

Couplings and Entropic Inequalities on the Majorization Lattice

Anuj Kumar Yadav, *Student Member, IEEE*, Yanina Y. Shkel, *Member, IEEE*

Abstract—Majorization is a stochastic ordering relation that compares the relative diversity of probability distributions with numerous applications in econometrics, spectral theory, and ecology. It is well known that the majorization partial order forms a complete lattice on the set of all ordered probability distributions on a fixed finite alphabet. In this work, we study the properties of entropic families on the majorization lattice via coupling methods. We establish a fundamental majorization relation between the comonotone and independent couplings. We derive a strengthened Tsallis entropy inequality involving the Gács–Körner common part that holds for every coupling whose bipartite support graph is a forest, including the comonotone coupling. Using these coupling results, we show that both Rényi and Tsallis entropies are subadditive on the majorization lattice for every order $\alpha \geq 0$. Additionally, we study the modular behavior of both entropic families. In particular, we show that Rényi entropy is supermodular on the majorization lattice for $\alpha \in \{0\} \cup [1, \infty]$ but neither supermodular nor submodular in general for $\alpha \in (0, 1)$. The Tsallis entropy satisfies supermodularity for every $\alpha \in [0, \infty)$. Furthermore, we demonstrate applications of our results to minimum Rényi entropy couplings, econometrics, and optimal transport. For the minimum Rényi entropy coupling problem with m marginals, we give a polynomial-time algorithm that constructs the comonotone coupling and show that its Rényi entropy is within $\log m$ nats of the optimum for every finite order α , while it is optimal for min-entropy. We also construct entropy-based distance metrics related to the generalized entropy index in econometrics. Finally, in optimal transport, we bound the excess transport cost introduced by entropic regularization in terms of the marginals alone.

Index Terms—majorization, probability couplings, entropic families, Lorenz curves, econometrics, optimal transport

I. INTRODUCTION

Majorization [2] is a widely studied mathematical concept for comparing the dispersion or inequality of the components of one vector relative to another. Its origin goes back to Schur’s study of symmetric convex functions and determinant inequalities [3] and the work of Hardy, Littlewood, and Pólya [4], [5]. Beyond pure inequality theory, majorization is closely connected to the Lorenz curve [6] and the Pigou–Dalton principle of transfers [7], [8], which provide foundational tools for comparing income and wealth distributions in economics. Thus, majorization offers a natural tool for comparing concentration, dispersion, and inequality across probability distributions and more general vectors.

Majorization-based ideas have appeared in several domains recently. In information theory, [9] studied the relation be-

tween Rényi divergences and majorization through Lorenz diagrams and Markov ordering. In [10], the authors studied matrix majorization in large samples and its catalytic formulation, while Elkouss et al. [11] derived finite sufficient conditions for catalytic majorization and thermal transformations. In wireless communications and signal processing, majorization has been used to compare eigenvalues and power allocation structures in MIMO transceiver design [12]. In quantum information theory, Nielsen [13] characterized pure-state entanglement transformations using majorization, while [14] developed a quantum-majorization framework for state transformations in quantum thermodynamics. Majorization methods also remain central in economics and statistics with recent work on grouped-data estimation and multivariate Lorenz comparisons [15], [16]. In [17], the authors developed Newton-Simpson-type inequalities using majorization theory. It has also been applied to study the minimum entropy couplings [18], functional representations [19]–[21], and causal inference [22]. In [20], the authors introduced a stronger notion of an information-spectrum-based majorization to derive information-theoretic lower bounds on minimum Rényi entropy couplings (see also [18], [23]).

Rényi and Tsallis entropic families are among the most widely studied one-parameter extensions of Shannon entropy. Rényi entropy, originally introduced in [24], appears naturally in source coding with exponential costs [25], guessing [26], privacy amplification [27], hypothesis testing and Rényi-divergence-based analysis [9], [28], and quantum information theory [13], [14]. Tsallis entropy, introduced in [29], is centrally studied in non-extensive statistical mechanics and systems with long-range interactions or memory effects [30], [31] and has also been used in image processing [32], cosmology [33], and machine learning [34].

Bapat [35] showed that the majorization order forms a complete lattice on the set of ordered probability distributions on a fixed finite alphabet. Entropy inequalities on lattices constitute another important line of work. Subadditivity, for instance, bounds a joint entropy using only the marginal entropies; for Shannon entropy, it implies that encoding two sources jointly never requires a higher optimal rate than encoding them separately. On the majorization lattice, the greatest lower bound plays the role of the joint distribution, since every coupling of two marginals is majorized by their greatest lower bound. More generally, submodularity and supermodularity are structural properties that provide monotonicity, convexity, and optimization principles on ordered spaces [36], [37]. On the Boolean lattice formed by

A. K. Yadav and Y. Y. Shkel are with the School of Computer and Communication Sciences (IC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (e-mails: anuj.yadav@epfl.ch; yanina.shkel@epfl.ch). This work has been accepted for publication in part at IEEE ITW 2026 [1].

subsets of a finite set of random variables, Shannon entropy was shown to be submodular [38]. This property underlies the polymatroidal structure of Shannon entropy and many basic information inequalities [38]. It also plays an important role in applications such as secret sharing and the entropy methods in combinatorics [39]. The authors in [40] initiated the analogous study on the majorization lattice, proving that Shannon entropy is supermodular and subadditive on this lattice. They also suggested that these inequalities may support an information-theoretic calculus within the lattice, analogous to joint entropy, conditional entropy, and mutual information in the classical setting. The majorization lattice has been studied in several related directions [41]–[43]. In [44], the authors extended our analysis to the two-parameter Sharma-Mittal entropic family [45]. Following our work, similar results were independently developed in [46], albeit using different proof techniques. More recently, bounds on the total variation distance and Hellinger affinity between ordered distributions were studied in [47].

A. Our Contributions

We study structural properties of couplings, Rényi entropy, and Tsallis entropy on the majorization lattice. Our results recover the subadditivity and supermodularity phenomena of Shannon entropy studied in [40]. Our approach is based on the analysis of couplings and Lorenz-curve geometry, yielding shorter and more conceptual proofs than the case-by-case analysis for the Shannon entropy in [40].

- **Ordering between couplings.** We show that for any two marginal probability mass functions (PMFs) arranged in non-increasing order of mass, the comonotone coupling (also known as the north-west coupling) majorizes the independent coupling.
- **Subadditivity property.** Using the ordering between the comonotone and independent couplings, we prove the subadditivity of Rényi entropy on the majorization lattice for every order $\alpha \in [0, \infty]$. For Tsallis entropy, we establish a stronger coupling inequality involving the entropy of the Gács–Körner common part under the comonotone coupling. As $\alpha \rightarrow 1$, the additional term reduces to the Gács–Körner common information. We further extend this inequality to every coupling whose bipartite support graph is a forest. This inequality enables us to prove the subadditivity of Tsallis entropy for every $\alpha \in [0, \infty)$, including the range $\alpha \in (0, 1)$, where the independent coupling argument does not hold.
- **Supermodularity property.** Using the geometry of couplings and Lorenz curves, we prove that Rényi entropy is supermodular on the majorization lattice for every $\alpha \in \{0\} \cup [1, \infty]$. However, we show that in general the Rényi entropy is neither supermodular nor submodular for every $\alpha \in (0, 1)$. In contrast, we also show that the Tsallis entropy is supermodular on the majorization lattice for every order $\alpha \in [0, \infty)$.
- **Applications.** We demonstrate applications of our results to the minimum Rényi entropy coupling problem,

econometrics, and optimal transport. In particular, given m marginal PMFs, we provide a polynomial-time construction of the comonotone coupling and show that its Rényi entropy exceeds the optimal value by at most $\log m$ nats for every finite order $\alpha \geq 0$. We further obtain a refined distribution-dependent approximation bound that is non-increasing in α , and establish optimality for min-entropy. Using our supermodularity results, we construct two families of information-theoretic distance metrics on ordered PMFs: a Rényi entropy-based family for $\alpha \in [1, \infty)$ and a Tsallis entropy-based family for $\alpha \in (0, \infty)$. Under these metrics, the distance of a PMF to the uniform distribution recovers transformations of the generalized entropy index, including the Theil index, used to measure income inequality. Finally, in optimal transport, we show that every vertex optimal transport plan satisfies our Gács–Körner-strengthened Tsallis entropy inequality, and we bound the additional transport cost due to Shannon entropy regularization in terms of the marginal PMFs alone.

II. NOTATIONS AND PRELIMINARIES

The set of all probability mass functions (PMFs) of size n is denoted by $\mathcal{P}'_n$. Further, we use $\mathcal{P}_n$ to denote the set of all ordered PMFs of size n , with masses in the non-increasing order. Clearly, $\mathcal{P}_n \subset \mathcal{P}'_n$. We denote PMFs by bold-face letters, say $\mathbf{p} := (p_1, p_2, \dots, p_n) \in \mathcal{P}'_n$. Given a random variable X , its support is denoted by $\mathcal{X}$, while a realization is denoted by lower case letter, for example, $x \in \mathcal{X}$. Given a collection of PMFs $\mathcal{S} = \{\mathbf{p}^{(i)}\}_{i=1}^m$, $\mathcal{C}(\mathcal{S})$ denotes the set of all couplings of the marginal PMFs in $\mathcal{S}$. We use $[n]$ to denote the set $\{1, 2, \dots, n\}$. Further, $[a, b]$ denotes the closed interval from a to b where $a \leq b$. Given $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, we use $\Pi(\mathbf{p}, \mathbf{q})$ and $\Pi^\downarrow(\mathbf{p}, \mathbf{q})$ to denote a coupling of $\mathbf{p}$ and $\mathbf{q}$ in the matrix and vector form (with masses arranged in the non-increasing order), respectively. Throughout the paper, $\log$ denotes the natural logarithm, and the entropies are measured in nats unless explicitly stated otherwise.

