{H}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbb{H}[\mathbf{E}]}{\mathbb{H}[\mathbf{X}]} \leq \mathbb{P}\left[\mathbf{E} = 0\right]$

The reconstruction principle on steroids: Fano's inequality.

The reconstruction principle is a special case of Fano's inequality:

Consider Markov chain $\underbrace{\mathbf{X}}_{\text{hidden RV}} \rightarrow \underbrace{\mathbf{Y}}_{\text{observation}} \rightarrow \underbrace{\hat{\mathbf{X}}}_{\text{guess of } \mathbf{X}}$.

Now, let $\mathbf{E}$ indicate whether $\hat{\mathbf{X}} = \mathbf{X}$.

$$\underbrace{\mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}\right]}_{\text{remaining uncertainty}} \leq \mathbb{H}\left[\mathbf{X}, \mathbf{E} \,\middle|\, \hat{\mathbf{X}}\right] \leq \mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}, \mathbf{E}\right] + \mathbb{H}\left[\mathbf{E}\right]$$

$$\leq \mathbb{P}\left[\mathbf{E} = 0\right] \mathbb{H}\left[\mathbf{X}\right] + \mathbb{H}\left[\mathbf{E}\right]$$

We obtain: $\frac{\mathbb{H}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbb{H}[\mathbf{E}]}{\mathbb{H}[\mathbf{X}]} \leq \mathbb{P}\left[\mathbf{E} = 0\right]$

The reconstruction principle on steroids: Fano's inequality.

The reconstruction principle is a special case of Fano's inequality:

Consider Markov chain $\underbrace{\mathbf{X}}_{\text{hidden RV}} \rightarrow \underbrace{\mathbf{Y}}_{\text{observation}} \rightarrow \underbrace{\hat{\mathbf{X}}}_{\text{guess of } \mathbf{X}}$.

Now, let $\mathbf{E}$ indicate whether $\hat{\mathbf{X}} = \mathbf{X}$.

$$\underbrace{\mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}\right]}_{\text{remaining uncertainty}} \leq \mathbb{H}\left[\mathbf{X}, \mathbf{E} \,\middle|\, \hat{\mathbf{X}}\right] \leq \mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}, \mathbf{E}\right] + \mathbb{H}[\mathbf{E}]$$

$$\leq \mathbb{P}[\mathbf{E} = 0]\, \mathbb{H}[\mathbf{X}] + \mathbb{H}[\mathbf{E}]$$

We obtain: $\frac{\mathbb{H}[\mathbf{X} \mid \hat{\mathbf{X}}] - \mathbb{H}[\mathbf{E}]}{\mathbb{H}[\mathbf{X}]} \leq \mathbb{P}[\mathbf{E} = 0]$

The reconstruction principle on steroids: Fano's inequality.

The reconstruction principle is a special case of Fano's inequality:

Consider Markov chain $\underbrace{\mathbf{X}}_{\text{hidden RV}} \rightarrow \underbrace{\mathbf{Y}}_{\text{observation}} \rightarrow \underbrace{\hat{\mathbf{X}}}_{\text{guess of } \mathbf{X}}$.

Now, let $\mathbf{E}$ indicate whether $\hat{\mathbf{X}} = \mathbf{X}$.

$$\underbrace{\mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}\right]}_{\text{remaining uncertainty}} \leq \mathbb{H}\left[\mathbf{X}, \mathbf{E} \,\middle|\, \hat{\mathbf{X}}\right] \leq \mathbb{H}\left[\mathbf{X} \,\middle|\, \hat{\mathbf{X}}, \mathbf{E}\right] + \mathbb{H}\left[\mathbf{E}\right]$$

$$\leq \mathbb{P}\left[\mathbf{E} = 0\right] \mathbb{H}\left[\mathbf{X}\right] + \mathbb{H}\left[\mathbf{E}\right]$$

We obtain: $\frac{\mathbb{H}\left[\mathbf{X} \,|\, \hat{\mathbf{X}}\right] - \mathbb{H}[\mathbf{E}]}{\mathbb{H}[\mathbf{X}]} \leq \mathbb{P}\left[\mathbf{E} = 0\right]$

The reconstruction principle on steroids: Fano's inequality.

The reconstruction principle is a special case of Fano's inequality:

Consider Markov chain $\underbrace{\mathbf{X}}_{\text{hidden RV}} \rightarrow \underbrace{\mathbf{Y}}_{\text{observation}} \rightarrow \underbrace{\hat{\mathbf{X}}}_{\text{guess of }\mathbf{X}}.$

Now, let $\mathbf{E}$ indicate whether $\hat{\mathbf{X}} = \mathbf{X}$.

In more convenient form:

$$\mathbb{P}\left[\mathbf{E} = 1\right] \leq \frac{\mathbb{I}\left[\mathbf{X}; \hat{\mathbf{X}}\right] + \mathbb{H}\left[\mathbf{E}\right]}{\mathbb{H}\left[\mathbf{X}\right]}.$$

One more involved example: detecting a biased coin.

Suppose we have coin C, which can be fair or biased $+\varepsilon, -\varepsilon$ (each equally likely).

Task: Flip the coin to figure out whether it is biased (i.e., learn the distribution its i.i.d. flips come from).

Question: How many coin flips $\mathbf{\Pi}_i$ do we need with *any* estimation method to be correct with $\mathbb{P}[\mathbf{E} = 1] \geq \frac{2}{3}$?

From Taylor expansion: $\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_i] \leq O(\varepsilon^2)$.

With this we obtain, using Fano's inequality:

$$\frac{2}{3} \leq \frac{\mathbb{I}[\mathbf{X}; \mathbf{\Pi}] + \mathbb{H}[\mathbf{E}]}{\log 3} \leq \frac{n\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_1] + \mathbb{H}[\mathbf{E}]}{\log 3} \approx \frac{n/\varepsilon^2 + 0.91}{1.58} \Leftrightarrow n = \Omega(1/\varepsilon^2).$$

One more involved example: detecting a biased coin.

Suppose we have coin C, which can be fair or biased $+\varepsilon, -\varepsilon$ (each equally likely).

Task: Flip the coin to figure out whether it is biased (i.e., learn the distribution its i.i.d. flips come from).

Question: How many coin flips $\mathbf{\Pi}_i$ do we need with *any* estimation method to be correct with $\mathbb{P}[\mathbf{E} = 1] \geq \frac{2}{3}$?

From Taylor expansion: $\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_i] \leq O(\varepsilon^2)$.

With this we obtain, using Fano's inequality:

$$\frac{2}{3} \leq \frac{\mathbb{I}[\mathbf{X}; \mathbf{\Pi}] + \mathbb{H}[\mathbf{E}]}{\log 3} \leq \frac{n\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_1] + \mathbb{H}[\mathbf{E}]}{\log 3} \approx \frac{n/\varepsilon^2 + 0.91}{1.58} \Leftrightarrow n = \Omega(1/\varepsilon^2).$$

One more involved example: detecting a biased coin.

Suppose we have coin C, which can be fair or biased $+\varepsilon, -\varepsilon$ (each equally likely).

Task: Flip the coin to figure out whether it is biased (i.e., learn the distribution its i.i.d. flips come from).

Question: How many coin flips $\mathbf{\Pi}_i$ do we need with *any* estimation method to be correct with $\mathbb{P}[\mathbf{E} = 1] \geq \frac{2}{3}$?

From Taylor expansion: $\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_i] \leq O(\varepsilon^2)$.

With this we obtain, using Fano's inequality:

$$\frac{2}{3} \leq \frac{\mathbb{I}[\mathbf{X}; \mathbf{\Pi}] + \mathbb{H}[\mathbf{E}]}{\log 3} \leq \frac{n\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_1] + \mathbb{H}[\mathbf{E}]}{\log 3} \approx \frac{n/\varepsilon^2 + 0.91}{1.58} \Leftrightarrow n = \Omega(1/\varepsilon^2).$$

One more involved example: detecting a biased coin.

Suppose we have coin C, which can be fair or biased $+\varepsilon, -\varepsilon$ (each equally likely).

Task: Flip the coin to figure out whether it is biased (i.e., learn the distribution its i.i.d. flips come from).

Question: How many coin flips $\mathbf{\Pi}_i$ do we need with *any* estimation method to be correct with $\mathbb{P}[\mathbf{E} = 1] \geq \frac{2}{3}$?

From Taylor expansion: $\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_i] \leq O(\varepsilon^2)$.

With this we obtain, using Fano's inequality:

$$\frac{2}{3} \leq \frac{\mathbb{I}[\mathbf{X}; \mathbf{\Pi}] + \mathbb{H}[\mathbf{E}]}{\log 3} \leq \frac{n\mathbb{I}[\mathbf{X}; \mathbf{\Pi}_1] + \mathbb{H}[\mathbf{E}]}{\log 3} \approx \frac{n/\varepsilon^2 + 0.91}{1.58} \Leftrightarrow n = \Omega(1/\varepsilon^2).$$

One more involved example: detecting a biased coin.

Suppose we have coin C, which can be fair or biased $+\varepsilon, -\varepsilon$ (each equally likely).

Task: Flip the coin to figure out whether it is biased (i.e., learn the distribution its i.i.d. flips come from).

Question: How many coin flips $\mathbf{\Pi}_i$ do we need with *any* estimation method to be correct with $\mathbb{P}\left[\mathbf{E} = 1\right] \geq \frac{2}{3}$?

From Taylor expansion: $\mathbb{I}\left[\mathbf{X}; \mathbf{\Pi}_i\right] \leq O(\varepsilon^2)$.

With this we obtain, using Fano's inequality:

$$\frac{2}{3} \leq \frac{\mathbb{I}\left[\mathbf{X}; \mathbf{\Pi}\right] + \mathbb{H}\left[\mathbf{E}\right]}{\log 3} \leq \frac{n\mathbb{I}\left[\mathbf{X}; \mathbf{\Pi}_1\right] + \mathbb{H}\left[\mathbf{E}\right]}{\log 3} \approx \frac{n/\varepsilon^2 + 0.91}{1.58} \Leftrightarrow n = \Omega(1/\varepsilon^2).$$

Information Theory + Extended Formulations

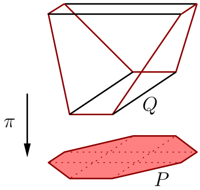
— Part 2 —

Sebastian Pokutta

Extended formulations - quick recap.

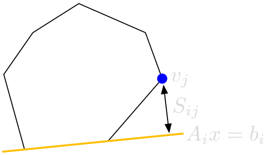
Definition (extension)

P, Q polytopes. Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

$\text{size}(Q) :=$ #facets of Q
$\text{xc}(P) := \min\{\text{size}(Q) \mid Q \text{ extension of } P\}$



Theorem (factorization thm [Yan.'91])

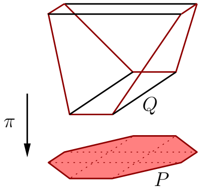
For every slack matrix S of P:

$$\text{xc}(P) = \text{rk}_+(S)$$

Extended formulations - quick recap.

Definition (extension)

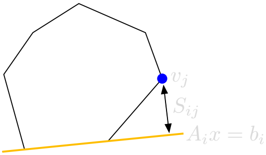
P, Q polytopes. Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

$\mathrm{size}(Q) :=$ #facets of Q

$\mathrm{xc}(P) := \min\{\mathrm{size}(Q) \mid Q \text{ extension of } P\}$



Theorem (factorization thm [Yan.'91])

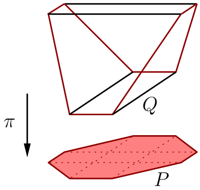
For every slack matrix S of P:

$$\mathrm{xc}(P) = \mathrm{rk}_+(S)$$

Extended formulations - quick recap.

