Minimum Rényi Entropy Couplings and their Applications

Anuj K. Yadav
EPFL

(Joint work with Yanina Shkel)

August 17, 2026

Outline

Outline

Minimum Rényi Entropy Coupling (M-REC)

Given : m marginal PMFs $\{P_i\}_{i=1}^m$

Find : coupling $C^\star \in \underset{C \in \Gamma(P_1,\ldots,P_m)}{\arg\min} H_\alpha(C), \qquad \alpha \geq 0$

where $\Gamma(P_1, \ldots, P_m)$ is the set of all couplings with marginals $\{P_i\}_{i=1}^m$.

However . . .

- Let $C^\star$ denote the optimal coupling.
- Computing $C^\star$ (or $H_\alpha(C^\star)$) is **NP-hard** in the support size of PMFs.

M. Kovačević et al., "On the entropy of couplings", Information and Computation, 2015.

However ...

- Let $C^\star$ denote the optimal coupling.
- Computing $C^\star$ (or $H_\alpha(C^\star)$) is **NP-hard** in the support size of PMFs.

- **Lower bounds on** $H_\alpha(C^\star)$ — Converse type results
- **Upper bounds on** $H_\alpha(C^\star)$ — Achievability type results

M. Kovačević et al., "On the entropy of couplings", Information and Computation, 2015.

Outline

Lower bounds (Converse Results)

What are the worst-case lower bounds on $H_\alpha(C^\star)$?

Prelude

Let P be a PMF such that $P = (p_1, p_2, p_3 \ldots,)$.

Information of P :

$$\imath_P = \log\left(\frac{1}{p_i}\right);$$

with probability p_i.

Information spectrum of P :

$$\mathbb{F}_{\imath_P}(t) = \mathbb{P}[\imath_P \leq t] ;$$

$$\forall t \in [0, \infty).$$

Prelude

Let P be a PMF such that $P = (p_1, p_2, p_3 \ldots,)$.

Information of P :

$$\imath_P = \log\left(\frac{1}{p_i}\right) ;$$

with probability p_i.

Information spectrum of P :

$$\mathbb{F}_{\imath_P}(t) = \mathbb{P}[\imath_P \leq t] ;$$

$$\forall t \in [0, \infty).$$

Shannon entropy of P :

$$H(P) = \mathbb{E}[\imath_P]$$

$$= \int_0^\infty \left(1 - \mathbb{F}_{\imath_P}(t)\right) dt.$$

Rényi entropy of P :

$$H_\alpha(P) = \frac{1}{1-\alpha} \log\left(\mathbb{E}\left[2^{(1-\alpha)\imath_P}\right]\right) ;$$

$$\forall \alpha \in [0, \infty).$$

Majorization ($\preceq_m$)

Definition

Given PMFs

$$Q = (q_1, q_2, q_3, \dots,) \ ; \quad q_1 \geq q_2 \geq \dots$$
$$P = (p_1, p_2, p_3, \dots,) \ ; \quad p_1 \geq p_2 \geq \dots$$

we say $\quad Q \preceq_m P$

if $$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k \geq 1.$$

Majorization ($\preceq_m$)

Definition

Given PMFs

$Q = (q_1, q_2, q_3, \dots,)$; $q_1 \geq q_2 \geq \dots$

$P = (p_1, p_2, p_3, \dots,)$; $p_1 \geq p_2 \geq \dots$

we say $Q \preceq_m P$

if $$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k \geq 1.$$

• $\preceq_m$ is a partial order on the set of all ordered PMFs. Moreover, it forms a complete lattice.

Majorization ($\preceq_m$)

Definition

Given PMFs

$$Q = (q_1, q_2, q_3, \dots,) \ ; \quad q_1 \geq q_2 \geq \dots$$
$$P = (p_1, p_2, p_3, \dots,) \ ; \quad p_1 \geq p_2 \geq \dots$$

we say $\quad Q \preceq_m P$

if $$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k \geq 1.$$

- $\preceq_m$ is a partial order on the set of all ordered PMFs. Moreover, it forms a complete lattice.

Greatest Lower Bound of $\{P_1, P_2, \dots, P_m\}$

Let $\mathcal{F} = \{Q \colon Q \preceq_m P_i \quad \forall i \in [m]\}$

$$\exists \bigwedge_{i=1}^{m} P_i \in \mathcal{F} \ \text{ s.t.}$$

$$Q \preceq_m \bigwedge_{i=1}^{m} P_i \preceq_m P_i \ ; \forall i, \ Q.$$

Majorization ($\preceq_m$)

Definition

Given PMFs

$Q = (q_1, q_2, q_3, \ldots,)$; $\quad q_1 \geq q_2 \geq \ldots$

$P = (p_1, p_2, p_3, \ldots,)$; $\quad p_1 \geq p_2 \geq \ldots$

we say $\quad Q \preceq_m P$

if $\qquad \sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k \geq 1.$

- $\preceq_m$ is a partial order on the set of all ordered PMFs. Moreover, it forms a complete lattice.