Definition 1 (Rényi entropy). *The Rényi entropy [24] of order $\alpha \in [0, \infty]$ of a random variable $X \sim \mathbf{p}$ is defined as*

$$H_\alpha(X) = H_\alpha(\mathbf{p}) \triangleq \frac{1}{1-\alpha} \log \left(\sum_{i=1}^n (p_i)^\alpha \right). \quad (1)$$

As $\alpha \rightarrow 1$, the Rényi entropy reduces to the well-known Shannon entropy. As $\alpha \rightarrow \infty$, we have $H_\infty(\mathbf{p}) = -\log \|\mathbf{p}\|_\infty$, where $\|\cdot\|_\infty$ denotes the infinity norm of a vector.

Definition 2 (Tsallis entropy). *The Tsallis entropy [29] of order $\alpha \in [0, \infty)$ of a random variable $X \sim \mathbf{p}$ is defined as*

$$T_\alpha(X) = T_\alpha(\mathbf{p}) \triangleq \frac{1 - \sum_{i=1}^n (p_i)^\alpha}{\alpha - 1}. \quad (2)$$

As $\alpha \rightarrow 1$, the Tsallis entropy also reduces to the Shannon entropy i.e., $T_1(\mathbf{p}) = H(\mathbf{p}) = -\sum_i p_i \log p_i$. An interesting property of Tsallis entropy which makes it different from Rényi

entropy is pseudo-additivity i.e., for two independent random variables X and Y , we have

$$T_\alpha(XY) = T_\alpha(X) + T_\alpha(Y) + (1 - \alpha)T_\alpha(X)T_\alpha(Y), \quad (3)$$

for every $\alpha \in [0, \infty)$.

At order zero, we use the convention $0^0 = 0$ in the entropy sums, so that $H_0(\mathbf{p}) = \log |\text{supp}(\mathbf{p})|$ and $T_0(\mathbf{p}) = |\text{supp}(\mathbf{p})| - 1$. Throughout the paper, we assume that the PMFs are arranged in the non-increasing order i.e., $\mathbf{p} \in \mathcal{P}_n$.

Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, and define the cumulative sums

$$P_i \triangleq \sum_{k=1}^i p_k, \quad \text{and} \quad Q_j \triangleq \sum_{k=1}^j q_k,$$

with the convention $P_0 = Q_0 = 0$. The **independent coupling** $\Pi_I(\mathbf{p}, \mathbf{q})$ is the joint PMF on $[n] \times [n]$ given by

$$\Pi_I(\mathbf{p}, \mathbf{q})(i, j) \triangleq p_i q_j.$$

The **comonotone coupling** $\Pi_c(\mathbf{p}, \mathbf{q})$ (also known as the north-west coupling) is the joint PMF defined by

$$\Pi_c(\mathbf{p}, \mathbf{q})(i, j) \triangleq [\min\{P_i, Q_j\} - \max\{P_{i-1}, Q_{j-1}\}]_+,$$

where $[x]_+ = \max\{x, 0\}$. Equivalently, $\Pi_c(\mathbf{p}, \mathbf{q})(i, j)$ is the length of the overlap between the intervals $(P_{i-1}, P_i]$ and $(Q_{j-1}, Q_j]$.

Definition 3 (Aggregation ($\sqsubseteq$)). Let $\mathbf{q} = (q_1, \dots, q_m)$ and $\mathbf{p} = (p_1, \dots, p_n)$ be probability mass functions with $m \leq n$. We say that $\mathbf{q}$ is an aggregation of $\mathbf{p}$, denoted by $\mathbf{p} \sqsubseteq \mathbf{q}$, if there exists a partition $\{A_1, \dots, A_m\}$ of $[n]$ such that

$$q_i = \sum_{j \in A_i} p_j, \quad (4)$$

for every $i \in [m]$.

Definition 4 (Partial Order ($\preceq$)). Let $\mathcal{P}$ be a set. A homogeneous binary relation on $\mathcal{P}$, denoted by $\preceq \subseteq \mathcal{P} \times \mathcal{P}$, is called a partial order if, $\forall x, y, z \in \mathcal{P}$, the following holds

- (a) **Reflexivity:** $(x, x) \in \preceq, \forall x \in \mathcal{P}$.
- (b) **Antisymmetry:** If $(x, y) \in \preceq$ and $(y, x) \in \preceq$, then $x = y$.
- (c) **Transitivity:** If $(x, y) \in \preceq$ and $(y, z) \in \preceq$, then $(x, z) \in \preceq$.

The pair $(\mathcal{P}, \preceq)$ is called the partially-ordered set (poset). In addition, if for every $x, y \in \mathcal{P}$, either $x \preceq y$ or $y \preceq x$, then $\preceq$ is called a total order.¹

Remark 1. A binary relation that is reflexive and transitive, but not necessarily antisymmetric is called a pre-order or quasi-order.

Now we define the greatest lower bound (glb) and the least upper bound (lub) with respect to a partial order $\preceq$.

Definition 5 (glb ($\wedge$) and lub ($\vee$)). Let $(\mathcal{P}, \preceq)$ be a poset, and let $\mathcal{Q} \subseteq \mathcal{P}$. An element $l \in \mathcal{P}$ is called a lower bound of $\mathcal{Q}$ if $\forall x \in \mathcal{Q}$, we have $l \preceq x$. Let

$$\mathcal{L}(\mathcal{Q}) := \{l \in \mathcal{P} : l \preceq x, \forall x \in \mathcal{Q}\}$$

denote the set of lower bounds of $\mathcal{Q}$. If $\mathcal{L}(\mathcal{Q})$ has a greatest element, then it is called the glb of $\mathcal{Q}$, denoted by $\bigwedge \mathcal{Q}$ i.e., $\bigwedge \mathcal{Q} \in \mathcal{L}(\mathcal{Q})$ and $l \preceq \bigwedge \mathcal{Q}$, for every $l \in \mathcal{L}(\mathcal{Q})$.

Similarly, an element $u \in \mathcal{P}$ is called an upper bound of $\mathcal{Q}$ if $\forall x \in \mathcal{Q}$, we have $x \preceq u$. Let

$$\mathcal{U}(\mathcal{Q}) := \{u \in \mathcal{P} : x \preceq u, \forall x \in \mathcal{Q}\}$$

denote the set of upper bounds of $\mathcal{Q}$. If $\mathcal{U}(\mathcal{Q})$ has a least element, then it is called the lub of $\mathcal{Q}$, denoted by $\bigvee \mathcal{Q}$ i.e., $\bigvee \mathcal{Q} \in \mathcal{U}(\mathcal{Q})$ and $\bigvee \mathcal{Q} \preceq u$, for every $u \in \mathcal{U}(\mathcal{Q})$.

Definition 6 (Lattice). Let $(\mathcal{P}, \preceq)$ be a poset. We say that $\mathcal{P}$ is a lattice if every pair of elements $x, y \in \mathcal{P}$ has both a glb and a lub in $\mathcal{P}$, denoted by $x \wedge y$ and $x \vee y$, respectively. In this case, we denote the lattice by the quadruple $(\mathcal{P}, \preceq, \wedge, \vee)$. If every subset $\mathcal{Q} \subseteq \mathcal{P}$ has both a glb and a lub, then $\mathcal{P}$ is called a complete lattice. Thus, every lattice is a poset, but not every poset is a lattice.

Definition 7 (Gács–Körner common information). Let $(X, Y) \sim \Pi$ be jointly distributed random variables on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. The Gács–Körner common information between X and Y is defined as

$$C_{\text{GK}}(X; Y) \triangleq \max_{\substack{U: H(U|X)=0 \\ H(U|Y)=0}} H(U), \quad (5)$$

where the maximization is over random variables U that can be recovered deterministically from X alone and from Y alone, i.e., $U = f(X) = g(Y)$ almost surely for some functions f and g . The joint distribution of (X, Y) is fixed, and we may take U to have an alphabet $\mathcal{U}$ satisfying $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}|\}$. A maximizing random variable K is called the Gács–Körner common part of X and Y . Consequently, $C_{\text{GK}}(X; Y) = H(K)$.

For $(X, Y) \sim \Pi_c(\mathbf{p}, \mathbf{q})$, we denote the common part by K_c and its PMF by $\mathbf{z}(\mathbf{p}, \mathbf{q})$, and thus $C_{\text{GK}}(\mathbf{p}, \mathbf{q}) = H(K_c)$.

III. THE MAJORIZATION LATTICE

In this section, we define the majorization partial order and state some related properties.

Definition 8 (Majorization). Let $\mathbf{p}$ and $\mathbf{q} \in \mathcal{P}_n$. Then, we say $\mathbf{q}$ majorizes $\mathbf{p}$, denoted by $\mathbf{p} \preceq \mathbf{q}$, if

$$\sum_{i=1}^k p_i \leq \sum_{i=1}^k q_i \quad (6)$$

holds for every $k \in [n]$.

If $\mathbf{p}$ and $\mathbf{q}$ have finite supports with different cardinalities, we pad the shorter PMF with zeros and apply the same definition. The notion of majorization can also be extended to PMFs with countably infinite supports.²

Majorization is a partial order on the set of ordered PMFs i.e., $\mathcal{P}_n$, while it is a pre-order on the set $\mathcal{P}'_n$. If a PMF $\mathbf{q}$ is

²Note that the majorization is not necessarily restricted to sorted PMFs. If the PMFs $\mathbf{p}$ and $\mathbf{q}$ are not ordered in the non-increasing order, we may compare the sums of the k largest probability masses instead of the partial sums.