Definition (extension)

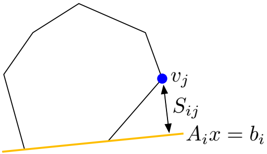
P, Q polytopes. Q is an extension of P if $\exists$ linear π with $\pi(Q) = P$



Definition (size and extension complexity)

$\mathrm{size}(Q) := \#$facets of Q

$\mathrm{xc}(P) := \min\{\mathrm{size}(Q) \mid Q \text{ extension of } P\}$



Theorem (factorization thm [Yan.'91])

For every slack matrix S of P:

$$\mathrm{xc}(P) = \mathrm{rk}_+(S)$$

Extended formulations - Sums of rank-1 matrices revisited.

Let M be a nonnegative matrix and consider a factorization

$$M = \sum_{\pi \in [r]} M_\pi$$

with M_π nonnegative rank-1 matrices.

Suppose that M is normalized so that $\sum_{a,b} M_{a,b} = 1$.
$\Rightarrow$ M is highly complicated probability distribution of (a, b)-pairs.
As distribution: $(\mathbf{A}, \mathbf{B}) \sim M / \|M\|_1$.

We want to sample from M via a set of product distributions.
$\Rightarrow$ Information has to go into the distribution of π.

Extended formulations - Sums of rank-1 matrices revisited.

Let M be a nonnegative matrix and consider a factorization

$$M = \sum_{\pi \in [r]} M_\pi$$

with M_π nonnegative rank-1 matrices.

Suppose that M is normalized so that $\sum_{a,b} M_{a,b} = 1$.
$\Rightarrow M$ is highly complicated probability distribution of (a, b)-pairs.
As distribution: $(\mathbf{A}, \mathbf{B}) \sim M / \|M\|_1$.

We want to sample from M via a set of product distributions.
$\Rightarrow$ Information has to go into the distribution of π.

Extended formulations - Sums of rank-1 matrices revisited.

Let M be a nonnegative matrix and consider a factorization

$$M = \sum_{\pi \in [r]} M_\pi$$

with M_π nonnegative rank-1 matrices.

Suppose that M is normalized so that $\sum_{a,b} M_{a,b} = 1$.
$\Rightarrow M$ is highly complicated probability distribution of (a, b)-pairs.
As distribution: $(\mathbf{A}, \mathbf{B}) \sim M / \|M\|_1$.

We want to sample from M via a set of product distributions.
$\Rightarrow$ Information has to go into the distribution of π.

Extended formulations - Sums of rank-1 matrices revisited.

Let M be a nonnegative matrix and consider a factorization

$$M = \sum_{\pi \in [r]} M_\pi$$

with M_π nonnegative rank-1 matrices.

Suppose that M is normalized so that $\sum_{a,b} M_{a,b} = 1$.
$\Rightarrow M$ is highly complicated probability distribution of (a, b)-pairs.
As distribution: $(\mathbf{A}, \mathbf{B}) \sim M / \|M\|_1$.

We want to sample from M via a set of product distributions.
$\Rightarrow$ Information has to go into the distribution of π.

Lemma (Matrices to distributions)

Let M be nonnegative and $(\mathbf{A}, \mathbf{B})$ be a random (row,col) of M, with

$$\mathbb{P}[\mathbf{A} = a, \mathbf{B} = b] = \frac{M(a,b)}{\sum_{x,y} M(x,y)}$$

Then $\exists$ discrete random variable $\mathbf{\Pi}$ with

1. $\mathbf{A}$ *and* $\mathbf{B}$ *are conditionally independent given* $\mathbf{\Pi}$,
2. $\mathbf{\Pi}$ *takes* $\mathrm{rk}_+(M)$ *distinct values.*

In particular, $\mathrm{rk}_+(M) \geq 2^{\mathbb{H}[\mathbf{\Pi}]}$.

Lemma (Matrices to distributions)

Let M be nonnegative and $(\mathbf{A}, \mathbf{B})$ be a random (row,col) of M, with

$$\mathbb{P}\left[\mathbf{A} = a, \mathbf{B} = b\right] = \frac{M(a,b)}{\sum_{x,y} M(x,y)}$$

Then $\exists$ discrete random variable $\mathbf{\Pi}$ with

1. *$\mathbf{A}$ and $\mathbf{B}$ are conditionally independent given $\mathbf{\Pi}$,*
2. *$\mathbf{\Pi}$ takes $\mathrm{rk}_+(M)$ distinct values.*

In particular, $\mathrm{rk}_+(M) \geq 2^{\mathbb{H}[\mathbf{\Pi}]}$.

Proof sketch.
Let a minimal factorization of M be given by

$$M(a,b) = \sum_{\pi} \alpha_\pi(a)\beta_\pi(b).$$

(1) Let $\mathbf{\Pi}$ be a RV running through $\pi \Rightarrow \mathrm{rk}_+(M)$ values.

(2) Define a new distribution of $\mathbf{A}, \mathbf{B}, \mathbf{\Pi}$ via

$$\mathbb{P}\left[\mathbf{A} = a, \mathbf{B} = b, \mathbf{\Pi} = \pi\right] = \frac{\alpha_\pi(a)\beta_\pi(b)}{\sum_{x,y} M(x,y)}.$$

Sum over π to verify that the distributions coincide for $(\mathbf{A}, \mathbf{B})$. Note that the product in the numerator ensures independence of $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$. $\square$

Proof sketch.
Let a minimal factorization of M be given by

$$M(a,b) = \sum_{\pi} \alpha_\pi(a)\beta_\pi(b).$$

(1) Let $\mathbf{\Pi}$ be a RV running through $\pi \Rightarrow \mathrm{rk}_+(M)$ values.

(2) Define a new distribution of $\mathbf{A}, \mathbf{B}, \mathbf{\Pi}$ via

$$\mathbb{P}\left[\mathbf{A} = a, \mathbf{B} = b, \mathbf{\Pi} = \pi\right] = \frac{\alpha_\pi(a)\beta_\pi(b)}{\sum_{x,y} M(x,y)}.$$

Sum over π to verify that the distributions coincide for $(\mathbf{A}, \mathbf{B})$. Note that the product in the numerator ensures independence of $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$. $\square$

Lemma (Cut-and-paste property for NMF)

Let M be nonnegative and $(\mathbf{A}, \mathbf{B}) \sim M$ with $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$. Then with $\mathbf{\Pi}_{a,b} := \mathbf{\Pi} | \mathbf{A} = a, \mathbf{B} = b$ we have:

$$\begin{aligned} &\sqrt{M(a_1, b_1) M(a_2, b_2)} \left(1 - h^2(\mathbf{\Pi}_{a_1,b_1}; \mathbf{\Pi}_{a_2,b_2})\right) \\ =&\sqrt{M(a_1, b_2) M(a_2, b_1)} \left(1 - h^2(\mathbf{\Pi}_{a_1,b_2}; \mathbf{\Pi}_{a_2,b_1})\right). \end{aligned}$$

In particular,

$$h^2(\mathbf{\Pi}_{a_1,b_1}; \mathbf{\Pi}_{a_2,b_2}) \geq 1 - \sqrt{\frac{M(a_1, b_2) M(a_2, b_1)}{M(a_1, b_1) M(a_2, b_2)}}.$$

Note: We care for distribution of $\mathbf{\Pi}$ conditioned on $\mathbf{A} = a, \mathbf{B} = b$ (and not vice versa). Allows us to beat traditional cut-and-paste.

Lemma (Cut-and-paste property for NMF)

Let M be nonnegative and $(\mathbf{A}, \mathbf{B}) \sim M$ with $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$. Then with $\mathbf{\Pi}_{a,b} := \mathbf{\Pi} | \mathbf{A} = a, \mathbf{B} = b$ we have:

$$\begin{aligned} &\sqrt{M(a_1, b_1) M(a_2, b_2)} \left(1 - h^2(\mathbf{\Pi}_{a_1,b_1}; \mathbf{\Pi}_{a_2,b_2})\right) \\ =&\sqrt{M(a_1, b_2) M(a_2, b_1)} \left(1 - h^2(\mathbf{\Pi}_{a_1,b_2}; \mathbf{\Pi}_{a_2,b_1})\right). \end{aligned}$$

In particular,

$$h^2(\mathbf{\Pi}_{a_1,b_1}; \mathbf{\Pi}_{a_2,b_2}) \geq 1 - \sqrt{\frac{M(a_1, b_2) M(a_2, b_1)}{M(a_1, b_1) M(a_2, b_2)}}.$$

Note: We care for distribution of $\mathbf{\Pi}$ conditioned on $\mathbf{A} = a, \mathbf{B} = b$ (and not vice versa). Allows us to beat traditional cut-and-paste.

Proof sketch (cut-and-paste property).
We have the distributions $\mathbf{\Pi}_{a,b}$ via:

$$\mathbf{\Pi}_{a,b}(\pi) = \begin{cases} \frac{\alpha_\pi(a)\beta_\pi(b)}{M(a,b)}, \quad \pi \in \mathbf{\Pi} & \text{for } M(a,b) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, for all rows a_1, a_2 and columns b_1, b_2:

$$\begin{aligned} & M(a_1,b_1)\mathbf{\Pi}_{a_1,b_1}(\pi) \cdot M(a_2,b_2)\mathbf{\Pi}_{a_2,b_2}(\pi) \\ = {} & M(a_1,b_2)\mathbf{\Pi}_{a_1,b_2}(\pi) \cdot M(a_2,b_1)\mathbf{\Pi}_{a_2,b_1}(\pi), \quad \pi \in \mathbf{\Pi}. \end{aligned}$$

Taking square root and summing up

$$\begin{aligned} & \sqrt{M(a_1,b_1)M(a_2,b_2)}\left(1 - h^2(\mathbf{\Pi}_{a_1,b_1};\mathbf{\Pi}_{a_2,b_2})\right) \\ = {} & \sqrt{M(a_1,b_2)M(a_2,b_1)}\left(1 - h^2(\mathbf{\Pi}_{a_1,b_2};\mathbf{\Pi}_{a_2,b_1})\right) \leq \sqrt{M(a_1,b_2)M(a_2,b_1)}. \end{aligned}$$

It also follows (assuming $M(a_1,b_1), M(a_2,b_2) > 0$)

$$h^2(\mathbf{\Pi}_{a_1,b_1};\mathbf{\Pi}_{a_2,b_2}) \geq 1 - \sqrt{\frac{M(a_1,b_2)M(a_2,b_1)}{M(a_1,b_1)M(a_2,b_2)}}.$$

□

Extended formulations - Common information and NMF.

Common information [Wyner, 75]

$$\mathbb{C}[M] := \min_{\mathbf{\Pi}: \mathbf{A} \perp \mathbf{B} | \mathbf{\Pi}} \mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}\right],$$

where $(\mathbf{A}, \mathbf{B}) \sim M / \|M\|_1$.

Common information captures the information about the correlation: once provided as seed, the sampling is independent.