Greatest Lower Bound of $\{P_1, P_2, \ldots, P_m\}$

Let $\mathcal{F} = \{Q \colon Q \preceq_m P_i \quad \forall i \in [m]\}$

$$\exists \bigwedge_{i=1}^{m} P_i \in \mathcal{F} \text{ s.t.}$$

$$Q \preceq_m \bigwedge_{i=1}^{m} P_i \preceq_m P_i \ ; \forall i,\ Q.$$

Schur Concavity

$$Q \preceq_m P \implies f(Q) \geq f(P)$$

Majorization ($\preceq_m$)

Definition

Given PMFs

$$Q = (q_1, q_2, q_3, \ldots,) \;; \quad q_1 \geq q_2 \geq \ldots$$
$$P = (p_1, p_2, p_3, \ldots,) \;; \quad p_1 \geq p_2 \geq \ldots$$

we say $\quad Q \preceq_m P$

if $$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k \geq 1.$$

• $\preceq_m$ is a partial order on the set of all ordered PMFs. Moreover, it forms a complete lattice.

Greatest Lower Bound of $\{P_1, P_2, \ldots, P_m\}$

Let $\mathcal{F} = \{Q \colon Q \preceq_m P_i \quad \forall i \in [m]\}$

$$\exists \bigwedge_{i=1}^{m} P_i \in \mathcal{F} \text{ s.t.}$$

$$Q \preceq_m \bigwedge_{i=1}^{m} P_i \preceq_m P_i \;; \forall i,\; Q.$$

Schur Concavity

$$Q \preceq_m P \implies f(Q) \geq f(P)$$

Rényi entropy is Schur concave.

Lower bound : A very basic one

- $C(P_1, \cdots, P_m) \sqsubseteq P_i \;\; ; \; \forall C, \; \forall i \in [m]$
- Aggregation implies Majorization.
- $C(P_1, \cdots, P_m) \preceq_m P_i \;\; ; \; \forall C, \; \forall i \in [m]$

$$\boxed{H_\alpha(C^\star) \geq \max_{i \in [m]} H_\alpha(P_i)}$$

Y. Y. Shkel, R. Blum, H. V. Poor, "Secrecy by design with applications to privacy and compression," IEEE Transactions on Information Theory, vol. 67, no. 2, pp. 824-843, 2021.

Lower bound : A very basic one

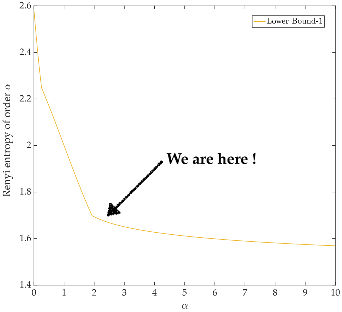
$$\boxed{H_\alpha(C^\star) \geq \max_{i \in [m]} H_\alpha(P_i)}$$

Toy Example with $(m = 3,\ n = 6)$:

$P_1 = (0.5, 0.125, 0.125, 0.125, 0.125, 0)$;

$P_2 = (0.4,\ 0.4,\ 0.1,\ 0.1,\ 0,\ 0)$;

$P_3 = (0.35, 0.35, 0.25, 0.04, 0.005, 0.005)$;



Lower bound : Based on Majorization ($\preceq_m$)

Recall,

$$C(P_1, \cdots, P_m) \preceq_m P_i \;\; ; \forall C, \, \forall i \in [m]$$

Thus,

$$C(P_1, \cdots, P_m) \;\preceq_m\; \bigwedge_{i=1}^{m} P_i \;\preceq_m\; P_i \;\; ; \quad \forall i \in [m]$$

$$\boxed{H_\alpha(C^\star) \geq H_\alpha\Big(\bigwedge_{i=1}^{m} P_i\Big)}$$

F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3436–3451, 2019.

Lower bound : Based on Majorization ($\preceq_m$)

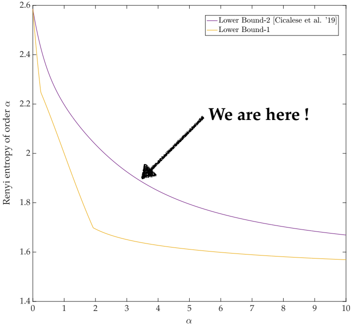
Recall,

$$C(P_1, \cdots, P_m) \preceq_m P_i \;\; ; \forall C, \, \forall i \in [m]$$

Thus,

$$C(P_1, \cdots, P_m) \;\preceq_m\; \bigwedge_{i=1}^{m} P_i \;\preceq_m\; P_i \;\; ; \;\; \forall i \in [m]$$

$$\boxed{H_\alpha(C^\star) \geq H_\alpha\Big(\bigwedge_{i=1}^{m} P_i\Big)}$$



F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3436–3451, 2019.

Majorization : Information-spectrum sense ($\preceq_\imath$)

We say : P majorizes Q in an information-spectrum sense, i.e.,

$$Q \preceq_\imath P$$

if $$\boxed{\mathbb{F}_{\imath_Q}(t) \leq \mathbb{F}_{\imath_P}(t), \quad \forall t \in [0, \infty)}$$

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations," in IEEE International Symposium on Information Theory (ISIT), 2023.

Information-spectrum based Lower Bound : Main Result 1

Theorem: Let $\mathcal{S} := \{P_1, \ldots, P_m\}$ be the set of m marginal distributions with support size at most n. Then,

$$C \preceq_{\imath} P_i \quad ; \quad \forall C, \; \forall i \in [m]$$

or

$$\mathbb{F}_{\imath_C}(t) \leq \mathbb{F}_{\imath_{P_i}}(t) \; ; \quad \forall t \in [0, \infty) \, , \forall C, \; \forall i \in [m]$$

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations," in IEEE International Symposium on Information Theory (ISIT), 2023.