¹Throughout the paper, we write $(a, b) \in \preceq$ equivalently as $a \preceq b$.

an aggregation of a PMF $\mathbf{p}$ i.e., $\mathbf{p} \sqsubseteq \mathbf{q}$ then $\mathbf{q}$ majorizes $\mathbf{p}$ i.e., $\mathbf{p} \preceq \mathbf{q}$. Thus, aggregation implies majorization.

Definition 9 (Schur-convex function). A function $f : \mathcal{P}_n \rightarrow \mathbb{R}$ is said to be Schur-convex if for every $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ satisfying $\mathbf{p} \preceq \mathbf{q}$, we have

$$f(\mathbf{p}) \leq f(\mathbf{q}) \quad (7)$$

Equivalently, f is called Schur-concave if $-f$ is Schur-convex. Rényi entropy ($H_\alpha(\mathbf{p})$) as well as Tsallis entropy $T_\alpha(\mathbf{p})$ of any order $\alpha \geq 0$ are Schur-concave functions.

Bapat [35] showed that the poset $(\mathcal{P}_n, \preceq)$ induced by majorization is a complete lattice. Cicalese and Vaccaro [40] gave explicit constructions for the greatest lower bound (glb) and the least upper bound (lub) of any two elements. Given $\mathbf{p}$ and $\mathbf{q}$ in $\mathcal{P}_n$, the glb $(\mathbf{p} \wedge \mathbf{q})$ is given by

$$\sum_{i=1}^k (\mathbf{p} \wedge \mathbf{q})(i) = \min \left\{ \sum_{i=1}^k p_i, \sum_{i=1}^k q_i \right\} \quad (8)$$

In order to define the lub, let $\beta_{\mathbf{p}, \mathbf{q}}$ be a vector taking values in $\mathbb{R}^n$, such that

$$\sum_{i=1}^k \beta_{\mathbf{p}, \mathbf{q}}(i) = \max \left\{ \sum_{i=1}^k p_i, \sum_{i=1}^k q_i \right\} \quad (9)$$

If $\beta_{\mathbf{p}, \mathbf{q}} \in \mathcal{P}_n$, then the lub $\mathbf{p} \vee \mathbf{q} = \beta_{\mathbf{p}, \mathbf{q}}$. Otherwise, $\mathbf{p} \vee \mathbf{q}$ is obtained by repeatedly replacing each maximal consecutive block of coordinates of $\beta_{\mathbf{p}, \mathbf{q}}$ that violates the non-increasing order by its average, until the resulting vector lies in $\mathcal{P}_n$. Thus, $\mathbf{p} \vee \mathbf{q}$ is the slope vector of the least concave majorant (LCM) of the cumulative sum curve of $\beta_{\mathbf{p}, \mathbf{q}}$.

Lorenz curves [2], [6] provide a geometric representation of PMFs and are particularly useful in the study of majorization: the majorization order is equivalently characterized by point-wise comparison of the corresponding Lorenz curves. We use this geometric perspective in the sequel to establish our results.

Definition 10 (Lorenz curve). Let $\mathbf{p} = (p_1, \dots, p_n) \in \mathcal{P}_n$. The Lorenz curve of $\mathbf{p}$ is the piecewise-linear concave function $L_{\mathbf{p}} : [0, n] \rightarrow [0, 1]$ obtained by linear interpolation of the points

$$\left(k, \sum_{i=1}^k p_i \right), \quad k \in [n],$$

where $\sum_{i=1}^0 p_i := 0$. Equivalently, $L_{\mathbf{p}}(0) = 0$ and, for $x \in [k-1, k]$ with $k \in [n]$,

$$L_{\mathbf{p}}(x) = \sum_{i=1}^{k-1} p_i + (x - k + 1)p_k.$$

Lemma 1. Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, and let $\Pi_c(\mathbf{p}, \mathbf{q})$ denote their comonotone coupling. Then, the glb and the lub of $\mathbf{p}$ and $\mathbf{q}$ in the majorization lattice are both canonically determined by $\Pi_c(\mathbf{p}, \mathbf{q})$. More precisely, $\mathbf{p} \wedge \mathbf{q}$ is an aggregation of the comonotone coupling $\Pi_c(\mathbf{p}, \mathbf{q})$, while $\mathbf{p} \vee \mathbf{q}$ is obtained by applying the least concave majorant (LCM) operation to an aggregation induced by $\Pi_c(\mathbf{p}, \mathbf{q})$.

Proof. The proof is deferred to Appendix A. It uses the quantile representation of the comonotone coupling and the characterizations of meet and join through cumulative sums. $\square$

Theorem 1. Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. Let $\Pi_c(\mathbf{p}, \mathbf{q})$ and $\Pi_1(\mathbf{p}, \mathbf{q})$ denote the comonotone coupling and the independent coupling of $\mathbf{p}$ and $\mathbf{q}$, respectively. Then, the comonotone coupling vector $\Pi_c^\downarrow$ majorizes the independent coupling vector $\Pi_1^\downarrow$ i.e.,

$$\Pi_1^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \Pi_c^\downarrow(\mathbf{p}, \mathbf{q}). \quad (10)$$

Proof. The proof is deferred to Appendix B. $\square$

Remark 2. Theorem 1 extends to any finite collection of marginal PMFs $\mathcal{S} := \{\mathbf{p}^{(i)}\}_{i=1}^m$, i.e., $\Pi_1^\downarrow(\mathcal{S}) \preceq \Pi_c^\downarrow(\mathcal{S})$. The proof follows the argument in Appendix B, with two modifications: the Young's diagram indexing the k largest masses of the independent coupling is replaced by a down-set $D \subseteq [n]^m$ in the product order, and the monotone-path count is replaced by a breakpoint count. Specifically, the box $[a_1] \times \dots \times [a_m]$, where

$$a_i \triangleq \max_{\mathbf{v} \in D} v_i,$$

has $\Pi_c(\mathcal{S})$ -mass $\min_{i \in [m]} P_{a_i}^{(i)}$ (where $P_k^{(i)} \triangleq \sum_{j=1}^k p_j^{(i)}$) and contains at most $\sum_{i=1}^m a_i - m + 1$ nonzero atoms of $\Pi_c(\mathcal{S})$, while

$$|D| \geq \sum_{i=1}^m a_i - m + 1.$$

IV. SUBADDITIVITY PROPERTY

Definition 11. A real-valued function $f : \mathcal{P} \rightarrow \mathbb{R}$ defined on a lattice $(\mathcal{P}, \preceq, \wedge, \vee)$ is subadditive if $\forall x, y \in \mathcal{P}$, we have

$$f(x \wedge y) \leq f(x) + f(y). \quad (11)$$

In the next theorem, we state a result on the subadditivity of Rényi entropy on the majorization lattice.

Theorem 2. For every $\alpha \in [0, \infty]$, the Rényi entropy H_α is subadditive on the majorization lattice $(\mathcal{P}_n, \preceq, \wedge, \vee)$. More specifically, for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$,

$$\max(H_\alpha(\mathbf{p}), H_\alpha(\mathbf{q})) \leq H_\alpha(\mathbf{p} \wedge \mathbf{q}) \quad (12)$$

$$\leq H_\alpha(\mathbf{p}) + H_\alpha(\mathbf{q}). \quad (13)$$

The equality in (13) holds if at least one of $\mathbf{p}$ and $\mathbf{q}$ is the deterministic PMF $(1, 0, \dots, 0)$.

Proof. From Lemma 1, we know that $\Pi_c(\mathbf{p}, \mathbf{q}) \sqsubseteq \mathbf{p} \wedge \mathbf{q}$. Since aggregation implies majorization, we have that

$$\Pi_c^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \mathbf{p} \wedge \mathbf{q} \quad (14)$$

Furthermore, using Theorem 1 we obtain

$$\Pi_1^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \Pi_c^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \mathbf{p} \wedge \mathbf{q}. \quad (15)$$

Since Rényi entropy is Schur-concave for every $\alpha \in [0, \infty]$,

$$H_\alpha(\mathbf{p} \wedge \mathbf{q}) \leq H_\alpha(\Pi_1^\downarrow(\mathbf{p}, \mathbf{q})) = H_\alpha(\mathbf{p}) + H_\alpha(\mathbf{q}) \quad (16)$$

The above proof can also be obtained without invoking Theorem 1. The independent coupling $\Pi_I(\mathbf{p}, \mathbf{q})$ has marginals $\mathbf{p}$ and $\mathbf{q}$; hence, by aggregation,

$$\Pi_I^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \mathbf{p} \quad \text{and} \quad \Pi_I^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \mathbf{q},$$

Thus $\Pi_I^\downarrow(\mathbf{p}, \mathbf{q})$ is a lower bound of $\mathbf{p}$ and $\mathbf{q}$. By the definition of the greatest lower bound,

$$\Pi_I^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \mathbf{p} \wedge \mathbf{q}. \quad (17)$$

Therefore, (16) follows directly. Nevertheless, Theorem 1 provides a sharper structural comparison between the independent and comonotone couplings, and we use it here to emphasize this relation.