Clearly,

$$\mathbb{C}[M] \leq \min_{\mathbf{\Pi}: \mathbf{A} \perp \mathbf{B} | \mathbf{\Pi}} \mathbb{H}\left[\mathbf{\Pi}\right] \leq \log \operatorname{rk}_+ M$$

Note: While useful, needs some adjustments for partial matrices and $\mathbb{C}[.]$ is not necessarily monotone under conditioning.

Extended formulations - Common information and NMF.

Common information [Wyner, 75]

$$\mathbb{C}[M] := \min_{\mathbf{\Pi}:\mathbf{A}\perp\mathbf{B}|\mathbf{\Pi}} \mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi}\right],$$

where $(\mathbf{A},\mathbf{B}) \sim M/\left\|M\right\|_1$.

Common information captures the information about the correlation: once provided as seed, the sampling is independent.

Clearly,

$$\mathbb{C}[M] \leq \min_{\mathbf{\Pi}:\mathbf{A}\perp\mathbf{B}|\mathbf{\Pi}} \mathbb{H}\left[\mathbf{\Pi}\right] \leq \log \operatorname{rk}_+ M$$

Note: While useful, needs some adjustments for partial matrices and $\mathbb{C}[.]$ is not necessarily monotone under conditioning.

Extended formulations - Conditioned Common information.

Conditioned Common information

$$\mathbb{C}[M \mid \mathfrak{Z}] := \min_{\substack{\mathbf{\Pi}: \mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi} \\ \mathbf{\Pi} \perp \mathfrak{Z} \mid (\mathbf{A}, \mathbf{B})}} \mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathfrak{Z}\right],$$

where $\mathfrak{Z}$ is a conditional.

Independence so that $\mathbf{\Pi}$ does not learn from conditional $\mathfrak{Z}$: a real factorization would not either.

Still

$$\mathbb{C}[M \mid \mathfrak{Z}] \leq \min_{\substack{\mathbf{\Pi}: \mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi} \\ \mathbf{\Pi} \perp \mathfrak{Z} \mid (\mathbf{A}, \mathbf{B})}} \mathbb{H}\left[\mathbf{\Pi} \mid \mathfrak{Z}\right] \leq \log \operatorname{rk}_+ M$$

Why? Allows us to fine-tune the distribution and deal with partial matrices.

Extended formulations - Conditioned Common information.

Conditioned Common information

$$\mathbb{C}[M \mid \mathcal{Z}] := \min_{\substack{\mathbf{\Pi}:\mathbf{A}\perp\mathbf{B}\mid\mathbf{\Pi} \\ \mathbf{\Pi}\perp\mathcal{Z}\mid(\mathbf{A},\mathbf{B})}} \mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi} \mid \mathcal{Z}\right],$$

where $\mathcal{Z}$ is a conditional.

Independence so that $\mathbf{\Pi}$ does not learn from conditional $\mathcal{Z}$: a real factorization would not either.

Still

$$\mathbb{C}[M \mid \mathcal{Z}] \leq \min_{\substack{\mathbf{\Pi}:\mathbf{A}\perp\mathbf{B}\mid\mathbf{\Pi} \\ \mathbf{\Pi}\perp\mathcal{Z}\mid(\mathbf{A},\mathbf{B})}} \mathbb{H}\left[\mathbf{\Pi} \mid \mathcal{Z}\right] \leq \log \operatorname{rk}_+ M$$

Why? Allows us to fine-tune the distribution and deal with partial matrices.

Extended formulations - Analysis of Common Information.

Lower bounds are now obtained via analyzing $\mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}]$.

General strategy:

1. Identify conditional $\mathfrak{Z}$, so that $\mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}]$ can be decomposed:

$$\mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathfrak{Z}] \geq \sum_{i \in [\ell]} \mathbb{I}[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}] \geq \ell \min_i \mathbb{I}[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}].$$

2. For subproblem, use polyhedral combinatorics to bound:

$$\mathbb{I}[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}] \geq \varepsilon$$

Nice side effects: we automatically also get polyhedral inapproximability results.

Extended formulations - Analysis of Common Information.

Lower bounds are now obtained via analyzing $\mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi}\right]$.

General strategy:

1. Identify conditional $\mathfrak{Z}$, so that $\mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi}\right]$ can be decomposed:

$$\mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi}\,|\,\mathfrak{Z}\right] \geq \sum_{i\in[\ell]} \mathbb{I}\left[\mathbf{A}_i,\mathbf{B}_i;\mathbf{\Pi}\,|\,\mathfrak{Z}\right] \geq \ell \min_i \mathbb{I}\left[\mathbf{A}_i,\mathbf{B}_i;\mathbf{\Pi}\,|\,\mathfrak{Z}\right].$$

2. For subproblem, use polyhedral combinatorics to bound:

$$\mathbb{I}\left[\mathbf{A}_i,\mathbf{B}_i;\mathbf{\Pi}\,|\,\mathfrak{Z}\right] \geq \varepsilon$$

Nice side effects: we automatically also get polyhedral inapproximability results.

Extended formulations - Analysis of Common Information.

Lower bounds are now obtained via analyzing $\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}\right]$.

General strategy:

1. Identify conditional $\mathfrak{Z}$, so that $\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}\right]$ can be decomposed:

$$\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathfrak{Z}\right] \geq \sum_{i \in [\ell]} \mathbb{I}\left[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}\right] \geq \ell \min_i \mathbb{I}\left[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}\right].$$

2. For subproblem, use polyhedral combinatorics to bound:

$$\mathbb{I}\left[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}\right] \geq \varepsilon$$

Nice side effects: we automatically also get polyhedral inapproximability results.

Extended formulations - Analysis of Common Information.

Lower bounds are now obtained via analyzing $\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}\right]$.

General strategy:

1. Identify conditional $\mathfrak{Z}$, so that $\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi}\right]$ can be decomposed:

$$\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathfrak{Z}\right] \geq \sum_{i \in [\ell]} \mathbb{I}\left[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}\right] \geq \ell \min_i \mathbb{I}\left[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}\right].$$

2. For subproblem, use polyhedral combinatorics to bound:

$$\mathbb{I}\left[\mathbf{A}_i, \mathbf{B}_i; \mathbf{\Pi} \mid \mathfrak{Z}\right] \geq \varepsilon$$

Nice side effects: we automatically also get polyhedral inapproximability results.

The Correlation Polytope

Correlation polytope: $\text{COR}(n) := \text{conv}\{bb^T \in \mathbb{R}^{n\times n} \mid b \in \{0,1\}^n\}$

Observation. For $a, b \in \{0,1\}^n$:

$$\begin{aligned}1 - \langle 2\text{diag}(a) - aa^T, bb^T\rangle &= 1 - 2\langle \text{diag}(a), bb^T\rangle + \langle aa^T, bb^T\rangle \\ &= 1 - 2\langle \text{diag}(a), \text{diag}(b)\rangle + \langle aa^T, bb^T\rangle \\ &= 1 - 2\, a^T b + (a^T b)^2 = (1 - a^T b)^2 =: M_{ab}\end{aligned}$$

Lemma (Key Lemma)

For every $a \in \{0,1\}^n$, the inequality

$$(\star) \qquad \langle 2\text{diag}(a) - aa^T, x\rangle \leqslant 1$$

is valid for COR(n). The slack of vertex bb^T w.r.t. ($\star$) is M_{ab}.

Note: (A variant of) the clique problem reduces to COR(n).

Correlation polytope: $\mathrm{COR}(n) := \mathrm{conv}\{bb^T \in \mathbb{R}^{n\times n} \mid b \in \{0,1\}^n\}$

Observation. For $a, b \in \{0,1\}^n$:

$$\begin{aligned} 1 - \langle 2\mathrm{diag}(a) - aa^T, bb^T\rangle &= 1 - 2\langle \mathrm{diag}(a), bb^T\rangle + \langle aa^T, bb^T\rangle \\ &= 1 - 2\langle \mathrm{diag}(a), \mathrm{diag}(b)\rangle + \langle aa^T, bb^T\rangle \\ &= 1 - 2\,a^Tb + (a^Tb)^2 = (1 - a^Tb)^2 =: M_{ab} \end{aligned}$$

Lemma (Key Lemma)

For every $a \in \{0,1\}^n$, the inequality

$(\star)$ $$\langle 2\mathit{diag}(a) - aa^T, x\rangle \leqslant 1$$

is valid for COR(n). The slack of vertex bb^T w.r.t. ($\star$) is M_{ab}.

Note: (A variant of) the clique problem reduces to COR(n).

Correlation polytope: $\mathrm{COR}(n) := \mathrm{conv}\{bb^T \in \mathbb{R}^{n\times n} \mid b \in \{0,1\}^n\}$

Observation. For $a, b \in \{0,1\}^n$:

$$\begin{aligned} 1 - \langle 2\mathrm{diag}(a) - aa^T, bb^T\rangle &= 1 - 2\langle \mathrm{diag}(a), bb^T\rangle + \langle aa^T, bb^T\rangle \\ &= 1 - 2\langle \mathrm{diag}(a), \mathrm{diag}(b)\rangle + \langle aa^T, bb^T\rangle \\ &= 1 - 2\,a^Tb + (a^Tb)^2 = (1 - a^Tb)^2 =: M_{ab} \end{aligned}$$

Lemma (Key Lemma)

For every $a \in \{0,1\}^n$, the inequality

$(\star)$ $$\langle 2\mathit{diag}(a) - aa^T, x\rangle \leqslant 1$$

is valid for COR(n). The slack of vertex bb^T w.r.t. ($\star$) is M_{ab}.

Note: (A variant of) the clique problem reduces to $\mathrm{COR}(n)$.

The slack matrix of the correlation polytope contains the so called UDISJ (partial) matrix $M \in \mathbb{R}_+^{2^n} \times \mathbb{R}_+^{2^n}$

$$M(a,b) = \begin{cases} 1 & \text{if} \quad |a \cap b| = 0 \\ 0 & \text{if} \quad |a \cap b| = 1. \end{cases}$$

Slack matrices of approximations of the correlation polytope contain its shift $M_\rho \in \mathbb{R}_+^{2^n} \times \mathbb{R}_+^{2^n}$

$$M_\rho(a,b) = \begin{cases} \rho & \text{if} \quad |a \cap b| = 0 \\ \rho - 1 & \text{if} \quad |a \cap b| = 1. \end{cases}$$

The slack matrix of the correlation polytope contains the so called UDISJ (partial) matrix $M \in \mathbb{R}_+^{2^n} \times \mathbb{R}_+^{2^n}$

$$M(a,b) = \begin{cases} 1 & \text{if} \quad |a \cap b| = 0 \\ 0 & \text{if} \quad |a \cap b| = 1. \end{cases}$$

Slack matrices of approximations of the correlation polytope contain its shift $M_\rho \in \mathbb{R}_+^{2^n} \times \mathbb{R}_+^{2^n}$

$$M_\rho(a,b) = \begin{cases} \rho & \text{if} \quad |a \cap b| = 0 \\ \rho - 1 & \text{if} \quad |a \cap b| = 1. \end{cases}$$

[de Wolf, 01] via [Razborov, 92]: $\mathrm{rk}_+ M \geq 2^{\Omega(n)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be captured by poly size LP.

[Braun, Fiorini, P., Steurer, 12]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-2\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1/2-\varepsilon}$ (similarly, PSD cone cannot be approximated).

[Braverman, Moitra, 13]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ matching Håstad's bound for $\mathrm{CLIQUE}(n)$.