Information-spectrum based Lower Bound : Main Result 1

$$\mathbb{F}_{\imath_{C^\star}}(t) \le \mathbb{F}_{\imath_{P_i}}(t) \; ; \quad \forall i \in [m]$$

$$\Downarrow$$

$$\begin{aligned} H(C^\star) = \mathbb{E}[\imath_{C^\star}(C^\star)] &= \int_0^\infty \Big(1 - \mathbb{F}_{\imath_{C^\star}}(t)\Big) dt \\ &\ge \int_0^\infty \max_{i\in[m]} \Big(1 - \mathbb{F}_{\imath_{P_i}}(t)\Big) dt \end{aligned}$$

$$\boxed{H(C^\star) \ge K_1(\mathcal{S})}$$

where $K_1(\mathcal{S}) := \int_0^\infty \max_{i\in[m]} \Big(1 - \mathbb{F}_{\imath_{P_i}}(t)\Big) dt$

Similar extension for Rényi Entropy
i.e., $H_\alpha(C^\star) \ge K_\alpha(\mathcal{S})$.

Information-spectrum based Lower Bound : Main Result 1

$$\mathbb{F}_{\imath_{C^\star}}(t) \leq \mathbb{F}_{\imath_{P_i}}(t) \; ; \quad \forall i \in [m]$$

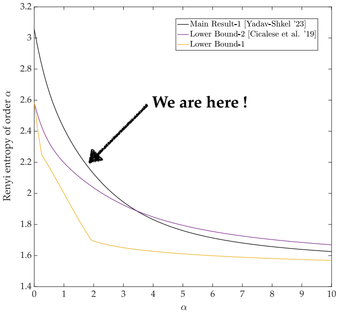
$$\Downarrow$$

$$H(C^\star) = \mathbb{E}[\imath_{C^\star}(C^\star)] = \int_0^\infty \Big(1 - \mathbb{F}_{\imath_{C^\star}}(t)\Big) dt$$

$$\geq \int_0^\infty \max_{i\in[m]} \Big(1 - \mathbb{F}_{\imath_{P_i}}(t)\Big) dt$$

$$\boxed{H(C^\star) \geq K_1(\mathcal{S})}$$

where $K_1(\mathcal{S}) := \int_0^\infty \max_{i\in[m]} \Big(1 - \mathbb{F}_{\imath_{P_i}}(t)\Big) dt$

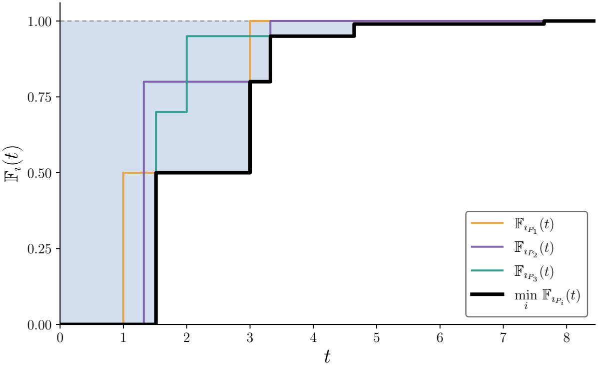


Similar extension for Rényi Entropy
i.e., $H_\alpha(C^\star) \geq K_\alpha(\mathcal{S})$.

Information-spectrum based Lower bound: Main Result 1

Example: Shannon entropy

- $H(C^\star) \geq K_1(\mathcal{S}) =$ area above the lower envelope.



Information-spectrum based Lower Bound : Main Result 1

Theorem: Let $\mathcal{S} := \{P_1, \dots, P_m\}$ be the set of m marginal distributions. Then, for any $\alpha \in [0, \infty)$, we have

$$H_\alpha(C^\star) \geq K_\alpha(\mathcal{S})$$

where,

$$K_\alpha(\mathcal{S}) = \begin{cases} \dfrac{1}{1-\alpha} \log \left[1 + \displaystyle\int_0^\infty J_\alpha(t)\, dt\right] & ; \text{ if } \alpha \in [0,1) \cup (1, \infty) \\ \displaystyle\int_0^\infty G(t)\, dt & ; \ \alpha = 1 \end{cases}$$

such that :

$$G(t) := \max_{i \in [m]} \left(1 - \mathbb{F}_{\imath_{P_i}}(t)\right)$$

$$J_\alpha(t) := (\ln 2)(1-\alpha)\, G(t)\, 2^{(1-\alpha)t}$$

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations," in IEEE International Symposium on Information Theory (ISIT), 2023.