(Conditions for equality). Now, it remains to establish the equality cases. First, if one of $\mathbf{p}$ and $\mathbf{q}$ is deterministic, say $\mathbf{p} = (1, 0, \dots, 0)$, then we have $\mathbf{p} \wedge \mathbf{q} = \mathbf{q}$. Consequently,

$$H_\alpha(\mathbf{p} \wedge \mathbf{q}) = H_\alpha(\mathbf{q}) = H_\alpha(\mathbf{p}) + H_\alpha(\mathbf{q}), \quad (18)$$

since $H_\alpha(\mathbf{p}) = 0$. A similar argument applies if $\mathbf{q}$ is deterministic. This completes our proof. $\square$

Furthermore, the Schur-concavity of Rényi entropy yields the following simple upper bound,

Remark 3. For all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ and $\alpha \in [0, \infty]$, we have

$$H_\alpha(\mathbf{p}) + H_\alpha(\mathbf{q}) \leq 2H_\alpha(\mathbf{p} \wedge \mathbf{q}). \quad (19)$$

For $\alpha \in (0, \infty)$, the equality holds if and only if $\mathbf{p} = \mathbf{q}$.

As a consequence of Theorem 2 and Remark 3, we have the following corollary which extends the above property to any finite number of PMFs,

Corollary 1. Let $\{\mathbf{p}^{(i)}\}_{i=1}^m$ be m PMFs in $\mathcal{P}_n$. Then, for Rényi entropy of any order $\alpha \in [0, \infty]$, we have that

$$H_\alpha\left(\bigwedge_{i=1}^m \mathbf{p}^{(i)}\right) \leq \sum_{i=1}^m H_\alpha(\mathbf{p}^{(i)}) \leq mH_\alpha\left(\bigwedge_{i=1}^m \mathbf{p}^{(i)}\right). \quad (20)$$

Proof. The first inequality follows by applying Theorem 2 (on the pair of PMFs) iteratively to the glb $\bigwedge_{i=1}^m \mathbf{p}^{(i)}$, for $m - 1$ times.

For the second inequality, since $\bigwedge_{i=1}^m \mathbf{p}^{(i)}$ is a lower bound, w.r.t majorization, for every $\mathbf{p}^{(i)}$, we have $\bigwedge_{i=1}^m \mathbf{p}^{(i)} \preceq \mathbf{p}^{(i)}$, for all i . By Schur-concavity of Rényi entropy, $H_\alpha(\mathbf{p}^{(i)}) \leq H_\alpha(\bigwedge_{i=1}^m \mathbf{p}^{(i)})$ for every i , and summing over $i \in [m]$ gives us the result. $\square$

It was shown in [48] that Tsallis entropy is subadditive for order $\alpha \geq 1$. In the next theorem, we show that the subadditivity property extends to all $\alpha \in [0, \infty)$ using our proof technique based on couplings involving the Gács–Körner common part. We first state a lemma which will be used to prove our subadditivity result.

Lemma 2. Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ such that $X \sim \mathbf{p}$, $Y \sim \mathbf{q}$, and $(X, Y) \sim \Pi_c(\mathbf{p}, \mathbf{q})$. Let K_c denote the Gács–Körner common

part of X and Y , and let $\mathbf{z}(\mathbf{p}, \mathbf{q})$ denote the PMF of K_c . Then, for every $\alpha \in [0, \infty)$, we have

$$T_\alpha(\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})) + T_\alpha(\mathbf{z}(\mathbf{p}, \mathbf{q})) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}). \quad (21)$$

For $\alpha \rightarrow 1$, $T_\alpha(\mathbf{z}) \rightarrow C_{\text{GK}}(\mathbf{p}, \mathbf{q})$, where $C_{\text{GK}}(\mathbf{p}, \mathbf{q})$ is the Gács–Körner common information [49] of $\mathbf{p}$ and $\mathbf{q}$, measured in nats, under the comonotone coupling. Furthermore, (21) is satisfied with equality for every $\mathbf{p}, \mathbf{q}$ if $\alpha = 0$. The equality also holds if either $\mathbf{p}$ or $\mathbf{q}$ is a deterministic PMF $(1, 0, \dots, 0)$.

Proof. The proof is deferred to Appendix C. $\square$

Every coupling Π between two marginal PMFs induces a bipartite support graph which includes vertex classes corresponding to the positive mass atoms of its marginals, with an edge between x_i and y_j if and only if $\Pi_{ij} > 0$. The bipartite support graph induced by the comonotone coupling is a forest [50]. In the next proposition, we show that the result in Lemma 2 can in fact be extended to every coupling Π whose bipartite support graph is a forest.

Proposition 1. Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, and let Π be a coupling of $X \sim \mathbf{p}$ and $Y \sim \mathbf{q}$ whose bipartite support graph is a forest. Let K_Π denote the Gács–Körner common part of X and Y under $(X, Y) \sim \Pi$, and let $\mathbf{z}_\Pi$ denote its PMF. Then, for every $\alpha \in [0, \infty)$, we have

$$T_\alpha(\Pi^\downarrow) + T_\alpha(\mathbf{z}_\Pi) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}). \quad (22)$$

Equality holds for $\alpha = 0$.

The proof is deferred to Appendix D. In the next result, we show that Tsallis entropy is subadditive on the majorization lattice for every $\alpha \geq 0$.

Theorem 3. For every $\alpha \in [0, \infty)$, the Tsallis entropy T_α is subadditive on the majorization lattice $(\mathcal{P}_n, \preceq, \wedge, \vee)$. More specifically, for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$,

$$\max(T_\alpha(\mathbf{p}), T_\alpha(\mathbf{q})) \leq T_\alpha(\mathbf{p} \wedge \mathbf{q}) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}). \quad (23)$$

The equality in (23) holds if at least one of $\mathbf{p}$ and $\mathbf{q}$ is the deterministic PMF $(1, 0, \dots, 0)$.

Proof. Unlike the Rényi entropy, the Tsallis entropy only satisfies pseudoadditivity under the independent coupling i.e., given PMFs $\mathbf{p}$ and $\mathbf{q}$, we have that $T_\alpha(\Pi_I^\downarrow(\mathbf{p}, \mathbf{q})) = T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}) + (1 - \alpha)T_\alpha(\mathbf{p})T_\alpha(\mathbf{q})$, for every $\alpha \in [0, \infty)$. Thus, the relation between the independent coupling and comonotone coupling can be leveraged to prove subadditivity on the majorization lattice only for $\alpha \geq 1$.

However, we bypass this approach and instead use Lemma 2 to prove our result. Since $T_\alpha(\cdot) \geq 0$, from (21) we have that

$$T_\alpha(\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}). \quad (24)$$

From Lemma 1, we know that $\Pi_c^\downarrow(\mathbf{p}, \mathbf{q}) \sqsubseteq \mathbf{p} \wedge \mathbf{q}$. Since aggregation implies majorization, we have that $\Pi_c^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \mathbf{p} \wedge \mathbf{q}$. Using Schur concavity of Tsallis entropy, for every $\alpha \in [0, \infty)$ we have that

$$T_\alpha(\mathbf{p} \wedge \mathbf{q}) \leq T_\alpha(\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})) \quad (25)$$

Consequently, from (24) and (25), we obtain that $T_\alpha(\mathbf{p} \wedge \mathbf{q}) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q})$, for every $\alpha \in [0, \infty)$.

(Conditions for equality). If one of $\mathbf{p}$ and $\mathbf{q}$ is deterministic, say $\mathbf{p} = (1, 0, \dots, 0)$, then we have $\mathbf{p} \wedge \mathbf{q} = \mathbf{q}$. Consequently, $T_\alpha(\mathbf{p} \wedge \mathbf{q}) = T_\alpha(\mathbf{q}) = T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q})$, since $T_\alpha(\mathbf{p}) = 0$. A similar argument applies if $\mathbf{q}$ is deterministic. This completes our proof. $\square$

Similar to the case of Rényi entropy in Corollary 1, the extension of subadditivity to the finite collection of PMFs also holds for the Tsallis entropy of any order $\alpha \in [0, \infty)$.

V. SUPERMODULARITY PROPERTY

Definition 12. A real-valued function $f : \mathcal{P} \rightarrow \mathbb{R}$ defined on a lattice $(\mathcal{P}, \preceq, \wedge, \vee)$ is said to be supermodular if $\forall x, y \in \mathcal{P}$, we have

$$f(x) + f(y) \leq f(x \wedge y) + f(x \vee y). \quad (26)$$

Equivalently, f is submodular iff $-f$ is supermodular.

In the next theorem, we state a result on the supermodularity of Rényi entropy on the majorization lattice,

Theorem 4. For every $\alpha \in \{0\} \cup [1, \infty]$, the Rényi entropy H_α is supermodular on the majorization lattice $(\mathcal{P}_n, \preceq, \wedge, \vee)$. That is, for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$,

$$H_\alpha(\mathbf{p}) + H_\alpha(\mathbf{q}) \leq H_\alpha(\mathbf{p} \wedge \mathbf{q}) + H_\alpha(\mathbf{p} \vee \mathbf{q}). \quad (27)$$

Moreover, for $\alpha \in \{0, \infty\}$, equality holds for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. Thus, $H_0(\cdot)$ and $H_\infty(\cdot)$ are modular on the majorization lattice.