[Braun, P., 13]: $\mathrm{rk}_+ \tilde{M}(n^\beta) \geq 2^{\Omega(n^{1-\beta})-o(1)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ in an average case sense.

[de Wolf, 01] via [Razborov, 92]: $\mathrm{rk}_+ M \geq 2^{\Omega(n)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be captured by poly size LP.

[Braun, Fiorini, P., Steurer, 12]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-2\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1/2-\varepsilon}$ (similarly, PSD cone cannot be approximated).

[Braverman, Moitra, 13]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ matching Håstad's bound for $\mathrm{CLIQUE}(n)$.

[Braun, P., 13]: $\mathrm{rk}_+ \tilde{M}(n^\beta) \geq 2^{\Omega(n^{1-\beta})-o(1)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ in an average case sense.

[de Wolf, 01] via [Razborov, 92]: $\mathrm{rk}_+ M \geq 2^{\Omega(n)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be captured by poly size LP.

[Braun, Fiorini, P., Steurer, 12]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-2\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1/2-\varepsilon}$ (similarly, PSD cone cannot be approximated).

[Braverman, Moitra, 13]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ matching Håstad's bound for $\mathrm{CLIQUE}(n)$.

[Braun, P., 13]: $\mathrm{rk}_+ \tilde{M}(n^\beta) \geq 2^{\Omega(n^{1-\beta})-o(1)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ in an average case sense.

[de Wolf, 01] via [Razborov, 92]: $\mathrm{rk}_+ M \geq 2^{\Omega(n)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be captured by poly size LP.

[Braun, Fiorini, P., Steurer, 12]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-2\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1/2-\varepsilon}$ (similarly, PSD cone cannot be approximated).

[Braverman, Moitra, 13]: $\mathrm{rk}_+ M_{n^\beta} \geq 2^{\Omega(n^{1-\beta})}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ matching Håstad's bound for $\mathrm{CLIQUE}(n)$.

[Braun, P., 13]: $\mathrm{rk}_+ \tilde{M}(n^\beta) \geq 2^{\Omega(n^{1-\beta})-o(1)}$

$\Rightarrow \mathrm{COR}(n)$ cannot be approximated by poly size LP within a factor of $n^{1-\varepsilon}$ in an average case sense.

Crossing over into numbers—our key estimations.

Pinsker's inequality: Let $\mathbf{A}, \mathbf{B}$ be discrete RVs with identical range. Then

$$\mathrm{D}\left(\mathbf{A} \,\|\, \mathbf{B}\right) \geq \frac{\log e}{2} \left\|p_{\mathbf{A}} - p_{\mathbf{B}}\right\|_1^2 = 2(\log e) \left(\max_{\mathcal{E}\text{: event}} \left|p_{\mathbf{A}}(\mathcal{E}) - p_{\mathbf{B}}(\mathcal{E})\right|\right)^2$$

Hellinger Distance: Let $\mathbf{A}, \mathbf{B}$ be discrete RVs with identical range. Then

$$\begin{aligned} h^2(\mathbf{A}; \mathbf{B}) &\coloneqq 1 - \sum_{a \in \operatorname{Range} \mathbf{A}} \sqrt{p_{\mathbf{A}}(a) p_{\mathbf{B}}(a)} \\ &= \frac{1}{2} \left\|\sqrt{p_{\mathbf{A}}} - \sqrt{p_{\mathbf{B}}}\right\|_2^2 \geq 0. \end{aligned}$$

Information-theoretic setup.

Note: Overall strategy similar to Bar-Yossef et al.

- Let $\mathbf{A}$, $\mathbf{B}$ be random subsets of $[n]$ conditionally independent given $\mathbf{\Pi}$ with $\mathbf{A}_i$ and $\mathbf{B}_i$ indicating $i \in \mathbf{A}$, $i \in \mathbf{B}$.
- Write the UDISJ distribution as

$$\mathbb{P}[\mathbf{A} = a, \mathbf{B} = b] = \begin{cases} c & \text{if } a \cap b = \emptyset \\ c(1-\varepsilon) & \text{if } |a \cap b| = 1 \end{cases}$$

- Take n fair coins $\mathbf{C}_1, \dots, \mathbf{C}_n$ independent of $\mathbf{A}, \mathbf{B}, \Pi$.
- New RVs $\mathbf{D}_1, \dots, \mathbf{D}_n$ with $\mathbf{D}_i = \mathbf{A}_i$ if $\mathbf{C}_i = 0$ and $\mathbf{D}_i = \mathbf{B}_i$ otherwise. Short: $\mathbf{D} := (\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n)$
- We will prove for any $\mathbf{\Pi}$ such that $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$

$$\mathbb{H}[\mathbf{\Pi}] \geq \mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \frac{\varepsilon n}{8}.$$

Information-theoretic setup.

Note: Overall strategy similar to Bar-Yossef et al.

- Let $\mathbf{A}$, $\mathbf{B}$ be random subsets of $[n]$ conditionally independent given $\mathbf{\Pi}$ with $\mathbf{A}_i$ and $\mathbf{B}_i$ indicating $i \in \mathbf{A}$, $i \in \mathbf{B}$.
- Write the UDISJ distribution as

$$\mathbb{P}\left[\mathbf{A} = a, \mathbf{B} = b\right] = \begin{cases} c & \text{if } a \cap b = \emptyset \\ c(1-\varepsilon) & \text{if } |\, a \cap b \,| = 1 \end{cases}$$

- Take n fair coins $\mathbf{C}_1, \dots, \mathbf{C}_n$ independent of $\mathbf{A}, \mathbf{B}, \Pi$.
- New RVs $\mathbf{D}_1, \dots, \mathbf{D}_n$ with $\mathbf{D}_i = \mathbf{A}_i$ if $\mathbf{C}_i = 0$ and $\mathbf{D}_i = \mathbf{B}_i$ otherwise. Short: $\mathbf{D} := (\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n)$
- We will prove for any $\mathbf{\Pi}$ such that $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$

$$\mathbb{H}\left[\mathbf{\Pi}\right] \geq \mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}\right] \geq \frac{\varepsilon n}{8}.$$

Information-theoretic setup.

Note: Overall strategy similar to Bar-Yossef et al.

- Let $\mathbf{A}$, $\mathbf{B}$ be random subsets of $[n]$ conditionally independent given $\mathbf{\Pi}$ with $\mathbf{A}_i$ and $\mathbf{B}_i$ indicating $i \in \mathbf{A}$, $i \in \mathbf{B}$.
- Write the UDISJ distribution as

$$\mathbb{P}[\mathbf{A} = a, \mathbf{B} = b] = \begin{cases} c & \text{if } a \cap b = \emptyset \\ c(1-\varepsilon) & \text{if } |a \cap b| = 1 \end{cases}$$

- Take n fair coins $\mathbf{C}_1, \ldots, \mathbf{C}_n$ independent of $\mathbf{A}, \mathbf{B}, \Pi$.
- New RVs $\mathbf{D}_1, \ldots, \mathbf{D}_n$ with $\mathbf{D}_i = \mathbf{A}_i$ if $\mathbf{C}_i = 0$ and $\mathbf{D}_i = \mathbf{B}_i$ otherwise. Short: $\mathbf{D} := (\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n)$
- We will prove for any $\mathbf{\Pi}$ such that $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$

$$\mathbb{H}[\mathbf{\Pi}] \geq \mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \frac{\varepsilon n}{8}.$$

Information-theoretic setup.

Note: Overall strategy similar to Bar-Yossef et al.

- Let $\mathbf{A}$, $\mathbf{B}$ be random subsets of $[n]$ conditionally independent given $\mathbf{\Pi}$ with $\mathbf{A}_i$ and $\mathbf{B}_i$ indicating $i \in \mathbf{A}$, $i \in \mathbf{B}$.
- Write the UDISJ distribution as

$$\mathbb{P}[\mathbf{A} = a, \mathbf{B} = b] = \begin{cases} c & \text{if } a \cap b = \emptyset \\ c(1-\varepsilon) & \text{if } |\, a \cap b \,| = 1 \end{cases}$$

- Take n fair coins $\mathbf{C}_1, \ldots, \mathbf{C}_n$ independent of $\mathbf{A}, \mathbf{B}, \Pi$.
- New RVs $\mathbf{D}_1, \ldots, \mathbf{D}_n$ with $\mathbf{D}_i = \mathbf{A}_i$ if $\mathbf{C}_i = 0$ and $\mathbf{D}_i = \mathbf{B}_i$ otherwise. Short: $\mathbf{D} := (\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n)$
- We will prove for any $\mathbf{\Pi}$ such that $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$

$$\mathbb{H}[\mathbf{\Pi}] \geq \mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \frac{\varepsilon n}{8}.$$

Information-theoretic setup.

Note: Overall strategy similar to Bar-Yossef et al.

- Let $\mathbf{A}$, $\mathbf{B}$ be random subsets of $[n]$ conditionally independent given $\mathbf{\Pi}$ with $\mathbf{A}_i$ and $\mathbf{B}_i$ indicating $i \in \mathbf{A}$, $i \in \mathbf{B}$.
- Write the UDISJ distribution as

$$\mathbb{P}[\mathbf{A} = a, \mathbf{B} = b] = \begin{cases} c & \text{if } a \cap b = \emptyset \\ c(1-\varepsilon) & \text{if } |\, a \cap b \,| = 1 \end{cases}$$

- Take n fair coins $\mathbf{C}_1, \dots, \mathbf{C}_n$ independent of $\mathbf{A}, \mathbf{B}, \Pi$.
- New RVs $\mathbf{D}_1, \dots, \mathbf{D}_n$ with $\mathbf{D}_i = \mathbf{A}_i$ if $\mathbf{C}_i = 0$ and $\mathbf{D}_i = \mathbf{B}_i$ otherwise. Short: $\mathbf{D} := (\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n)$
- We will prove for any $\mathbf{\Pi}$ such that $\mathbf{A} \perp \mathbf{B} \mid \mathbf{\Pi}$

$$\mathbb{H}[\mathbf{\Pi}] \geq \mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \frac{\varepsilon n}{8}.$$

Reduction to case $n = 1$.

Note that the pairs $\{(\mathbf{A}_j, \mathbf{B}_j) : j \in [n]\}$ are independent given $\mathbf{D} = 0, \mathbf{C}$, and hence

$$\mathbb{I}[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \sum_{j \in [n]} \mathbb{I}[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}]$$

Observe that the distribution of $\mathbf{A}_j, \mathbf{B}_j, \mathbf{\Pi}, \mathbf{D}_j, \mathbf{C}_j$ given $\mathbf{D}_i = 0, \mathbf{C}_i : i \neq j$ satisfies the assumptions for the case $n = 1$ (possibly with a modified c). Thus

$$\mathbb{I}[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \frac{\varepsilon}{8},$$

so that

$$\sum_{j \in [n]} \mathbb{I}[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \mid \mathbf{D} = 0, \mathbf{C}] \geq \frac{\varepsilon n}{8}$$

It remains to prove the case $n = 1$.

Reduction to case $n = 1$.

Note that the pairs $\{(\mathbf{A}_j, \mathbf{B}_j) : j \in [n]\}$ are independent given $\mathbf{D} = 0, \mathbf{C}$, and hence

$$\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] \geq \sum_{j \in [n]} \mathbb{I}\left[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right]$$

Observe that the distribution of $\mathbf{A}_j, \mathbf{B}_j, \mathbf{\Pi}, \mathbf{D}_j, \mathbf{C}_j$ given $\mathbf{D}_i = 0, \mathbf{C}_i : i \neq j$ satisfies the assumptions for the case $n = 1$ (possibly with a modified c). Thus

$$\mathbb{I}\left[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] \geq \frac{\varepsilon}{8},$$

so that

$$\sum_{j \in [n]} \mathbb{I}\left[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] \geq \frac{\varepsilon n}{8}$$

It remains to prove the case $n = 1$.