Majorization : Information-spectrum sense ($\preceq_\imath$)

We say :

$$Q \preceq_\imath P$$
$$\text{if} \quad \mathbb{F}_{\imath_Q}(t) \leq \mathbb{F}_{\imath_P}(t), \quad \forall t \in [0, \infty)$$

Recall that :

$$Q \preceq_m P$$
$$\text{if} \quad \sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i \quad \forall k \geq 1$$

- **Lemma 1:** $Q \preceq_\imath P \implies Q \preceq_m P$

Majorization : Information-spectrum sense ($\preceq_\imath$)

We say :

$$Q \preceq_\imath P$$
if $\mathbb{F}_{\imath_Q}(t) \leq \mathbb{F}_{\imath_P}(t), \quad \forall t \in [0, \infty)$

Recall that :

$$Q \preceq_m P$$
if $\displaystyle\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i \quad \forall k \geq 1$

- **Lemma 1:** $Q \preceq_\imath P \implies Q \preceq_m P$
- Let $\mathcal{F} = \{Q \colon Q \preceq_\imath P_i \; ; \; \forall i \in [m]\}$

$$\nexists \bigwedge_{i=1}^{m} P_i \in \mathcal{F} \quad \text{s.t.} \quad Q \preceq_\imath \bigwedge_{i=1}^{m} P_i \preceq_\imath P_i \; ; \; \forall i \in [m] \, , \; Q \in \mathcal{F}$$

- $\preceq_\imath$ does not form a lattice; the greatest lower bound does not exist.

Majorization : Information-spectrum sense ($\preceq_\imath$)

We say :

$$Q \preceq_\imath P$$
if $\mathbb{F}_{\imath_Q}(t) \leq \mathbb{F}_{\imath_P}(t), \quad \forall t \in [0, \infty)$

Recall that :

$$Q \preceq_m P$$
if $\displaystyle\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i \quad \forall k \in [m]$

- **Lemma 1:** $Q \preceq_\imath P \implies Q \preceq_m P$
- **Lemma 2:** Let $\mathcal{F} = \{Q\colon Q \preceq_\imath P_i \quad \forall i \in [m]\}$

$$\exists Q^* \in \mathcal{F} \text{ s.t. } Q \preceq_m Q^* \preceq_i P_i\ ; \quad \forall Q \in \mathcal{F},\, i \in [m].$$

- Q^* is computable in polynomial time $\mathcal{O}(mn)$.

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations," in IEEE International Symposium on Information Theory (ISIT), 2023.

Information-spectrum based Lower bound: Main Result II

Recall : $C \preceq_\imath P_i$; $\forall C,\ \forall i \in [m]$

Define : $\mathcal{F} = \{Q\colon Q \preceq_\imath P_i\ ;\ i \in [m]\}$

$\forall C$, we have that $C \in \mathcal{F}$. From Lemma 2,

$$\exists Q^* \in \mathcal{F} \text{ s.t. } Q \preceq_m Q^* \preceq_\imath P_i \ ;$$

Therefore,

$$C^\star \preceq_m Q^* \preceq_\imath P_i \ ;\ \forall i \in [m].$$

$$\boxed{H_\alpha(C^\star) \geq H_\alpha(Q^*)}$$

*Y. Y. Shkel, and *A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations," in IEEE International Symposium on Information Theory (ISIT), 2023.

Information-spectrum based Lower bound: Main Result II

Recall : $C \preceq_\imath P_i \;\; ; \qquad \forall C, \; \forall i \in [m]$

Define : $\mathcal{F} = \{Q \colon Q \preceq_\imath P_i \; ; \; i \in [m]\}$

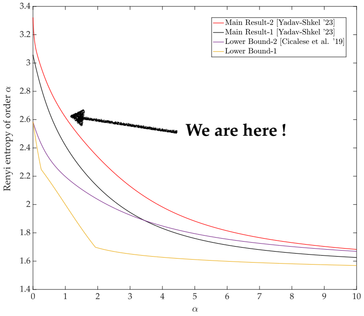
$\forall C$, we have that $C \in \mathcal{F}$. From Lemma 2,

$$\exists Q^* \in \mathcal{F} \;\; \text{s.t.} \;\; Q \preceq_m Q^* \preceq_\imath P_i \;\; ;$$

Therefore,

$$C^\star \preceq_m Q^* \preceq_\imath P_i \; ; \; \forall i \in [m].$$

$$\boxed{H_\alpha(C^\star) \geq H_\alpha(Q^*)}$$

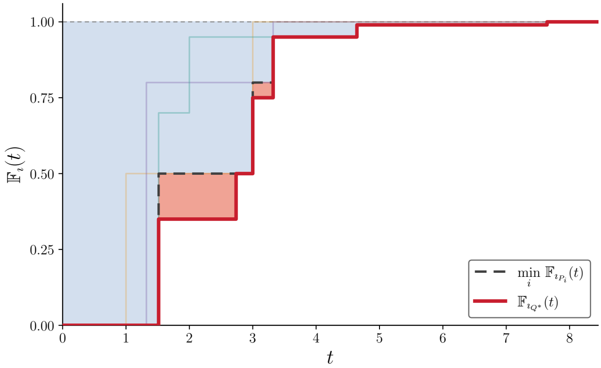


[*] Y. Y. Shkel, and [*] A. K. Yadav, "Information-spectrum converse for minimum entropy couplings and functional representations," in IEEE International Symposium on Information Theory (ISIT), 2023.

Information-spectrum based Lower bound: Main Result 2

Example: Shannon entropy

- Consider $\mathcal{F} = \{Q : Q \preceq_\imath P_i \; ; \; \forall i \in [m]\}$
- Q^*: the $\preceq_m$-greatest element of $\mathcal{F}$ [Lemma 2]
- $H(C^\star) \geq H(Q^*) \geq K_1(\mathcal{S})$: Q^* recovers strictly more area.



Outline

Upper bounds (Achievability Results)

Can we construct 'nearly-optimal' couplings and give approximation guarantees w.r.t $H_\alpha(C^\star)$?