Proof. We leverage the idea of Lorenz curves to represent the glb and lub. We define the following shorthands: $\mathbf{w} := \mathbf{p} \wedge \mathbf{q}$, $\mathbf{v} := \mathbf{p} \vee \mathbf{q}$, and $S_\alpha(\mathbf{p}) := \sum_{i=1}^n p_i^\alpha$. Thus, we need to show that for $\alpha > 1$,

$$\begin{aligned} & \frac{1}{1-\alpha} \left(\log(S_\alpha(\mathbf{p})) + \log(S_\alpha(\mathbf{q})) \right) \\ & \leq \frac{1}{1-\alpha} \left(\log(S_\alpha(\mathbf{w})) + \log(S_\alpha(\mathbf{v})) \right) \end{aligned} \quad (28)$$

i.e., $S_\alpha(\mathbf{p})S_\alpha(\mathbf{q}) \geq S_\alpha(\mathbf{w})S_\alpha(\mathbf{v})$. From Definition 10 of Lorenz curve, we note that for every $\mathbf{x} \in \mathcal{P}_n$,

$$S_\alpha(\mathbf{x}) = \int_0^n L'_x(t)^\alpha dt \quad (29)$$

Let $L_{\mathbf{p}}$ and $L_{\mathbf{q}}$ denote the Lorenz curves of $\mathbf{p}$ and $\mathbf{q}$, respectively. Then, we have that for all $k \in [n]$, the Lorenz curves of $\mathbf{w}$ and $\mathbf{v}$ satisfy

$$L_{\mathbf{w}}(k) = \min\{L_{\mathbf{p}}(k), L_{\mathbf{q}}(k)\}, \quad (30)$$

$$L_{\mathbf{v}}(k) = [\text{LCM}(F_{\mathbf{v}})](k), \quad (31)$$

where $F_{\mathbf{v}}(k) := \max\{L_{\mathbf{p}}(k), L_{\mathbf{q}}(k)\}$ and $\text{LCM}(\cdot)$ denotes the least concave majorant operation. Further, let $m(t)$ and $M(t)$ be the continuous envelopes defined for all $t \in [0, n]$,

$$m(t) = \min\{L_{\mathbf{p}}(t), L_{\mathbf{q}}(t)\}, \quad (32)$$

$$M(t) = \max\{L_{\mathbf{p}}(t), L_{\mathbf{q}}(t)\}. \quad (33)$$

Within any interval of integer points or crossings between $L_{\mathbf{p}}$ and $L_{\mathbf{q}}$, we have that

$$\begin{aligned} m &= L_{\mathbf{p}}, \quad M = L_{\mathbf{q}}, \quad \text{if } L_{\mathbf{p}} \leq L_{\mathbf{q}} \\ m &= L_{\mathbf{q}}, \quad M = L_{\mathbf{p}}, \quad \text{if } L_{\mathbf{q}} \leq L_{\mathbf{p}}. \end{aligned}$$

Thus, within any continuous interval $(m'(t), M'(t))$ is a permutation of $(L'_{\mathbf{p}}(t), L'_{\mathbf{q}}(t))$. Consequently, for almost every $t \in [0, n]$,

$$(m'(t))^\alpha + (M'(t))^\alpha = (L'_{\mathbf{p}}(t))^\alpha + (L'_{\mathbf{q}}(t))^\alpha. \quad (34)$$

On integrating both sides with respect to $t \in [0, n]$,

$$\int_0^n (m'(t))^\alpha dt + \int_0^n (M'(t))^\alpha dt = S_\alpha(\mathbf{p}) + S_\alpha(\mathbf{q}). \quad (35)$$

Fix an interval $I_k \triangleq (k-1, k)$. Then, m is a piece-wise linear concave function within I_k while $L_{\mathbf{w}}$ is a straight line joining $m(k-1)$ to $m(k)$. Thus,

$$L'_{\mathbf{w}}(t) = \frac{m(k) - m(k-1)}{1} = \int_{k-1}^k m'(t) dt$$

Since $u \rightarrow u^\alpha$ is a convex map for $\alpha \geq 1$, using Jensen's inequality we have

$$(L'_{\mathbf{w}}(t))^\alpha \leq \int_{k-1}^k m'(t)^\alpha dt \quad (36)$$

Now, considering M in the interval I_k . Unlike m , M is a piece-wise linear (not necessarily concave) function while $F_{\mathbf{v}}$ is a straight line joining $M(k-1)$ to $M(k)$. If I_k has no non-integer crossings, then $F'_{\mathbf{v}}(t) = M'(t)$.

However, if there are crossings then M is convex and we need to concavify it to get $F_{\mathbf{v}}$. Thus,

$$F'_{\mathbf{v}}(t) = \frac{M(k) - M(k-1)}{1} = \int_{k-1}^k M'(t) dt$$

Since $u \rightarrow u^\alpha$ is a convex map for $\alpha > 1$, using Jensen's inequality we have

$$(F'_{\mathbf{v}}(t))^\alpha \leq \int_{k-1}^k (M'(t))^\alpha dt \quad (37)$$

Now, if $F_{\mathbf{v}}$ is a piecewise-linear concave function, then it is the lub $L_{\mathbf{v}}$. Otherwise, it needs to be concavified and consequently the lub $L_{\mathbf{v}}$ is the least concave majorant (LCM) of $F_{\mathbf{v}}$.

Therefore, on applying Jensen again over the pooled intervals used by the LCM operation, we get

$$S_\alpha(\mathbf{v}) = \int_0^n (L'_{\mathbf{v}}(t))^\alpha dt \quad (38)$$

$$\leq \int_0^n (F'_{\mathbf{v}}(t))^\alpha dt \quad (39)$$

$$\leq \int_0^n (M'(t))^\alpha dt. \quad (40)$$

On summing (36) over all intervals I_k for $k \in [n]$ and using (40), we obtain

$$S_\alpha(\mathbf{w}) + S_\alpha(\mathbf{v}) \leq \int_0^n (m'(t))^\alpha dt + \int_0^n (M'(t))^\alpha dt \quad (41)$$

Therefore, from (35) we have

$$S_\alpha(\mathbf{w}) + S_\alpha(\mathbf{v}) \leq S_\alpha(\mathbf{p}) + S_\alpha(\mathbf{q}). \quad (42)$$

Recall that $\mathbf{w} \preceq \mathbf{p}$, $\mathbf{q} \preceq \mathbf{v}$. Since S_α is schur-convex for $\alpha \geq 1$, it implies that $S_\alpha(\mathbf{w}) \leq S_\alpha(\mathbf{p})$, $S_\alpha(\mathbf{q}) \leq S_\alpha(\mathbf{v})$. Let $x, y, z \geq 0$ such that

$$\begin{aligned} S_\alpha(\mathbf{p}) &= S_\alpha(\mathbf{w}) + x \\ S_\alpha(\mathbf{q}) &= S_\alpha(\mathbf{w}) + y \\ S_\alpha(\mathbf{v}) &= S_\alpha(\mathbf{w}) + z \end{aligned}$$

Then, from (42) we have $x + y \geq z$. Now,

$$\begin{aligned} S_\alpha(\mathbf{p})S_\alpha(\mathbf{q}) - S_\alpha(\mathbf{w})S_\alpha(\mathbf{v}) &= (S_\alpha(\mathbf{w}) + x)(S_\alpha(\mathbf{w}) + y) \\ &\quad - S_\alpha(\mathbf{w})(S_\alpha(\mathbf{w}) + z) \quad (43) \\ &= S_\alpha(\mathbf{w})(x + y - z) + xy \quad (44) \\ &\geq 0, \quad (45) \end{aligned}$$

where (45) follows from noting that $x, y \geq 0$. The case $\alpha = 1$ follows by taking the limit $\alpha \rightarrow 1$, since Rényi entropy is continuous in α on finite alphabets.

(Modularity for $\alpha = 0$ and $\alpha = \infty$). Let $a := |\text{supp}(\mathbf{p})|$ and $b := |\text{supp}(\mathbf{q})|$, where $a, b \leq n$. For $\alpha \rightarrow 0$, we have $|\text{supp}(\mathbf{p} \wedge \mathbf{q})| = \max(a, b)$, and $|\text{supp}(\mathbf{p} \vee \mathbf{q})| = \min(a, b)$. Therefore,

$$\begin{aligned} H_0(\mathbf{p} \wedge \mathbf{q}) + H_0(\mathbf{p} \vee \mathbf{q}) &= \log \max(a, b) + \log \min(a, b) \\ &= \log a + \log b \quad (46) \\ &= H_0(\mathbf{p}) + H_0(\mathbf{q}). \quad (47) \end{aligned}$$

Now consider $\alpha \rightarrow \infty$. Since $\mathbf{p}$ and $\mathbf{q}$ are arranged in non-increasing order, thus $p_1 = \max_i p_i$, and $q_1 = \max_i q_i$. Then

$$(\mathbf{p} \wedge \mathbf{q})(1) = \min(p_1, q_1), \quad \text{and} \quad (\mathbf{p} \vee \mathbf{q})(1) = \max(p_1, q_1).$$

Consequently, we have that

$$\begin{aligned} H_\infty(\mathbf{p} \wedge \mathbf{q}) + H_\infty(\mathbf{p} \vee \mathbf{q}) &= -\log \min(p_1, q_1) - \log \max(p_1, q_1) \\ &= -\log p_1 - \log q_1 \\ &= H_\infty(\mathbf{p}) + H_\infty(\mathbf{q}). \end{aligned}$$

Thus, the Rényi entropy is modular for $\alpha \in \{0, \infty\}$. This completes our proof. $\square$

Remark 4. Unlike Corollary 1, the supermodularity property does not extend in the same way to more than two PMFs. In particular, for $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \mathbf{p}^{(3)} \in \mathcal{P}_n$, it is not generally true that

$$\sum_{i=1}^3 H_\alpha(\mathbf{p}^{(i)}) \leq H_\alpha\left(\bigwedge_{i=1}^3 \mathbf{p}^{(i)}\right) + H_\alpha\left(\bigvee_{i=1}^3 \mathbf{p}^{(i)}\right).$$

For example, on taking $\mathbf{p}^{(1)} = \mathbf{p}^{(2)} = \mathbf{p}^{(3)} = \mathbf{p}$ for any non-deterministic PMF $\mathbf{p}$, the left-hand side equals $3H_\alpha(\mathbf{p})$, whereas the right-hand side equals $2H_\alpha(\mathbf{p})$. Thus, the inequality fails whenever $H_\alpha(\mathbf{p}) > 0$.