Reduction to case $n = 1$.

Note that the pairs $\{(\mathbf{A}_j, \mathbf{B}_j) : j \in [n]\}$ are independent given $\mathbf{D} = 0, \mathbf{C}$, and hence

$$\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] \geq \sum_{j \in [n]} \mathbb{I}\left[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right]$$

Observe that the distribution of $\mathbf{A}_j, \mathbf{B}_j, \mathbf{\Pi}, \mathbf{D}_j, \mathbf{C}_j$ given $\mathbf{D}_i = 0, \mathbf{C}_i : i \neq j$ satisfies the assumptions for the case $n = 1$ (possibly with a modified c). Thus

$$\mathbb{I}\left[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] \geq \frac{\varepsilon}{8},$$

so that

$$\sum_{j \in [n]} \mathbb{I}\left[\mathbf{A}_j, \mathbf{B}_j; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] \geq \frac{\varepsilon n}{8}$$

It remains to prove the case $n = 1$.

The case $n = 1$.

$$\mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi} \,|\, \mathbf{D}=0,\mathbf{C}\right] = \frac{\mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi} \,|\, \mathbf{A}_1=0\right] + \mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi} \,|\, \mathbf{B}_1=0\right]}{2}$$

Let $\mathbf{\Pi}_{ab}$ denote the distribution of $\mathbf{\Pi}$ given $\mathbf{A}_1 = a$ and $\mathbf{B}_1 = b$. As $\mathbf{A}_1, \mathbf{B}_1$ is a uniform binary variable given either $\mathbf{A}_1 = 0$ or $\mathbf{B}_1 = 0$ via Bar-Yossef et al. lemma:

$$\mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi} \,|\, \mathbf{A}_1=0\right] \geq h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{01}),$$
$$\mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi} \,|\, \mathbf{B}_1=0\right] \geq h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{10}).$$

Not a good idea: separate estimation. $h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{01}) = 0$ possible as $00, 01$ can be in the same rank-1 factor. Similar for $h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{10})$.

The case $n = 1$.

$$\mathbb{I}\left[\mathbf{A}, \mathbf{B}; \mathbf{\Pi} \,|\, \mathbf{D} = 0, \mathbf{C}\right] = \frac{\mathbb{I}\left[\mathbf{A}_1, \mathbf{B}_1; \mathbf{\Pi} \,|\, \mathbf{A}_1 = 0\right] + \mathbb{I}\left[\mathbf{A}_1, \mathbf{B}_1; \mathbf{\Pi} \,|\, \mathbf{B}_1 = 0\right]}{2}$$

Let $\mathbf{\Pi}_{ab}$ denote the distribution of $\mathbf{\Pi}$ given $\mathbf{A}_1 = a$ and $\mathbf{B}_1 = b$. As $\mathbf{A}_1, \mathbf{B}_1$ is a uniform binary variable given either $\mathbf{A}_1 = 0$ or $\mathbf{B}_1 = 0$ via Bar-Yossef et al. lemma:

$$\mathbb{I}\left[\mathbf{A}_1, \mathbf{B}_1; \mathbf{\Pi} \,|\, \mathbf{A}_1 = 0\right] \geq h^2(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{01}),$$
$$\mathbb{I}\left[\mathbf{A}_1, \mathbf{B}_1; \mathbf{\Pi} \,|\, \mathbf{B}_1 = 0\right] \geq h^2(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{10}).$$

Not a good idea: separate estimation. $h^2(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{01}) = 0$ possible as $00, 01$ can be in the same rank-1 factor. Similar for $h^2(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{10})$.

The case $n = 1$.

$$\mathbb{I}\left[\mathbf{A},\mathbf{B};\mathbf{\Pi}\,|\,\mathbf{D}=0,\mathbf{C}\right] = \frac{\mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi}\,|\,\mathbf{A}_1=0\right] + \mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi}\,|\,\mathbf{B}_1=0\right]}{2}$$

Let $\mathbf{\Pi}_{ab}$ denote the distribution of $\mathbf{\Pi}$ given $\mathbf{A}_1 = a$ and $\mathbf{B}_1 = b$. As $\mathbf{A}_1, \mathbf{B}_1$ is a uniform binary variable given either $\mathbf{A}_1 = 0$ or $\mathbf{B}_1 = 0$ via Bar-Yossef et al. lemma:

$$\mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi}\,|\,\mathbf{A}_1=0\right] \geq h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{01}),$$
$$\mathbb{I}\left[\mathbf{A}_1,\mathbf{B}_1;\mathbf{\Pi}\,|\,\mathbf{B}_1=0\right] \geq h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{10}).$$

Not a good idea: separate estimation. $h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{01}) = 0$ possible as $00, 01$ can be in the same rank-1 factor. Similar for $h^2(\mathbf{\Pi}_{00};\mathbf{\Pi}_{10})$.

Simultaneous estimation via Cauchy-Schwarz and Δ-inequality.

$$\begin{aligned}
&\frac{\mathbb{I}\left[\mathbf{A}_1, \mathbf{B}_1; \mathbf{\Pi} \mid \mathbf{A}_1 = 0\right] + \mathbb{I}\left[\mathbf{A}_1, \mathbf{B}_1; \mathbf{\Pi} \mid \mathbf{B}_1 = 0\right]}{2} \\
&\geq \frac{h^2(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{01}) + h^2(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{10})}{2} \geq \frac{\left(h(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{01}) + h(\mathbf{\Pi}_{00}; \mathbf{\Pi}_{10})\right)^2}{4} \\
&\geq \frac{h^2(\mathbf{\Pi}_{01}; \mathbf{\Pi}_{10})}{4},
\end{aligned}$$

we simply apply cut-and-paste:

$$\begin{aligned}
\sqrt{M(a_1, b_1) M(a_2, b_2)} &\geq \sqrt{M(a_1, b_1) M(a_2, b_2)} \left(1 - h^2(\mathbf{\Pi}_{a_1, b_1}; \mathbf{\Pi}_{a_2, b_2})\right) \\
&= \sqrt{M(a_1, b_2) M(a_2, b_1)} \left(1 - h^2(\mathbf{\Pi}_{a_1, b_2}; \mathbf{\Pi}_{a_2, b_1})\right)
\end{aligned}$$

and hence

$$h^2(\mathbf{\Pi}_{01}; \mathbf{\Pi}_{10}) \geq 1 - \sqrt{\frac{M(0,0) M(1,1)}{M(0,1) M(1,0)}} \geq 1 - \sqrt{1 - \varepsilon} \geq \varepsilon / 2.$$

□

Theorem

Let $\mathbf{A}$*,* $\mathbf{B}$ *be random subsets of* $[n]$*, conditionally independent given* $\mathbf{\Pi}$*. Assume that*

$$\mathbb{P}\left[\mathbf{A} = a, \mathbf{B} = b\right] = \begin{cases} \rho & \text{if } a \cap b = \emptyset, \\ \rho - 1 & \text{if } |\, a \cap b \,| = 1 \end{cases} \tag{1}$$

for all $a, b \subseteq [n]$ *for some* $\rho \geq 1$*. Then* $\mathbb{H}\left[\mathbf{\Pi}\right] \geq \frac{n}{8\rho}$*.*

Approach is extremely robust w.r.t. changes in matrix.

Perturbation	$\log \mathrm{rk}_+ \geq$	Remarks
(0) UDISJ	$\frac{6-3\log 3}{4} n$	Optimal estimation
(1) Shifts of UDISJ	$\frac{1}{8\rho} n$	$(\rho - 1)$-shift
(2) Sets of fixed size $\frac{n}{4} + O(n^{1-\varepsilon})$	$\frac{n}{8\rho} - O(n^{1-\varepsilon})$	
Removing a fraction of rows and columns (remaining dimension indicated)		
(3) Random $2^{(1-\alpha)n} \times 2^{(1-\beta)n}$	$(\frac{1}{8\rho} - \alpha - \beta)n$	in expectation
(4) Adversarial $(1-\alpha)2^n \times (1-\beta)2^n$	$(\frac{1}{8\rho} - \alpha - \beta)n - \log 3$	removal of fractions per size
Flipping of a fraction τ of DISJ entries and NDISJ entries of (1)		
(5) Random	$\frac{1-2\tau}{8(\rho-\tau)} n - O(1)$	with high probability
(6) Adversarial	$\frac{\rho(1-10\tau)}{8(\rho-\tau)^2} n - O(1)$	with mild restrictions

The Matching Polytope

The matching problem.

We consider the matching polytope

$$P_{PM}(n) := \operatorname{conv}(\left\{\chi_M \in \mathbb{R}^{\binom{n}{2}} \,\middle|\, M \text{ is a perfect matching in } K_n\right\}).$$

Inequalities of interest for the (perfect) matching polytope:

$$Q(n) := \left\{x \in \mathbb{R}^{\binom{n}{2}} \,\middle|\, x(E[U]) \leq \frac{|U|-1}{2} \; \forall U \subseteq V : |U| \text{ odd}\right\}.$$

Folklore: PSRS for the matching problem.

For $\rho > 1$ consider the polytope

$$K_n = \left\{ x \,\middle|\, x(\delta(v)) \leq 1 \;\forall v, x(E[U]) \leq \rho \frac{|U| - 1}{2} \;\forall U : \mathsf{odd}, x \geq 0 \right\} \subseteq \rho P_M(n).$$

We have $P_{PM}(n) \subseteq K_n \subseteq \rho Q(n)$: For $U \subseteq [n]$ with odd $|U| > \frac{\rho}{\rho - 1}$:

$$\rho \frac{|U| - 1}{2} = \frac{|U| + |U|(\rho - 1) - \rho}{2} \geq \frac{|U|}{2}.$$

Thus $x(E[U]) \leq \rho \frac{|U|-1}{2}$ is dominated by $x(E[U]) \leq \frac{|U|}{2}$ which arises from positive combinations of $x(\delta(v)) \leq 1$ for $v \in U$.

$\Rightarrow K_n$ is defined by roughly $O(n^{\rho/(\rho-1)})$ inequalities

Folklore: PSRS for the matching problem.

For $\rho > 1$ consider the polytope

$$K_n = \left\{ x \,\middle|\, x(\delta(v)) \leq 1 \ \forall v, x(E[U]) \leq \rho \frac{|U| - 1}{2} \ \forall U : \text{odd}, x \geq 0 \right\} \subseteq \rho P_M(n).$$

We have $P_{PM}(n) \subseteq K_n \subseteq \rho Q(n)$: For $U \subseteq [n]$ with odd $|U| > \frac{\rho}{\rho - 1}$:

$$\rho \frac{|U| - 1}{2} = \frac{|U| + |U|(\rho - 1) - \rho}{2} \geq \frac{|U|}{2}.$$

Thus $x(E[U]) \leq \rho \frac{|U|-1}{2}$ is dominated by $x(E[U]) \leq \frac{|U|}{2}$ which arises from positive combinations of $x(\delta(v)) \leq 1$ for $v \in U$.