Approximation Algorithms

- Greedy Coupling Algorithm [Painsky et al.]
  - Let C_G denote the output of the algorithm.
- Comonotone Coupling Algorithm [Yadav-Shkel]
  - Let C_M denote the output of the algorithm.

A. Painsky, S. Rosset, and M. Feder, "Memoryless representation of Markov processes", in IEEE International Symposium on Information Theory (ISIT), 2013.

A. K. Yadav, and Y. Y. Shkel, "Couplings and Entropic Inequalities on the Majorization Lattice," 2026. (In part, accepted at IEEE Information Theory Workshop (ITW), 2026.)

Approximation Algorithms

- Greedy Coupling Algorithm [Painsky et al.]
  - Let C_G denote the output of the algorithm.
- Comonotone Coupling Algorithm [Yadav-Shkel]
  - Let C_M denote the output of the algorithm.

- $H_\alpha(\bigwedge \mathcal{S})/K_\alpha(\mathcal{S}) \leq H_\alpha(Q^*) \leq H_\alpha(C^\star)$ - - [from the Lower bound]
- $H_\alpha(C^\star) \leq H_\alpha(C_G)/H_\alpha(C_M)$ - - [problem's nature]

Approximation Algorithms

- Greedy Coupling Algorithm [Painsky et al.]
  - Let C_G denote the output of the algorithm.
- Comonotone Coupling Algorithm [Yadav-Shkel]
  - Let C_M denote the output of the algorithm.
- $H_\alpha(\bigwedge \mathcal{S})/K_\alpha(\mathcal{S}) \leq H_\alpha(Q^*) \leq H_\alpha(C^\star)$ - - [from the Lower bound]
- $H_\alpha(C^\star) \leq H_\alpha(C_G)/H_\alpha(C_M)$ - - [problem's nature]
- **Our Goal :** $H_\alpha(C_G)/H_\alpha(C_M) \leq H_\alpha(C^\star) + Q$; (smallest Q for every $\alpha \in [0, \infty]$).

Greedy Coupling Algorithm

- **Input :** m PMFs $\{P_i\}_{i=1}^m$, each with $\leq n$ states
- **Output :** Coupling $C_G := (c_1, c_2, \ldots, c_T)$
- **Algorithm :**
  - Sort each PMF in the non-increasing order
  - Find the minimum of maximum of each PMF i.e., $q = \min_i(P_i(1))$

$$\left.\begin{array}{l} \circledast \text{ Append } q \text{ as the next state of } C_G \\ \circledast \text{ Update the maximum state of each PMF} \\ \quad \text{i.e., } P_i(1) = \big(P_i(1) - q\big)\,,\ \forall\, i \leq m \\ \circledast \text{ Sort each PMF in non-increasing order} \\ \circledast \text{ Find } q = \min_i(P_i(1)) \end{array}\right\} \quad \begin{array}{c} \textbf{Repeat} \\ \textbf{until} \\ \boldsymbol{q > 0} \end{array}$$

Greedy Coupling Algorithm : Example

- **Input :** $\{P_1 = (0.5, 0.4, 0.1);\ P_2 = (0.6, 0.2, 0.2)\};\ \ (m = 2\,,\ n = 3)$

Iteration (t)	**Current PMFs**	**q**	**Updated PMFs**	C_G
1	(0.5, 0.4, 0.1) (0.6, 0.2, 0.2)	0.5	(0, 0.4, 0.1) (0.1, 0.2, 0.2)	(0.5)
2	(0.4, 0.1, 0) (0.2, 0.2, 0.1)	0.2	(0.2, 0.1, 0) (0, 0.2, 0.1)	(0.5, 0.2)
3	(0.2, 0.1, 0) (0.2, 0.1, 0)	0.2	(0, 0.1, 0) (0, 0.1, 0)	(0.5, 0.2, 0.2)
$\mathrm{T} = 4$	(0.1, 0, 0) (0.1, 0, 0)	0.1	(0, 0, 0) (0, 0, 0)	(0.5, 0.2, 0.2, 0.1)
5	(0, 0, 0) (0, 0, 0)	0	(0, 0, 0) (0, 0, 0)	

- **Output :** Coupling $C_G = (0.5, 0.2, 0.2, 0.1)$

Main Results

Theorem: Let $\mathcal{S} := \{P_1, \ldots, P_m\}$ be the set of m marginal distributions of support size n. Then, for any $\alpha \in [0, \infty)$, we have

$$H_\alpha(C_G) \leq K_\alpha(\mathcal{S}) + F(\alpha, m)$$

where,

$$F(\alpha, m) = \frac{1}{\alpha - 1} \log \left[r(\alpha, m) \right]$$
$$r(\alpha, m) = \max(0, 1 - \tilde{r}(\alpha, m)).$$

where, $$\tilde{r}(\alpha, m) := \begin{cases} \displaystyle\max_{\substack{w_1=0;\ w_{m+1}=1; \\ w_1<w_2\leq w_3\leq\ldots\leq w_m<w_{m+1}}} \sum_{k=2}^{m} w_k \left(w_k^{\alpha-1} - w_{k+1}^{\alpha-1}\right); & \text{for } \alpha \in [0,1), \\ \displaystyle\min_{\substack{w_1=0;\ w_{m+1}=1; \\ w_1<w_2\leq w_3\leq\ldots\leq w_m<w_{m+1}}} \sum_{k=2}^{m} w_k \left(w_k^{\alpha-1} - w_{k+1}^{\alpha-1}\right); & \text{for } \alpha \in (1,\infty). \end{cases}$$

A. K. Yadav, and Y. Y. Shkel, "Approximation Guarantees for Minimum Rényi Entropy Functional Representations", in IEEE International Symposium on Information Theory (ISIT), 2025.