Theorem 5. For $n \geq 3$ and every $\alpha \in (0, 1)$, the Rényi entropy H_α is neither supermodular nor submodular on the

majorization lattice $(\mathcal{P}_n, \preceq, \wedge, \vee)$.

Proof. The proof is deferred to Appendix E. $\square$

In the next theorem, we prove the supermodularity of Tsallis entropy on the majorization lattice,

Theorem 6. For every $\alpha \in [0, \infty)$, the Tsallis entropy T_α is supermodular on the majorization lattice $(\mathcal{P}_n, \preceq, \wedge, \vee)$. That is, for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$,

$$T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}) \leq T_\alpha(\mathbf{p} \wedge \mathbf{q}) + T_\alpha(\mathbf{p} \vee \mathbf{q}). \quad (48)$$

Moreover, for $\alpha = 0$, equality holds for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. Thus, $T_0(\cdot)$ is modular on the majorization lattice.

Proof. The proof is based on the idea of Lorenz curves, similar to the one in Theorem 4. We defer the details to Appendix F. $\square$

Remark 5. It is stated in [48] (Theorem 3.2) that the Tsallis entropy is submodular on the majorization lattice, for every order $\alpha \in [0, \infty)$. This contradicts our Theorem 6. We note that there is a sign error in one of the steps in the proof in [48]. The argument therein controls the power sum $S_\alpha(\mathbf{p}) := \sum_{i=1}^n p_i^\alpha$, which is submodular on the majorization lattice for $\alpha > 1$ and supermodular for $\alpha < 1$ (by convexity, resp. concavity, of $u \mapsto u^\alpha$). The Tsallis entropy is the affine transform of $S_\alpha(\cdot)$ i.e., $T_\alpha = \frac{1-S_\alpha}{\alpha-1}$, whose slope $\frac{-1}{\alpha-1}$ changes sign at $\alpha = 1$. This sign change exactly compensates the change in the modularity direction of S_α . Consequently, T_α is supermodular for every $\alpha \in [0, \infty)$ — rather than submodular.

As an explicit example, consider $\mathbf{p} = (0.6, 0.2, 0.2)$ and $\mathbf{q} = (0.45, 0.4, 0.15)$, such that $\mathbf{p} \wedge \mathbf{q} = (0.45, 0.35, 0.2)$ and $\mathbf{p} \vee \mathbf{q} = (0.6, 0.25, 0.15)$. On both sides of $\alpha = 1$, we have that

$$\begin{aligned} T_2(\mathbf{p} \wedge \mathbf{q}) + T_2(\mathbf{p} \vee \mathbf{q}) &= 1.190 \\ &> 1.175 = T_2(\mathbf{p}) + T_2(\mathbf{q}), \end{aligned}$$

and

$$\begin{aligned} T_{1/2}(\mathbf{p} \wedge \mathbf{q}) + T_{1/2}(\mathbf{p} \vee \mathbf{q}) &\approx 2.743 \\ &> 2.719 \approx T_{1/2}(\mathbf{p}) + T_{1/2}(\mathbf{q}), \end{aligned}$$

each violating submodularity as stated in [48] and therefore is consistent with our Theorem 6.

VI. APPLICATIONS

A. Minimum Rényi entropy couplings

Given a collection of m marginal PMFs, say $\mathcal{S} := \{\mathbf{p}^{(i)}\}_{i=1}^m$, the minimum Rényi entropy coupling problem [51] refers to finding the joint distribution (coupling) with the minimum joint Rényi entropy $H_\alpha^*(\text{OPT}_{\mathcal{S}})$. However, it was shown in [52] that the problem is NP-Hard in the support size of the marginal PMFs, for every order $\alpha \in [0, \infty)$. Therefore, the existing literature involves computing information-theoretic lower bounds on $H_\alpha^*(\text{OPT}_{\mathcal{S}})$ (converse-type results) and designing approximation algorithms to output a coupling with Rényi entropy close to $H_\alpha^*(\text{OPT}_{\mathcal{S}})$. For ease in comparison, we state the following approximation guarantees from the existing literature in bits.

Cicalese et al. [51] derived the lower bound $H_\alpha(\bigwedge \mathcal{S})$ on the minimum coupling entropy using the greatest lower bound (glb) under majorization. Later, Shkel and Yadav [20] obtained the stronger lower bound $H_\alpha(Q^*)$ using information-spectrum majorization (see also Compton et al. [23] for concurrent work). Therefore, we have

$$H_\alpha\left(\bigwedge \mathcal{S}\right) \leq H_\alpha(Q^*) \leq H_\alpha^*(\text{OPT}_\mathcal{S}) \quad (49)$$

for every order $\alpha \in [0, \infty]$. Several algorithms have been proposed in the literature to compute the sub-optimal couplings. In [51], the authors proposed a polynomial-time algorithm to construct a coupling and showed that its Shannon entropy was within $\lceil \log_2 m \rceil$ bits from the optimal solution (i.e., within 1 bit for $m = 2$). Li [18] introduced a new coupling algorithm based on the geometric splitting technique. Kocaoglu et al. proposed a greedy algorithm to construct a coupling in [22]. Compton et al. [23] showed that the output of the greedy algorithm is always within $\frac{\log_2 e}{e} \approx 0.53$ bits of the optimal Shannon entropy for $m = 2$ and always within $\frac{1 + \log_2 e}{2} \approx 1.22$ bits for an arbitrary m . Yadav and Shkel [21] further provided approximation guarantees for Rényi entropy of any order $\alpha \in [0, \infty)$, showing that the greedy algorithm is optimal for the min-entropy ($\alpha = \infty$).

Motivated by our results in Lemma 1 and Theorem 1, we propose a polynomial-time algorithm (Algorithm 1) for constructing the probability vector induced by a comonotone coupling of a collection of marginal distributions. We use this construction to obtain worst-case additive approximation guarantees for the minimum Rényi entropy coupling problem and show that the comonotone coupling obtained via Algorithm 1 is optimal for min-entropy. Although the greedy coupling algorithm [19], [22] exhibits stronger worst-case guarantees, we show, through various examples, that the comonotone coupling can outperform greedy on structured families of marginal PMFs. We further provide sufficient conditions under which the proposed algorithm is optimal.

Algorithm 1 Comonotone Coupling Algorithm

Input: m marginal distributions $\mathcal{S} = \{\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(m)}\}$, each with n states.

Output: Coupling $\Pi_c^\downarrow(\mathcal{S}) := (c_1, c_2, \dots, c_T)$.

Sort each marginal PMF in non-increasing order.

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

 Compute the set of all breakpoints $\mathcal{B}_i \triangleq \left\{x \in (0, 1) : x = \sum_{j=1}^k p_j^{(i)}, k \in [n-1]\right\}$

end for

Set $\mathcal{B} = \{0, 1\} \cup \bigcup_{i=1}^m \mathcal{B}_i$.

Sort distinct elements of $\mathcal{B}$ in the increasing order i.e., $0 = b_0 < b_1 < b_2 < \dots < b_T = 1$.

for $t = 1, \dots, T$ **do**

 Set $c_t = b_t - b_{t-1}$.

end for

Sort $\Pi_c^\downarrow(\mathcal{S})$ in the non-increasing order.

return Sorted coupling vector $\Pi_c^\downarrow(\mathcal{S})$.

Algorithm 1 outputs at most $T \leq m(n-1) + 1$ masses, so the output size is $O(mn)$. Sorting the m marginal PMFs takes

$O(mn \log n)$, and computing their cumulative breakpoints is $O(mn)$. The breakpoint set $\mathcal{B}$ is the union of m already-sorted lists of total size $O(mn)$, which can be merged in $O(mn \log mn)$ time by a generic sort, similar for $\Pi_c^\downarrow(\mathcal{S})$. Hence the total running time is $O(mn \log(mn))$, polynomial in the number of marginals m and states n , while the space complexity is $O(mn)$.

Now, we state the result on the worst-case approximation guarantee of our proposed algorithm.

Theorem 7. *Given any collection of marginal PMFs $\mathcal{S} \triangleq \{\mathbf{p}^{(i)}\}_{i=1}^m$, let $\Pi_c^\downarrow(\mathcal{S})$ denote the output of Algorithm 1. Then, we have*

$$H_\alpha^*(\text{OPT}_\mathcal{S}) \leq H_\alpha(\Pi_c^\downarrow(\mathcal{S})) \leq H_\alpha^*(\text{OPT}_\mathcal{S}) + \log m \quad (50)$$

for every order $\alpha \in [0, \infty)$. Furthermore, for $\alpha \rightarrow \infty$, Algorithm 1 is optimal i.e., $H_\infty^*(\text{OPT}_\mathcal{S}) = H_\infty(\Pi_c^\downarrow(\mathcal{S}))$.

Proof. By extending the result of Lemma 1 to a finite collection of PMFs, we have that the glb of any finite collection of PMFs i.e., $\bigwedge \mathcal{S}$ is an aggregation of the comonotone coupling constructed in Algorithm 1, $\Pi_c(\mathcal{S})$, i.e., $\Pi_c^\downarrow(\mathcal{S}) \sqsubseteq \bigwedge \mathcal{S}$ thus $\Pi_c^\downarrow(\mathcal{S}) \preceq \bigwedge \mathcal{S}$. Here, we will show that if each element of $\bigwedge \mathcal{S}$ is split equally into m parts then the resulting vector is majorized by the comonotone coupling vector i.e.,

$$\bigwedge \mathcal{S} \otimes \left(\frac{1}{m}, \dots, \frac{1}{m}\right) \preceq \Pi_c^\downarrow(\mathcal{S}) \preceq \bigwedge \mathcal{S}. \quad (51)$$

Using the Schur concavity of the Rényi entropy allows us to prove our result. Detailed proof appears in Appendix G. $\square$

In the next corollary, we strengthen the result of Theorem 7 by providing a distribution-dependent worst-case approximation guarantee on the comonotone coupling in Algorithm 1.