$\Rightarrow K_n$ is defined by roughly $O(n^{\rho/(\rho-1)})$ inequalities

Folklore: PSRS for the matching problem.

For $\rho > 1$ consider the polytope

$$K_n = \left\{ x \,\middle|\, x(\delta(v)) \leq 1 \;\forall v, x(E[U]) \leq \rho \frac{|U| - 1}{2} \;\forall U : \text{odd}, x \geq 0 \right\} \subseteq \rho P_M(n).$$

We have $P_{PM}(n) \subseteq K_n \subseteq \rho Q(n)$: For $U \subseteq [n]$ with odd $|U| > \frac{\rho}{\rho - 1}$:

$$\rho \frac{|U| - 1}{2} = \frac{|U| + |U|(\rho - 1) - \rho}{2} \geq \frac{|U|}{2}.$$

Thus $x(E[U]) \leq \rho \frac{|U|-1}{2}$ is dominated by $x(E[U]) \leq \frac{|U|}{2}$ which arises from positive combinations of $x(\delta(v)) \leq 1$ for $v \in U$.

$\Rightarrow K_n$ is defined by roughly $O(n^{\rho/(\rho-1)})$ inequalities

FPSRS for the matching polytope.

Note that $n^{\rho/(\rho-1)}$ is polynomial for any *fixed* ρ.

However, for $\rho = 1 + 1/n$ we have $n^{n\cdot(1+\frac{1}{n})} = n^{n+1} = \omega(poly(n))$.

Thus:

Matching Polytope	: exponential xc (Rothvoss 2013)
ρ-approx Matching Polytope (ρ fixed)	: polynomial xc

Does the matching polytope admit an FPSRS, i.e., (a family of) approximate linear programming formulations of size $poly(n, \frac{1}{\varepsilon})$?

FPSRS for the matching polytope.

Note that $n^{\rho/(\rho-1)}$ is polynomial for any *fixed* ρ.

However, for $\rho = 1 + 1/n$ we have $n^{n\cdot(1+\frac{1}{n})} = n^{n+1} = \omega(poly(n)$.

Thus:

Matching Polytope	: exponential xc (Rothvoss 2013)
ρ-approx Matching Polytope (ρ fixed)	: polynomial xc

Does the matching polytope admit an FPSRS, i.e., (a family of) approximate linear programming formulations of size $poly(n, \frac{1}{\varepsilon})$?

FPSRS for the matching polytope.

Note that $n^{\rho/(\rho-1)}$ is polynomial for any *fixed* ρ.

However, for $\rho = 1 + 1/n$ we have $n^{n\cdot(1+\frac{1}{n})} = n^{n+1} = \omega(poly(n)$.

Thus:

Matching Polytope	: exponential xc (Rothvoss 2013)
ρ-approx Matching Polytope (ρ fixed)	: polynomial xc

Does the matching polytope admit an FPSRS, i.e., (a family of) approximate linear programming formulations of size $poly(n, \frac{1}{\varepsilon})$?

Ruling out FPSRS for matching—setup.

Slack matrix of interest (U odd set, M matching):

$$S^{+\varepsilon}_{M,U} \coloneqq |\delta(U) \cap M| - 1 + \varepsilon.$$

Suppose NMF $S^{+\varepsilon} = \sum_{i \in [r]} a_i b_i^\intercal$ inducing (K normalization constant)

$$\mathbb{P}\left[\mathbf{M} = m, \mathbf{U} = u, \mathbf{\Pi} = i\right] = K \cdot a_i(m) b_i(u).$$

Marginal distribution of $\mathbf{M}, \mathbf{U}$ independent of factorization:

$$\mathbb{P}\left[\mathbf{M} = m, \mathbf{U} = u\right] = K \cdot S^{+\varepsilon}_{m,u}$$

Ruling out FPSRS for matching—setup.

Slack matrix of interest (U odd set, M matching):

$$S^{+\varepsilon}_{M,U} \coloneqq |\delta(U) \cap M| - 1 + \varepsilon.$$

Suppose NMF $S^{+\varepsilon} = \sum_{i \in [r]} a_i b_i^\intercal$ inducing (K normalization constant)

$$\mathbb{P}\left[\mathbf{M} = m, \mathbf{U} = u, \mathbf{\Pi} = i\right] = K \cdot a_i(m) b_i(u).$$

Marginal distribution of $\mathbf{M}, \mathbf{U}$ independent of factorization:

$$\mathbb{P}\left[\mathbf{M} = m, \mathbf{U} = u\right] = K \cdot S^{+\varepsilon}_{m,u}$$

Ruling out FPSRS for matching—setup.

Slack matrix of interest (U odd set, M matching):

$$S_{M,U}^{+\varepsilon} \coloneqq |\delta(U) \cap M| - 1 + \varepsilon.$$

Suppose NMF $S^{+\varepsilon} = \sum_{i \in [r]} a_i b_i^\intercal$ inducing (K normalization constant)

$$\mathbb{P}\left[\mathbf{M} = m, \mathbf{U} = u, \mathbf{\Pi} = i\right] = K \cdot a_i(m) b_i(u).$$

Marginal distribution of $\mathbf{M}, \mathbf{U}$ independent of factorization:

$$\mathbb{P}\left[\mathbf{M} = m, \mathbf{U} = u\right] = K \cdot S_{m,u}^{+\varepsilon}$$

Ruling out FPSRS for matching—setup.

We construct a conditional $\mathfrak{Z}$ to make the problem decompose:

1. Choose $\mathbf{H}$ 3-matching between disjoint subsets $\mathbf{C_H}$ and $\mathbf{D_H}$.
   Goal of $\mathfrak{Z}$: only pairs $(\mathbf{M}, \mathbf{U})$ with $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$ with $\mathbf{C_H} \subseteq \mathbf{U}$ and $\mathbf{U} \cap \mathbf{D_H} = \emptyset$.
2. Partition the remaining vertices not covered by $\mathbf{H}$ into chunks $\mathbf{T_1}, \ldots, \mathbf{T_m}$ of size $2(k-3)$ (put residual into $\mathbf{L}$).
3. Split $\mathbf{T_i}$ into disjoint sets $\mathbf{C_i}$ and $\mathbf{D_i}$ of size $k-3$.
4. $\mathbf{T}$ collection of $\mathbf{C_1}, \mathbf{D_1}, \ldots, \mathbf{C_m}, \mathbf{D_m}, \mathbf{C_H}, \mathbf{D_H}, \mathbf{L}$.
5. $\mathbf{T}$ and $\mathbf{H}$ be jointly uniformly distributed independent of $\mathbf{M}$ and $\mathbf{U}$.

Ruling out FPSRS for matching—setup.

We construct a conditional $\mathfrak{Z}$ to make the problem decompose:

1. Choose $\mathbf{H}$ 3-matching between disjoint subsets $\mathbf{C_H}$ and $\mathbf{D_H}$. Goal of $\mathfrak{Z}$: only pairs $(\mathbf{M}, \mathbf{U})$ with $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$ with $\mathbf{C_H} \subseteq \mathbf{U}$ and $\mathbf{U} \cap \mathbf{D_H} = \emptyset$.
2. Partition the remaining vertices not covered by $\mathbf{H}$ into chunks $\mathbf{T_1}, \ldots, \mathbf{T_m}$ of size $2(k-3)$ (put residual into $\mathbf{L}$).
3. Split $\mathbf{T_i}$ into disjoint sets $\mathbf{C_i}$ and $\mathbf{D_i}$ of size $k-3$.
4. $\mathbf{T}$ collection of $\mathbf{C_1}, \mathbf{D_1}, \ldots, \mathbf{C_m}, \mathbf{D_m}, \mathbf{C_H}, \mathbf{D_H}, \mathbf{L}$.
5. $\mathbf{T}$ and $\mathbf{H}$ be jointly uniformly distributed independent of $\mathbf{M}$ and $\mathbf{U}$.

Ruling out FPSRS for matching—setup.

We construct a conditional $\mathfrak{Z}$ to make the problem decompose:

1. Choose $\mathbf{H}$ 3-matching between disjoint subsets $\mathbf{C_H}$ and $\mathbf{D_H}$. Goal of $\mathfrak{Z}$: only pairs $(\mathbf{M}, \mathbf{U})$ with $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$ with $\mathbf{C_H} \subseteq \mathbf{U}$ and $\mathbf{U} \cap \mathbf{D_H} = \emptyset$.
2. Partition the remaining vertices not covered by $\mathbf{H}$ into chunks $\mathbf{T_1}, \ldots, \mathbf{T_m}$ of size $2(k-3)$ (put residual into $\mathbf{L}$).
3. Split $\mathbf{T_i}$ into disjoint sets $\mathbf{C_i}$ and $\mathbf{D_i}$ of size $k-3$.
4. $\mathbf{T}$ collection of $\mathbf{C_1}, \mathbf{D_1}, \ldots, \mathbf{C_m}, \mathbf{D_m}, \mathbf{C_H}, \mathbf{D_H}, \mathbf{L}$.
5. $\mathbf{T}$ and $\mathbf{H}$ be jointly uniformly distributed independent of $\mathbf{M}$ and $\mathbf{U}$.

Ruling out FPSRS for matching—setup.

We construct a conditional $\mathfrak{Z}$ to make the problem decompose:

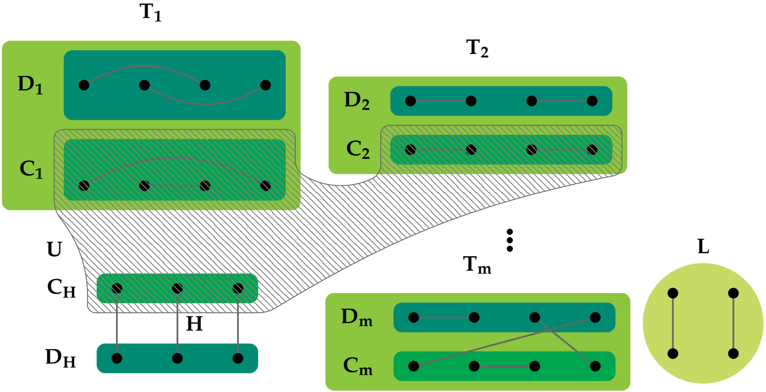
1. Choose $\mathbf{H}$ 3-matching between disjoint subsets $\mathbf{C_H}$ and $\mathbf{D_H}$.
   Goal of $\mathfrak{Z}$: only pairs $(\mathbf{M}, \mathbf{U})$ with $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$ with $\mathbf{C_H} \subseteq \mathbf{U}$ and $\mathbf{U} \cap \mathbf{D_H} = \emptyset$.
2. Partition the remaining vertices not covered by $\mathbf{H}$ into chunks $\mathbf{T_1}, \ldots, \mathbf{T_m}$ of size $2(k-3)$ (put residual into $\mathbf{L}$).
3. Split $\mathbf{T_i}$ into disjoint sets $\mathbf{C_i}$ and $\mathbf{D_i}$ of size $k-3$.
4. $\mathbf{T}$ collection of $\mathbf{C_1}, \mathbf{D_1}, \ldots, \mathbf{C_m}, \mathbf{D_m}, \mathbf{C_H}, \mathbf{D_H}, \mathbf{L}$.
5. $\mathbf{T}$ and $\mathbf{H}$ be jointly uniformly distributed independent of $\mathbf{M}$ and $\mathbf{U}$.

Ruling out FPSRS for matching—setup.