Main Results

Corollary 1 : Let $\mathcal{S} := \{P_1, P_2\}$ be the set of two marginal distributions of support size n. Then, for any $\alpha \in [0, \infty)$, we have

$$H_\alpha(C_G) \leq K_\alpha(\mathcal{S}) + F(\alpha, 2)$$

where, $$F(\alpha, 2) = \frac{1}{\alpha - 1} \log \left[1 + \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} \right].$$

A. K. Yadav, and Y. Y. Shkel, "Approximation Guarantees for Minimum Rényi Entropy Functional Representations", in IEEE International Symposium on Information Theory (ISIT), 2025.

Main Results

Corollary 1 : Let $\mathcal{S} := \{P_1, P_2\}$ be the set of two marginal distributions of support size n. Then, for any $\alpha \in [0, \infty)$, we have

$$H_\alpha(C_G) \leq K_\alpha(\mathcal{S}) + F(\alpha, 2)$$

where, $$F(\alpha, 2) = \frac{1}{\alpha - 1} \log \left[1 + \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} \right].$$

$F(\alpha, m)$ does not have a closed-form solution, in general!

A. K. Yadav, and Y. Y. Shkel, "Approximation Guarantees for Minimum Rényi Entropy Functional Representations", in IEEE International Symposium on Information Theory (ISIT), 2025.

Main Results

Recall that $F(\alpha, m) = \dfrac{1}{\alpha - 1} \log \left[r(\alpha, m) \right]$; where
$r(\alpha, m) = \max(0, 1 - \tilde{r}(\alpha, m))$ such that

$$\tilde{r}(\alpha, m) := \begin{cases} \max\limits_{\substack{w_1=0;\ w_{m+1}=1; \\ w_1 < w_2 \le w_3 \le \ldots \le w_m < w_{m+1}}} \sum_{k=2}^{m} w_k \left(w_k^{\alpha-1} - w_{k+1}^{\alpha-1} \right); & \text{for } \alpha \in [0, 1), \\ \min\limits_{\substack{w_1=0;\ w_{m+1}=1; \\ w_1 < w_2 \le w_3 \le \ldots \le w_m < w_{m+1}}} \sum_{k=2}^{m} w_k \left(w_k^{\alpha-1} - w_{k+1}^{\alpha-1} \right); & \text{for } \alpha \in (1, \infty). \end{cases}$$

Lemma: For every $\alpha \in [0, \infty)$, $F(\alpha, m)$ is a non-decreasing function of m.

As $m \to \infty$, $F(\alpha, m)$ approaches $\dfrac{1}{\alpha - 1} \log \left[\max \left(0, \dfrac{2\alpha - 1}{\alpha} \right) \right]$.

A. K. Yadav, and Y. Y. Shkel, "Approximation Guarantees for Minimum Rényi Entropy Functional Representations", in IEEE International Symposium on Information Theory (ISIT), 2025.

Main Results

Corollary 2 : Let $\mathcal{S} := \{P_1, \ldots, P_m\}$ be the set of m marginal distributions of support size n. Then, for any $\alpha \in [0, \infty)$, we have

$$H_\alpha(C_G) \leq K_\alpha(\mathcal{S}) + \lim_{m\to\infty} F(\alpha, m)$$

$$= K_\alpha(\mathcal{S}) + \frac{1}{\alpha - 1} \log\left[\max\left(0, \frac{2\alpha - 1}{\alpha}\right)\right].$$

A. K. Yadav, and Y. Y. Shkel, "Approximation Guarantees for Minimum Rényi Entropy Functional Representations", in IEEE International Symposium on Information Theory (ISIT), 2025.

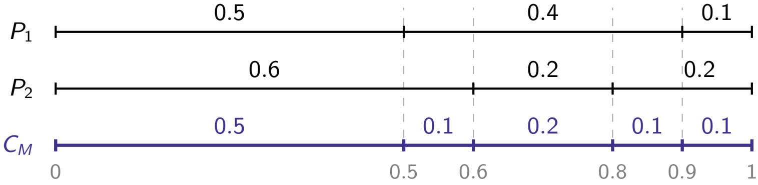
Comonotone Coupling Algorithm

- **Input :** m PMFs $\{P_i\}_{i=1}^m$, each with $\leq n$ states
- **Output :** Coupling $C_M := (c_1, c_2, \ldots, c_W)$
- **Algorithm :**
  - Sort each PMF in the non-increasing order
  - Compute the cumulative breakpoints of each PMF i.e.,
    $\mathcal{B}_i = \left\{ \sum_{j=1}^{k} P_i(j) \,:\, k \in [n-1] \right\}$
  - Merge and sort the distinct breakpoints:
    $\mathcal{B} = \{0, 1\} \cup \bigcup_{i=1}^{m} \mathcal{B}_i \quad$ i.e., $\quad 0 = b_0 < b_1 < \cdots < b_W = 1$
  - Set $c_t = b_t - b_{t-1} \;\; ; \;\; \forall\, t \in [W]$
  - Sort $(c_1, \ldots, c_W)$ in the non-increasing order

A. K. Yadav, and Y. Y. Shkel, "Couplings and Entropic Inequalities on the Majorization Lattice," 2026. (In part, accepted at IEEE Information Theory Workshop (ITW), 2026.)