Corollary 2. *Under the assumptions of Theorem 7, let $\bigwedge \mathcal{S} = (w_1, \dots, w_n)$ denote the glb of $\mathcal{S}$ and $s = |\text{supp}(\bigwedge \mathcal{S})|$. Then, for every order $\alpha \in [0, \infty]$,*

$$\begin{aligned} H_\alpha^*(\text{OPT}_\mathcal{S}) &\leq H_\alpha(\Pi_c^\downarrow(\mathcal{S})) \\ &\leq H_\alpha^*(\text{OPT}_\mathcal{S}) + \varepsilon_\alpha\left(\bigwedge \mathcal{S}\right), \end{aligned} \quad (52)$$

where, for $0 < \alpha < \infty$ with $\alpha \neq 1$,

$$\varepsilon_\alpha\left(\bigwedge \mathcal{S}\right) = \frac{1}{1-\alpha} \log \left(\frac{w_1^\alpha + m^{1-\alpha} \sum_{j=2}^n w_j^\alpha}{\sum_{j=1}^n w_j^\alpha} \right),$$

and

$$\varepsilon_0\left(\bigwedge \mathcal{S}\right) = \log \frac{1 + m(s-1)}{s},$$

$$\varepsilon_1\left(\bigwedge \mathcal{S}\right) = (1 - w_1) \log m,$$

$$\varepsilon_\infty\left(\bigwedge \mathcal{S}\right) = 0.$$

We have that $\varepsilon_\alpha(\bigwedge \mathcal{S})$ is monotonically non-increasing in α and $\varepsilon_\alpha(\bigwedge \mathcal{S}) \leq \log m$, for every order $\alpha \in [0, \infty]$.

The proof appears in Appendix H. Now, we state sufficient conditions and present examples of collections of marginal PMFs for which Algorithm 1 is optimal.

Corollary 3. Let $\mathcal{S} = \{\mathbf{p}^{(i)}\}_{i=1}^m$ be a collection of ordered PMFs, and let $\mathcal{B}_i$ denote the cumulative breakpoint set of $\mathbf{p}^{(i)}$ i.e., $\mathcal{B}_i = \left\{x \in (0, 1) : x = \sum_{j=1}^k p_j^{(i)}, k \in [n-1]\right\}$. Suppose there exists an $r \in [m]$ such that $\mathcal{B}_r = \bigcup_{i=1}^m \mathcal{B}_i$. Then, Algorithm 1 is optimal.

Proof. Let us suppose that $\mathcal{B}_r = \bigcup_{i=1}^m \mathcal{B}_i$, for some $r \in [m]$. Then, the breakpoint set in our Algorithm 1 is exactly $\{0, 1\} \cup \mathcal{B}_r$. Therefore, the consecutive differences between these breakpoints are the positive entries of $\mathbf{p}^{(r)}$. Consequently, the output coupling has sorted PMF vector $\Pi_c^\downarrow(\mathcal{S}) = \mathbf{p}^{(r)}$.

On the other hand, for every feasible coupling Π of the marginal PMFs in the set $\mathcal{S}$, the marginal $\mathbf{p}^{(r)}$ is an aggregation of Π . Therefore, we have $\Pi^\downarrow \preceq \mathbf{p}^{(r)}$. By Schur concavity of the Rényi entropy, we have

$$H_\alpha(\Pi) \geq H_\alpha(\mathbf{p}^{(r)}) = H_\alpha(\Pi_c^\downarrow(\mathcal{S})), \quad \forall \alpha \in [0, \infty].$$

Thus, Algorithm 1 minimizes the joint Rényi entropy for every order $\alpha \in [0, \infty]$. $\square$

The condition in Corollary 3 corresponds to the case where the partition induced by one marginal PMF refines the partitions induced by all the other marginal PMFs. Under this condition, Algorithm 1 does not split any positive mass of the finest marginal and therefore attains the marginal entropy lower bound. However, the above condition does not guarantee the optimality of the greedy coupling algorithm for finite orders of the Rényi entropy. Example 4 in Appendix I explicitly provides an instance in which the condition holds, but the greedy coupling algorithm splits an atom of the finest marginal and produces a coupling with strictly larger entropy than the comonotone coupling in Algorithm 1.

Furthermore, the condition in Corollary 3 is sufficient but not necessary for optimality of Algorithm 1. In Examples 1 and 2 of Appendix I, we show that no single marginal breakpoint set contains all the others (i.e., the condition of Corollary 3 is not satisfied), yet the sorted comonotone coupling vector coincides with the information-spectrum lower-bound distribution Q^* , which proves that Algorithm 1 can be optimal even when the condition in Corollary 3 does not hold. However, we also show via Example 3 (Appendix I) that Algorithm 1 can remain suboptimal for Rényi entropies of every finite positive order $\alpha > 0$.

B. Information-theoretic distance metrics

Let $\mathbf{x}, \mathbf{y} \in \mathcal{P}_n$. Inspired by the work in [53], we define the following distance

$$d_\alpha^R(\mathbf{x}, \mathbf{y}) \triangleq H_\alpha(\mathbf{x}) + H_\alpha(\mathbf{y}) - 2H_\alpha(\mathbf{x} \vee \mathbf{y}) \quad (53)$$

$$d_\alpha^T(\mathbf{x}, \mathbf{y}) \triangleq T_\alpha(\mathbf{x}) + T_\alpha(\mathbf{y}) - 2T_\alpha(\mathbf{x} \vee \mathbf{y}) \quad (54)$$

where $H_\alpha(\cdot)$ and $T_\alpha(\cdot)$ denote the Rényi entropy and Tsallis entropy of order $\alpha \in [0, \infty)$. Using Schur concavity and supermodularity results in previous sections, we can show that $d_\alpha^R(\cdot, \cdot)$ and $d_\alpha^T(\cdot, \cdot)$ are distance metrics on $\mathcal{P}_n$ for the ranges of α stated in the following theorem.

Theorem 8. Let $d_\alpha^R(\mathbf{x}, \mathbf{y})$ and $d_\alpha^T(\mathbf{x}, \mathbf{y})$ be the distances as defined in (53) and (54), respectively. Then, $d_\alpha^R(\mathbf{x}, \mathbf{y})$ is a

metric for every $\alpha \in [1, \infty)$ while $d_\alpha^T(\mathbf{x}, \mathbf{y})$ is a metric for every $\alpha \in (0, \infty)$.

Proof. The proof is deferred to Appendix J. $\square$

For $\alpha \rightarrow 1$, the distances $d_\alpha^R(\mathbf{x}, \mathbf{y})$ and $d_\alpha^T(\mathbf{x}, \mathbf{y})$ reduce to the Shannon-entropy based distance $d(\mathbf{x}, \mathbf{y}) \triangleq H(\mathbf{x}) + H(\mathbf{y}) - 2H(\mathbf{x} \vee \mathbf{y})$, studied in [53]. This metric family also admits a natural interpretation as a measure of wealth inequality in a population. These entropy-based measures are also used to quantify biological diversity [54]. Let $\mathbf{u}_n = (1/n, \dots, 1/n)$ denote the uniform PMF. Since, the lub $\mathbf{x} \vee \mathbf{u}_n = \mathbf{x}$, we have

$$I_\alpha^R(\mathbf{x}) = d_\alpha^R(\mathbf{x}, \mathbf{u}_n) = \log n - H_\alpha(\mathbf{x}), \quad (55)$$

$$I_\alpha^T(\mathbf{x}) = d_\alpha^T(\mathbf{x}, \mathbf{u}_n) = \frac{1 - n^{1-\alpha}}{\alpha - 1} - T_\alpha(\mathbf{x}). \quad (56)$$

Thus, a larger value of $I_\alpha^R(\mathbf{x})$ or $I_\alpha^T(\mathbf{x})$ suggests that the income in a population is more concentrated and therefore indicates higher income inequality in the population. Here $\mathbf{x}$ denotes the normalized wealth of the population. We also observe that I_α^R and I_α^T are transformations of the generalized entropy index [55], whose special case includes the Theil index used in analyses of income inequality [56], [57]. Let $\mathbf{w}$ denote the un-normalized wealth vector and let $\mathbf{x} \in \mathcal{P}_n$ (where $x_i = w_i / \sum_{j=1}^n w_j$) be the corresponding normalized wealth vector, then for finite $\alpha > 0$, we have

$$I_\alpha^R(\mathbf{x}) = \frac{1}{\alpha - 1} \log(1 + \alpha(\alpha - 1) \text{GE}_\alpha(\mathbf{w})) \quad (57)$$

$$I_\alpha^T(\mathbf{x}) = \alpha n^{1-\alpha} \text{GE}_\alpha(\mathbf{w}) \quad (58)$$

where $\text{GE}_\alpha(\mathbf{w})$ is the generalized entropy index of the population. We interpret the above expressions by continuity at $\alpha = 1$, in the limit.

The quantities $I_\alpha^R(\mathbf{x})$ and $I_\alpha^T(\mathbf{x})$ compare a population only to the egalitarian (uniform) benchmark $\mathbf{u}_n$, and therefore, recover generalized-entropy-type inequality index. In contrast, the distance metrics $d_\alpha^R(\mathbf{x}, \mathbf{y})$ and $d_\alpha^T(\mathbf{x}, \mathbf{y})$ compare two arbitrary populations $\mathbf{x}$ and $\mathbf{y}$. Unlike the generalized entropy index, which compares a population only to the egalitarian benchmark, the proposed metrics form a geometry on the space of income distributions, allowing comparison between any two populations directly. This enables clustering of populations, comparison of inequality profiles across countries or time, and detection of distributional differences that may be hidden by a single index such as generalized entropy index, Gini index, or Atkinson's index. A systematic empirical study of these metrics would be a future direction. More broadly, the distance metrics d_α^R and d_α^T can be used to compare arbitrary PMFs, making them useful even beyond econometrics for PMF truncation, guessing, and quantum entanglement.