We construct a conditional $\mathfrak{Z}$ to make the problem decompose:

1. Choose $\mathbf{H}$ 3-matching between disjoint subsets $\mathbf{C_H}$ and $\mathbf{D_H}$.
   Goal of $\mathfrak{Z}$: only pairs $(\mathbf{M}, \mathbf{U})$ with $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$ with $\mathbf{C_H} \subseteq \mathbf{U}$ and $\mathbf{U} \cap \mathbf{D_H} = \emptyset$.
2. Partition the remaining vertices not covered by $\mathbf{H}$ into chunks $\mathbf{T_1}, \ldots, \mathbf{T_m}$ of size $2(k-3)$ (put residual into $\mathbf{L}$).
3. Split $\mathbf{T_i}$ into disjoint sets $\mathbf{C_i}$ and $\mathbf{D_i}$ of size $k-3$.
4. $\mathbf{T}$ collection of $\mathbf{C_1}, \mathbf{D_1}, \ldots, \mathbf{C_m}, \mathbf{D_m}, \mathbf{C_H}, \mathbf{D_H}, \mathbf{L}$.
5. $\mathbf{T}$ and $\mathbf{H}$ be jointly uniformly distributed independent of $\mathbf{M}$ and $\mathbf{U}$.

Ruling out FPSRS for matching—setup.



Ruling out FPSRS for matching—setup.

Baseline event $\mathcal{E}_0$ (keep the setup clean):

$$\mathcal{E}_0 := \begin{cases} \mathbf{U} \cap \mathbf{T_i} \in \{\emptyset, \mathbf{C_i}\}, & \mathbf{C_H} \subseteq \mathbf{U} \subseteq \mathbf{C_H} \cup \bigcup_{i \in [m]} \mathbf{C_i} \\ \mathbf{H} \subseteq \mathbf{M}, & \mathbf{M} \subseteq \mathbf{H} \cup E[\mathbf{L}] \cup \bigcup_{i \in [m]} E[\mathbf{T_i}] \end{cases}$$

In particular, given $\mathcal{E}_0$ we have $\mathbf{L} \cap \mathbf{U} = \emptyset$. (Actually, the sole role of $\mathbf{L}$ is to collect the vertices not fitting into the scheme.)

Mutually independent random fair coins $\mathbf{N} = \mathbf{N_1}, \ldots, \mathbf{N_m}$, which are also independent of the random variables introduced before.

$\mathcal{E}$ to switch cases via coins (recall UDISJ):

$$\mathcal{E} := \mathcal{E}_0 \wedge \begin{cases} \mathbf{U} \cap \mathbf{T_i} = \emptyset, & \text{if } \mathbf{N_i} = 0, \\ \delta(\mathbf{C_i}) \cap \mathbf{M} = \emptyset, & \text{if } \mathbf{N_i} = 1 \end{cases}$$

The event $\mathcal{E}$ ensures $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$.

Ruling out FPSRS for matching—setup.

Baseline event $\mathcal{E}_0$ (keep the setup clean):

$$\mathcal{E}_0 := \begin{cases} \mathbf{U} \cap \mathbf{T_i} \in \{\emptyset, \mathbf{C_i}\}, & \mathbf{C_H} \subseteq \mathbf{U} \subseteq \mathbf{C_H} \cup \bigcup_{i \in [m]} \mathbf{C_i} \\ \mathbf{H} \subseteq \mathbf{M}, & \mathbf{M} \subseteq \mathbf{H} \cup E[\mathbf{L}] \cup \bigcup_{i \in [m]} E[\mathbf{T_i}] \end{cases}$$

In particular, given $\mathcal{E}_0$ we have $\mathbf{L} \cap \mathbf{U} = \emptyset$. (Actually, the sole role of $\mathbf{L}$ is to collect the vertices not fitting into the scheme.)

Mutually independent random fair coins $\mathbf{N} = \mathbf{N_1}, \ldots, \mathbf{N_m}$, which are also independent of the random variables introduced before.

$\mathcal{E}$ to switch cases via coins (recall UDISJ):

$$\mathcal{E} := \mathcal{E}_0 \wedge \begin{cases} \mathbf{U} \cap \mathbf{T_i} = \emptyset, & \text{if } \mathbf{N_i} = 0, \\ \delta(\mathbf{C_i}) \cap \mathbf{M} = \emptyset, & \text{if } \mathbf{N_i} = 1 \end{cases}$$

The event $\mathcal{E}$ ensures $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$.

Ruling out FPSRS for matching—setup.

Baseline event $\mathcal{E}_0$ (keep the setup clean):

$$\mathcal{E}_0 := \begin{cases} \mathbf{U} \cap \mathbf{T_i} \in \{\emptyset, \mathbf{C_i}\}, \quad \mathbf{C_H} \subseteq \mathbf{U} \subseteq \mathbf{C_H} \cup \bigcup_{i \in [m]} \mathbf{C_i} \\ \mathbf{H} \subseteq \mathbf{M}, \qquad \mathbf{M} \subseteq \mathbf{H} \cup E[\mathbf{L}] \cup \bigcup_{i \in [m]} E[\mathbf{T_i}] \end{cases}$$

In particular, given $\mathcal{E}_0$ we have $\mathbf{L} \cap \mathbf{U} = \emptyset$. (Actually, the sole role of $\mathbf{L}$ is to collect the vertices not fitting into the scheme.)

Mutually independent random fair coins $\mathbf{N} = \mathbf{N_1}, \ldots, \mathbf{N_m}$, which are also independent of the random variables introduced before.

$\mathcal{E}$ to switch cases via coins (recall UDISJ):

$$\mathcal{E} := \mathcal{E}_0 \wedge \begin{cases} \mathbf{U} \cap \mathbf{T_i} = \emptyset, & \text{if } \mathbf{N_i} = 0, \\ \delta(\mathbf{C_i}) \cap \mathbf{M} = \emptyset, & \text{if } \mathbf{N_i} = 1 \end{cases}$$

The event $\mathcal{E}$ ensures $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$.

Ruling out FPSRS for matching—reduction to $m = 1$ and $L = \emptyset$.

We will show

$$\log \operatorname{rk}_+ S^{+\varepsilon} \geq \mathbb{C}[S^{+\varepsilon}] \geq \min_{\mathbf{\Pi}\text{: seed}} \mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \mid \mathbf{T}, \mathbf{N}, \mathbf{H}, \mathcal{E}\right] \geq c_{k,\varepsilon} m = \Theta(n).$$

Reduction to $m = 1$ and $L = 0$:

1. Recall $\mathcal{E}$ ensures $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$.
2. Thus, as the probability of a pair $(\mathbf{M}, \mathbf{U})$ depends only on the number of crossing edges, $(\mathbf{M}, \mathbf{U})$ is uniformly distributed given $\mathcal{E}$.
3. The matching $\mathbf{M}$ decomposes into $\mathbf{M_i} := \mathbf{M} \cap E[\mathbf{T_i}]$ for $i \in [m]$, together with $\mathbf{M_L} := \mathbf{M} \cap E[\mathbf{L}]$ and $\mathbf{H}$. Similarly, the set $\mathbf{U}$ decomposes as $\mathbf{U} = \mathbf{C_H} \cup \bigcup_{i \in [m]} \mathbf{U_i}$ with $\mathbf{U_i} := \mathbf{U} \cap \mathbf{T_i}$.

The pairs $(\mathbf{M_i}, \mathbf{U_i})$ together with $(\mathbf{M_L}, \emptyset)$ are mutually independent, therefore by the direct sum property

$$\mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \mid \mathbf{T}, \mathbf{N}, \mathbf{H}, \mathcal{E}\right] \geq \sum_{i \in [m]} \mathbb{I}\left[\mathbf{M_i}, \mathbf{U_i}; \mathbf{\Pi} \mid \mathbf{T}, \mathbf{N}, \mathbf{H}, \mathcal{E}\right] \geq c_{k,\varepsilon} m,$$

where the last inequality is concluded from the local case.

Ruling out FPSRS for matching—reduction to $m = 1$ and $L = \emptyset$.

We will show

$$\log \operatorname{rk}_+ S^{+\varepsilon} \geq \mathbb{C}[S^{+\varepsilon}] \geq \min_{\mathbf{\Pi}\text{: seed}} \mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{T}, \mathbf{N}, \mathbf{H}, \mathcal{E}\right] \geq c_{k,\varepsilon} m = \Theta(n).$$

Reduction to $m = 1$ and $L = 0$:

1. Recall $\mathcal{E}$ ensures $\delta(\mathbf{U}) \cap \mathbf{M} = \mathbf{H}$.
2. Thus, as the probability of a pair $(\mathbf{M}, \mathbf{U})$ depends only on the number of crossing edges, $(\mathbf{M}, \mathbf{U})$ is uniformly distributed given $\mathcal{E}$.
3. The matching $\mathbf{M}$ decomposes into $\mathbf{M_i} \coloneqq \mathbf{M} \cap E[\mathbf{T_i}]$ for $i \in [m]$, together with $\mathbf{M_L} \coloneqq \mathbf{M} \cap E[\mathbf{L}]$ and $\mathbf{H}$. Similarly, the set $\mathbf{U}$ decomposes as $\mathbf{U} = \mathbf{C_H} \cup \bigcup_{i \in [m]} \mathbf{U_i}$ with $\mathbf{U_i} \coloneqq \mathbf{U} \cap \mathbf{T_i}$.

The pairs $(\mathbf{M_i}, \mathbf{U_i})$ together with $(\mathbf{M_L}, \emptyset)$ are mutually independent, therefore by the direct sum property

$$\mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{T}, \mathbf{N}, \mathbf{H}, \mathcal{E}\right] \geq \sum_{i \in [m]} \mathbb{I}\left[\mathbf{M_i}, \mathbf{U_i}; \mathbf{\Pi} \,|\, \mathbf{T}, \mathbf{N}, \mathbf{H}, \mathcal{E}\right] \geq c_{k,\varepsilon} m,$$

where the last inequality is concluded from the local case.

Ruling out FPSRS for matching—the local case.

Cleaning up the setup:

1. $\mathbf{C} \coloneqq \mathbf{C_1} \cup \mathbf{C_H}$ and $\mathbf{D} \coloneqq \mathbf{D_1} \cup \mathbf{D_H}$
2. $\mathbf{C}, \mathbf{D}$ and $\mathbf{H}$ are uniformly distributed (independently of $\mathbf{M}, \mathbf{U}, \mathbf{\Pi}, \mathbf{N}$), and together determine the $\mathbf{C_1}, \mathbf{D_1}, \mathbf{C_H}, \mathbf{D_H}$.
3. Introduce $\mathbf{F}$ as a uniformly random extension of $\mathbf{H}$ into a full matching between $\mathbf{C}$ and $\mathbf{D}$, depending only on $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{H}$.

This independence ensures that adding it as condition to the mutual information has no effect:

$$
\begin{aligned}
&\mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{T}, \mathbf{H}, \mathbf{N}, \mathcal{E}\right] = \mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{T}, \mathbf{H}, \mathbf{F}, \mathbf{N}, \mathcal{E}\right] \\
= \;& \mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{C}, \mathbf{D}, \mathbf{F}, \mathbf{H}, \mathbf{N}, \mathcal{E}\right] \\
= \;& \mathbb{E}_{C \sim \mathbf{C}, D \sim \mathbf{D}, F \sim \mathbf{F} | \mathcal{E}}\left[\mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{C} = C, \mathbf{D} = D, \mathbf{F} = F, \mathbf{H}, \mathbf{N}, \mathcal{E}\right]\right]
\end{aligned}
$$

Ruling out FPSRS for matching—the local case.