Comonotone Coupling Algorithm : Example

- **Input :** $\{P_1 = (0.5, 0.4, 0.1);\ P_2 = (0.6, 0.2, 0.2)\};\ \ (m = 2\,,\ n = 3)$



- $\mathcal{B}_1 = \{0.5,\ 0.9\};\ \ \mathcal{B}_2 = \{0.6,\ 0.8\} \implies \mathcal{B} = \{0,\ 0.5,\ 0.6,\ 0.8,\ 0.9,\ 1\}$

- **Output :** Coupling $C_M = (0.5,\ 0.2,\ 0.1,\ 0.1,\ 0.1)$

Comonotone Coupling Algorithm : Worst-case Guarantee

Theorem : Let $\mathcal{S} := \{P_1, \ldots, P_m\}$ be the set of m marginal distributions and let C_M denote the output of the comonotone coupling algorithm. Then, for every $\alpha \in [0, \infty)$, we have

$$H_\alpha(C_M) \ \leq \ H_\alpha\Big(\bigwedge_{i=1}^m P_i \Big) + \log_2(m)$$

A. K. Yadav, and Y. Y. Shkel, "Couplings and Entropic Inequalities on the Majorization Lattice," 2026. (In part, accepted at IEEE Information Theory Workshop (ITW), 2026.)

Comonotone Coupling Algorithm : Worst-case Guarantee

Theorem : Let $\mathcal{S} := \{P_1, \ldots, P_m\}$ be the set of m marginal distributions and let C_M denote the output of the comonotone coupling algorithm. Then, for every $\alpha \in [0, \infty)$, we have

$$H_\alpha(C_M) \;\leq\; H_\alpha\Big(\textstyle\bigwedge_{i=1}^m P_i \Big) + \log_2(m)$$

Proof idea :

$$\Big(\bigwedge_{i=1}^{m} P_i \Big) \otimes \Big(\tfrac{1}{m}, \ldots, \tfrac{1}{m} \Big) \;\preceq_m\; C_M \;\preceq_m\; \bigwedge_{i=1}^{m} P_i$$

Schur concavity of H_α; the uniform factor contributes the $\log_2(m)$.

A. K. Yadav, and Y. Y. Shkel, "Couplings and Entropic Inequalities on the Majorization Lattice," 2026. (In part, accepted at IEEE Information Theory Workshop (ITW), 2026.)

Summary of Lower and Upper Bounds

- **Converse type results (Lower Bounds) :**
  - Two lower bounds based on ‘Information-spectrum majorization’.

Summary of Lower and Upper Bounds

- **Converse type results (Lower Bounds) :**
  - Two lower bounds based on 'Information-spectrum majorization'.
  - $K_\alpha(\mathcal{S})$ is better for lower values of α.
  - $H_\alpha(Q^*)$ is better than all the previously known lower bounds for any $\alpha \in [0, \infty)$.

Summary of Lower and Upper Bounds

- **Converse type results (Lower Bounds) :**
  - Two lower bounds based on 'Information-spectrum majorization'.
  - $K_\alpha(\mathcal{S})$ is better for lower values of α.
  - $H_\alpha(Q^*)$ is better than all the previously known lower bounds for any $\alpha \in [0, \infty)$.
- **Achievability type results (Upper Bounds) :**
  - Approx. analysis between the Rényi entropy of the 'output of the coupling algorithms' and the 'optimal coupling'.

Summary of Lower and Upper Bounds

- **Converse type results (Lower Bounds) :**
  - Two lower bounds based on 'Information-spectrum majorization'.
  - $K_\alpha(\mathcal{S})$ is better for lower values of α.
  - $H_\alpha(Q^*)$ is better than all the previously known lower bounds for any $\alpha \in [0, \infty)$.

- **Achievability type results (Upper Bounds) :**
  - Approx. analysis between the Rényi entropy of the 'output of the coupling algorithms' and the 'optimal coupling'.
  - Analysis of the 'Greedy Coupling algorithm'

    For Shannon entropy: additive gap ≈ 0.53 ($m = 2$)
  - Analysis of the 'Comonotone Coupling algorithm'

    For Shannon entropy: additive gap ≈ 1 ($m = 2$)

Summary of Lower and Upper Bounds

- **Converse type results (Lower Bounds) :**
  - Two lower bounds based on 'Information-spectrum majorization'.
  - $K_\alpha(\mathcal{S})$ is better for lower values of α.
  - $H_\alpha(Q^*)$ is better than all the previously known lower bounds for any $\alpha \in [0, \infty)$.

- **Achievability type results (Upper Bounds) :**
  - Approx. analysis between the Rényi entropy of the 'output of the coupling algorithms' and the 'optimal coupling'.
  - Analysis of the 'Greedy Coupling algorithm'

    For Shannon entropy: additive gap ≈ 0.53 ($m = 2$)
  - Analysis of the 'Comonotone Coupling algorithm'

    For Shannon entropy: additive gap ≈ 1 ($m = 2$)
  - Neither algorithm instance-wise dominates the other
  - Both the algorithms are optimal for min-entropy i.e., $\alpha \to \infty$.