From now on fix C, D, F and drop from conditional (we average over all specific choices).

Cleaning up the setup (now the events):

$$\mathcal{E}_0 \coloneqq \{\mathbf{U} \in \{C, C(\mathbf{H})\}, \mathbf{H} \subseteq \mathbf{M}\} \quad \mathcal{E} \coloneqq \begin{cases} \mathbf{U} = C(\mathbf{H}), & \text{if } \mathbf{N} = 0 \\ \delta(C) \cap \mathbf{M} = \mathbf{H}, & \text{if } \mathbf{N} = 1. \end{cases}$$

Here and below for a 3-matching $h \subseteq F$, let $C(h)$ denote the endpoints of the edges of h lying in C. With this:

$$\mathbb{I}[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \mid \mathbf{H}, \mathbf{N}, \mathcal{E}] = \mathbb{E}_{\mathbf{\Pi}, \mathbf{H}, \mathbf{N} \mid \mathcal{E}} [\mathrm{D}(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N}, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N}, \mathcal{E})]$$
$$= \sum_{i \in \{0,1\}} \mathbb{P}[\mathbf{N} = i \mid \mathcal{E}] \cdot \mathbb{E}_{\mathbf{\Pi}, \mathbf{H} \mid \mathbf{N} = i, \mathcal{E}} [\mathrm{D}(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N} = i, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N} = i, \mathcal{E})].$$

Ruling out FPSRS for matching—the local case.

From now on fix C, D, F and drop from conditional (we average over all specific choices).

Cleaning up the setup (now the events):

$$\mathcal{E}_0 \coloneqq \{\mathbf{U} \in \{C, C(\mathbf{H})\}, \mathbf{H} \subseteq \mathbf{M}\} \quad \mathcal{E} \coloneqq \begin{cases} \mathbf{U} = C(\mathbf{H}), & \text{if } \mathbf{N} = 0 \\ \delta(C) \cap \mathbf{M} = \mathbf{H}, & \text{if } \mathbf{N} = 1. \end{cases}$$

Here and below for a 3-matching $h \subseteq F$, let $C(h)$ denote the endpoints of the edges of h lying in C. With this:

$$\mathbb{I}\left[\mathbf{M}, \mathbf{U}; \mathbf{\Pi} \,|\, \mathbf{H}, \mathbf{N}, \mathcal{E}\right] = \mathbb{E}_{\mathbf{\Pi}, \mathbf{H}, \mathbf{N} | \mathcal{E}}\left[\mathrm{D}\left(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N}, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N}, \mathcal{E}\right)\right]$$
$$= \sum_{i \in \{0,1\}} \mathbb{P}\left[\mathbf{N} = i \,|\, \mathcal{E}\right] \cdot \mathbb{E}_{\mathbf{\Pi}, \mathbf{H} | \mathbf{N} = i, \mathcal{E}}\left[\mathrm{D}\left(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N} = i, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N} = i, \mathcal{E}\right)\right].$$

Ruling out FPSRS for matching—the local case.

We analyze the relative entropy term

$$I \coloneqq \mathrm{D}\left(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N} = i, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N} = i, \mathcal{E}\right).$$

When is $I \approx 0$?

Whenever the distribution of matchings and odd sets on the whole slack matrix is close the one of the rank-1 factor under consideration.

These factors do not contribute to the lower bound and we care for those where the distribution is markedly different.

A pair (π, h) is $\mathbf{M}$-*good* if for all matchings $m \supseteq h$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]} \leq 1 + \delta.$$

A pair (π, h) is $\mathbf{U}$-*good* if for $u = C(h)$ and $u = C$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]} \leq 1 + \delta.$$

Ruling out FPSRS for matching—the local case.

We analyze the relative entropy term

$$I \coloneqq \mathrm{D}\left(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N} = i, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N} = i, \mathcal{E}\right).$$

When is $I \approx 0$?

Whenever the distribution of matchings and odd sets on the whole slack matrix is close the one of the rank-1 factor under consideration.

These factors do not contribute to the lower bound and we care for those where the distribution is markedly different.

A pair (π, h) is $\mathbf{M}$-*good* if for all matchings $m \supseteq h$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]} \leq 1 + \delta.$$

A pair (π, h) is $\mathbf{U}$-*good* if for $u = C(h)$ and $u = C$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]} \leq 1 + \delta.$$

Ruling out FPSRS for matching—the local case.

We analyze the relative entropy term

$$I := \mathrm{D}\left(\mathbf{M}, \mathbf{U} \mid \mathbf{\Pi}, \mathbf{H}, \mathbf{N} = i, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N} = i, \mathcal{E}\right).$$

When is $I \approx 0$?

Whenever the distribution of matchings and odd sets on the whole slack matrix is close the one of the rank-1 factor under consideration.

These factors do not contribute to the lower bound and we care for those where the distribution is markedly different.

A pair (π, h) is $\mathbf{M}$-*good* if for all matchings $m \supseteq h$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]} \leq 1 + \delta.$$

A pair (π, h) is $\mathbf{U}$-*good* if for $u = C(h)$ and $u = C$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]} \leq 1 + \delta.$$

Ruling out FPSRS for matching—the local case.

We analyze the relative entropy term

$$I := \mathrm{D}\left(\mathbf{M}, \mathbf{U} \mid \boldsymbol{\Pi}, \mathbf{H}, \mathbf{N} = i, \mathcal{E} \,\|\, \mathbf{M}, \mathbf{U} \mid \mathbf{H}, \mathbf{N} = i, \mathcal{E}\right).$$

When is $I \approx 0$?

Whenever the distribution of matchings and odd sets on the whole slack matrix is close the one of the rank-1 factor under consideration.

These factors do not contribute to the lower bound and we care for those where the distribution is markedly different.

A pair (π, h) is $\mathbf{M}$-*good* if for all matchings $m \supseteq h$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{M} = m \,|\, \boldsymbol{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{M} = m \,|\, \mathbf{H} = h, \mathbf{N} = 0, \mathcal{E}\right]} \leq 1 + \delta.$$

A pair (π, h) is $\mathbf{U}$-*good* if for $u = C(h)$ and $u = C$

$$1 - \delta \leq \frac{\mathbb{P}\left[\mathbf{U} = u \,|\, \boldsymbol{\Pi} = \pi, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]}{\mathbb{P}\left[\mathbf{U} = u \,|\, \mathbf{H} = h, \mathbf{N} = 1, \mathcal{E}\right]} \leq 1 + \delta.$$

Ruling out FPSRS for matching—the local case.

Via Pinsker's inequality:

$$\mathbb{E}_{\mathbf{\Pi},\mathbf{H}|\mathbf{N}=0,\mathcal{E}}\left[\mathrm{D}\left(\mathbf{M},\mathbf{U} \mid \mathbf{\Pi},\mathbf{H},\mathbf{N}=0,\mathcal{E} \,\|\, \mathbf{M},\mathbf{U} \mid \mathbf{H},\mathbf{N}=0,\mathcal{E}\right)\right] \geq \mathbb{P}\left[\mathbf{M}\text{-BAD}(\mathbf{\Pi},\mathbf{H}) \mid \mathbf{N}=0,\mathcal{E}\right] 2(\log e)(\delta\alpha)^2$$

as given $\mathbf{\Pi}$, the variables $\mathbf{N}$ and $\mathbf{U}$ are independent of $\mathbf{M}$. Similarly,

$$\mathbb{E}_{\mathbf{\Pi},\mathbf{H}|\mathbf{N}=1,\mathcal{E}}\left[\mathrm{D}\left(\mathbf{M},\mathbf{U} \mid \mathbf{\Pi},\mathbf{H},\mathbf{N}=1,\mathcal{E} \,\|\, \mathbf{M},\mathbf{U} \mid \mathbf{H},\mathbf{N}=1,\mathcal{E}\right)\right] \geq \mathbb{P}\left[\mathbf{U}\text{-BAD}(\mathbf{\Pi},\mathbf{H}) \mid \mathbf{N}=1,\mathcal{E}\right] 2(\log e)\left(\frac{\delta}{2}\right)^2$$

. . . some technical computations . . .

$$\min\left\{\mathbb{P}\left[\mathbf{M}\text{-BAD}(\mathbf{\Pi},\mathbf{H}) \mid \mathbf{N}=0,\mathcal{E}\right], \mathbb{P}\left[\mathbf{U}\text{-BAD}(\mathbf{\Pi},\mathbf{H}) \mid \mathbf{N}=1,\mathcal{E}\right]\right\} \geq B_{k,\varepsilon} > 0.$$

Ruling out FPSRS for matching—the local case.

Via Pinsker's inequality:

$$\mathbb{E}_{\mathbf{\Pi},\mathbf{H}|\mathbf{N}=0,\mathcal{E}}\left[\mathrm{D}\left(\mathbf{M},\mathbf{U}\mid\mathbf{\Pi},\mathbf{H},\mathbf{N}=0,\mathcal{E}\,\|\,\mathbf{M},\mathbf{U}\mid\mathbf{H},\mathbf{N}=0,\mathcal{E}\right)\right] \geq \mathbb{P}\left[\mathbf{M}\text{-BAD}(\mathbf{\Pi},\mathbf{H})\,|\,\mathbf{N}=0,\mathcal{E}\right]2(\log e)(\delta\alpha)^2$$

as given $\mathbf{\Pi}$, the variables $\mathbf{N}$ and $\mathbf{U}$ are independent of $\mathbf{M}$. Similarly,

$$\mathbb{E}_{\mathbf{\Pi},\mathbf{H}|\mathbf{N}=1,\mathcal{E}}\left[\mathrm{D}\left(\mathbf{M},\mathbf{U}\mid\mathbf{\Pi},\mathbf{H},\mathbf{N}=1,\mathcal{E}\,\|\,\mathbf{M},\mathbf{U}\mid\mathbf{H},\mathbf{N}=1,\mathcal{E}\right)\right] \geq \mathbb{P}\left[\mathbf{U}\text{-BAD}(\mathbf{\Pi},\mathbf{H})\,|\,\mathbf{N}=1,\mathcal{E}\right]2(\log e)\left(\frac{\delta}{2}\right)^2$$

...some technical computations ...

$$\min\left\{\mathbb{P}\left[\mathbf{M}\text{-BAD}(\mathbf{\Pi},\mathbf{H})\,|\,\mathbf{N}=0,\mathcal{E}\right],\mathbb{P}\left[\mathbf{U}\text{-BAD}(\mathbf{\Pi},\mathbf{H})\,|\,\mathbf{N}=1,\mathcal{E}\right]\right\} \geq B_{k,\varepsilon} > 0.$$

Ruling out FPSRS for matching.

Theorem

Let $0 < \varepsilon < 1$ be fixed and n even. Then $\mathrm{xc}(P_{PM}(n), Q^{+\varepsilon}(n)) = 2^{\Theta(n)}$. In particular, the extension complexity of the ρ-approximation of the perfect matching polytope is $\mathrm{xc}(P_{PM}(n), \rho Q) = 2^{\Theta(n)}$ for $\rho \leq 1 + \varepsilon/n$, and $\mathrm{xc}(P_{PM}(n)) = 2^{\Theta(n)}$. Thus, the perfect matching polytope does not admit an FPSRS.

Thank you!