Outline

Applications : Functional Representation Lemma (FRL)

- Given jointly distributed discrete random variables (X, Y).
- **FRL:** There exists a random variable Z such that

$$X = g(Y, Z) \quad \text{and} \quad Y \perp Z$$

- Applications in network information theory, privacy, and causal inference.

A. El Gamal and Y. H. Kim, "Network Information Theory", Cambridge University Press, 2011.

Applications : Functional Representation Lemma (FRL)

- Given jointly distributed discrete random variables (X, Y).
- **FRL:** There exists a random variable Z such that

$$X = g(Y, Z) \quad \text{and} \quad Y \perp Z$$

- Applications in network information theory, privacy, and causal inference.
- **Goal:** What is the most efficient representation of Z i.e., $\min H_\alpha(Z)$?

Applications : Functional Representation Lemma (FRL)

- Given jointly distributed discrete random variables (X, Y).
- **FRL:** There exists a random variable Z such that

$$X = g(Y, Z) \quad \text{and} \quad Y \perp Z$$

- Applications in network information theory, privacy, and causal inference.
- **Goal:** What is the most efficient representation of Z i.e., $\min H_\alpha(Z)$?
- Solving for $\min H_\alpha(Z)$ is equivalent to solving M-REC on $\{P_{X|Y=y_i}\}_{i=1}^m$.
- An Optimal Z can be chosen to be isomorphic to a minimum Rényi entropy coupling on $\{P_{X|Y=y_i}\}_{i=1}^m$.

Applications : Causal Inference

- Given jointly distributed discrete random variables (X, Y).
- **Goal:** Identify the direction of causation i.e., $X \to Y$ or $Y \to X$?
- Entropy based approach to causal identifiability [Kocaoglu et al. '17]

$X \to Y$

Find exogenous random variable E s.t. $X \perp E$ and $Y = f(X, E)$ with minimum $H_\alpha(E)$.

$Y \to X$

Find exogenous random variable $\tilde{E}$ s.t. $Y \perp \tilde{E}$ and $X = g(Y, \tilde{E})$ with minimum $H_\alpha(\tilde{E})$.

Murat Kocaoglu, Alexandros G. Dimakis, Sriram Vishwanath, and Babak Hassibi, "Entropic causal inference", In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI), 2017.

Applications : Causal Inference

- Given jointly distributed discrete random variables (X, Y).
- **Goal:** Identify the direction of causation i.e., $X \to Y$ or $Y \to X$?
- Entropy based approach to causal identifiability [Kocaoglu et al. '17]

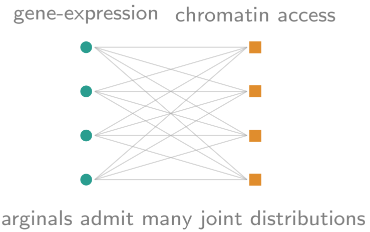
$X \to Y$	$Y \to X$
Find exogenous random variable E s.t. $X \perp E$ and $Y = f(X, E)$ with minimum $H_\alpha(E)$.	Find exogenous random variable $\tilde{E}$ s.t. $Y \perp \tilde{E}$ and $X = g(Y, \tilde{E})$ with minimum $H_\alpha(\tilde{E})$.

- $X \to Y$ if $H_\alpha(X) + H_\alpha(E^*) < H_\alpha(Y) + H_\alpha(\tilde{E}^*)$ and vice-versa.

Murat Kocaoglu, Alexandros G. Dimakis, Sriram Vishwanath, and Babak Hassibi, "Entropic causal inference", In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI), 2017.

Applications : Multi-Modal Learning

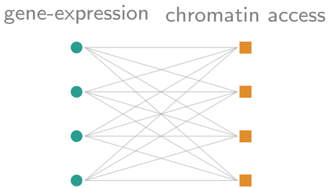
- Real-world multimodal data is often **unpaired** : unpaired image domains ; RNA and ATAC profiles without known cross-modal correspondences.
- Each modality is observed only through its **marginal** — which joint distribution should we learn?



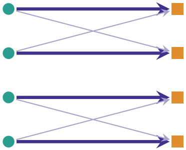
M. Bounoua, G. Franzese, and P. Michiardi, "Learning to Match Unpaired Data with Minimum Entropy Coupling," International Conference on Machine Learning (ICML), 2025.

Applications : Multi-Modal Learning

- Real-world multimodal data is often **unpaired** : unpaired image domains ; RNA and ATAC profiles without known cross-modal correspondences.
- Each modality is observed only through its **marginal** — which joint distribution should we learn?



marginals admit many joint distributions



minimum entropy coupling — the most informative dependence

- Continuous MEC via two cooperative conditional **diffusion models** (DDMEC).

M. Bounoua, G. Franzese, and P. Michiardi, "Learning to Match Unpaired Data with Minimum Entropy Coupling," International Conference on Machine Learning (ICML), 2025.

Other Applications

- Partial Secrecy
- Perfectly Secure Steganography
- Dimensionality Reduction
- Network Information Theory (FRL)
- Contingency Tables
- Transportation Polytopes

Other Applications

- Partial Secrecy
- Perfectly Secure Steganography
- Dimensionality Reduction
- Network Information Theory (FRL)
- Contingency Tables
- Transportation Polytopes

Thank You!