C. Optimal Transport and Entropic Regularization

Our results also have applications to discrete optimal transport (OT). Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. Further, let $\mathcal{C}(\mathbf{p}, \mathbf{q})$ denote the set of all couplings of $\mathbf{p}$ and $\mathbf{q}$, and let $C \in \mathbb{R}^{n \times n}$ be a cost matrix. The OT problem $\min_{\Pi \in \mathcal{C}(\mathbf{p}, \mathbf{q})} \langle C, \Pi \rangle$ admits an optimal solution at a vertex of the transportation polytope. Since the bipartite support graph of every vertex is a forest [50,

Sec. 3.4.1], Proposition 1 applies to every vertex optimal plan Π^* , including those obtained by the network simplex method. Therefore, for every $\alpha \in [0, \infty)$ we have

$$T_\alpha(\Pi^{\star\downarrow}) + T_\alpha(\mathbf{z}_{\Pi^*}) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}) \quad (59)$$

where $\mathbf{z}_{\Pi^*}$ is the PMF of the Gács–Körner common part under the coupling $(X, Y) \sim \Pi^*$. Since $T_\alpha(\mathbf{z}_{\Pi^*}) \geq 0$, we also have $T_\alpha(\Pi^*) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q})$. For $\alpha \in [0, 1)$, the function (with $t \geq 0$)

$$\varphi_\alpha(t) = \frac{1}{1-\alpha} \log(1 + (1-\alpha)t),$$

is increasing and concave, with $\varphi_\alpha(0) = 0$, and is therefore subadditive. Using $H_\alpha = \varphi_\alpha(T_\alpha)$, we obtain

$$H_\alpha(\Pi^*) \leq H_\alpha(\mathbf{p}) + H_\alpha(\mathbf{q}), \quad (60)$$

for $\alpha \in [0, 1]$, where $\alpha = 1$ follows from the subadditivity of the Shannon entropy. The restriction to vertex optimal plans is relevant here because Rényi entropy is not subadditive over arbitrary couplings for $\alpha \in (0, 1)$. In particular when $C = 0$, every coupling is cost-optimal, so (60) need not hold for every optimal transport plan.

Next, we consider Shannon entropy regularization in optimal transport [58]. For $\varepsilon > 0$, let

$$\Pi_\varepsilon = \arg \min_{\Pi \in \mathcal{C}(\mathbf{p}, \mathbf{q})} \{ \langle C, \Pi \rangle - \varepsilon H(\Pi) \}, \quad (61)$$

and let Π^* now denote any cost-optimal plan for the classical OT. By the optimality of Π_ε and Π^* for their respective problems, we have that

$$0 \leq \langle C, \Pi_\varepsilon \rangle - \langle C, \Pi^* \rangle \leq \varepsilon [H(\Pi_\varepsilon) - H(\Pi^*)]. \quad (62)$$

We already know that every coupling aggregates to $\mathbf{p}$ and $\mathbf{q}$, and hence $\Pi^\downarrow \preceq \mathbf{p} \wedge \mathbf{q}$. Consequently, Schur concavity gives us $H(\Pi^*) \geq H(\mathbf{p} \wedge \mathbf{q})$. Combining this with $H(\Pi_\varepsilon) \leq H(\mathbf{p}) + H(\mathbf{q})$ yields

$$0 \leq \langle C, \Pi_\varepsilon \rangle - \langle C, \Pi^* \rangle \leq \varepsilon [H(\mathbf{p}) + H(\mathbf{q}) - H(\mathbf{p} \wedge \mathbf{q})]. \quad (63)$$

Thus, the majorization lower bound on coupling entropy allows us to upper bound the additional transport cost introduced by regularization, using only the marginal PMFs. Moreover, $H(\mathbf{p} \wedge \mathbf{q}) \geq \max\{H(\mathbf{p}), H(\mathbf{q})\}$, so the upper bound in (63) is no larger than $\varepsilon \min\{H(\mathbf{p}), H(\mathbf{q})\}$. The improvement is strict whenever $\mathbf{p}$ and $\mathbf{q}$ are incomparable i.e., $\mathbf{p} \wedge \mathbf{q} \prec \mathbf{p}$ and $\mathbf{p} \wedge \mathbf{q} \prec \mathbf{q}$, by strict Schur concavity of the Shannon entropy. The upper bound in (63) can be further improved by replacing $H(\mathbf{p} \wedge \mathbf{q})$ with $H(Q^*)$, where Q^* is the distribution defining the tighter information-spectrum based lower bound derived for the minimum entropy coupling in [20].

Finally, our Theorem 1 relates the limiting OT plans of (61) when the comonotone coupling $\Pi_c(\mathbf{p}, \mathbf{q})$ is the unique cost-optimal plan. A sufficient condition is that the cost function C is strictly Monge in the indexing used to construct $\Pi_c(\mathbf{p}, \mathbf{q})$, i.e.,

$$C_{ij} + C_{i'j'} < C_{ij'} + C_{i'j}, \quad i < i', \quad j < j'.$$

Under this condition, the comonotone coupling is an optimal

OT plan [59]. The strictness condition rules out crossing pairs (i.e., cells (i, j') and (i', j) with $i < i'$, $j < j'$, both in the support) in the support of any optimal plan, and the comonotone coupling is the unique coupling without such crossings. Recall that our indexing places masses in the marginal PMFs in non-increasing order. For one-dimensional costs $c(x, y) = |x - y|^r$, with $r > 1$, the optimal plan is the comonotone coupling in the spatial order of the points in the support set. This agrees with our $\Pi_c(\mathbf{p}, \mathbf{q})$ when both marginal PMFs are non-increasing in that spatial order.

When $\Pi_c(\mathbf{p}, \mathbf{q})$ is the unique cost-optimal plan, Π_ε converges to $\Pi_c(\mathbf{p}, \mathbf{q})$ as $\varepsilon \rightarrow 0$, and to the independent coupling $\Pi_l(\mathbf{p}, \mathbf{q})$ as $\varepsilon \rightarrow \infty$ [50, Prop. 4.1]. Our Theorem 1 orders these limiting OT plans through majorization, i.e., $\Pi_l^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \Pi_c^\downarrow(\mathbf{p}, \mathbf{q})$. Therefore, we have

$$H_\alpha(\Pi_c(\mathbf{p}, \mathbf{q})) \leq H_\alpha(\Pi_l(\mathbf{p}, \mathbf{q})) \quad (64)$$

and

$$T_\alpha(\Pi_c(\mathbf{p}, \mathbf{q})) \leq T_\alpha(\Pi_l(\mathbf{p}, \mathbf{q})), \quad (65)$$

for $\alpha \in [0, \infty]$ and $\alpha \in [0, \infty)$, respectively. For Shannon entropy, the difference between these endpoint entropies is the mutual information under the comonotone coupling. If the cost-optimal plan is not unique, the limit as $\varepsilon \rightarrow 0$ is instead the optimal transport plan with the maximum Shannon entropy [50, Prop. 4.1].

VII. CONCLUSION

In this paper, we studied entropic inequalities on the majorization lattice using couplings and the geometry of Lorenz curves. We established a majorization relation between the comonotone and independent couplings of ordered marginal PMFs and proved a strengthened Tsallis entropy inequality involving the Gács–Körner common part for the couplings which induce a forest-supported bipartite graph. Using these results, we showed that both Rényi and Tsallis entropies are subadditive on the majorization lattice. We also characterized their supermodularity properties, showing that the two families behave differently for orders $\alpha \in (0, 1)$.

Further, we also considered applications of our results to minimum Rényi entropy couplings, distance metrics, and optimal transport. For the coupling problem with m marginal PMFs, we provided a polynomial-time construction of the comonotone coupling whose Rényi entropy is within $\log m$ nats of the optimum for every finite order. Further, we obtained a sharper distribution-dependent bound and established optimality for min-entropy. Our supermodularity results helped us to establish two families of distance metrics on ordered PMFs, with connections to the generalized entropy index used to measure income inequality in econometrics. In optimal transport, we showed that every vertex optimal transport plan satisfies the strengthened Tsallis inequality, and we upper bounded the additional transport cost under the Shannon entropy regularization in terms of the marginal PMFs.

A direction for the future work is to study how the order of the Rényi as well as Tsallis entropy affects the behavior of the proposed distance metrics and their usefulness in comparing income distributions and diversity of populations. It would

also be interesting to identify conditions under which the strengthened Tsallis inequality holds for couplings whose support graphs contain cycles for $\alpha \in (0, 1)$, and to investigate whether the coupling inequalities developed here yield useful guarantees for OT under different regularizers.

ACKNOWLEDGMENT

This work is supported by the Swiss National Science Foundation (SNSF) grant number 211337.

REFERENCES

- [1] A. K. Yadav and Y. Y. Shkel, “Geometry of rényi entropy on the majorization lattice,” 2026. [Online]. Available: <https://arxiv.org/abs/2605.09655v1>
- [2] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, ser. Springer Series in Statistics. Springer New York, 2010. [Online]. Available: <https://books.google.ch/books?id=jE11p6zLJ0C>
- [3] I. Schur, “Über eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie,” *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, vol. 22, pp. 9–20, 1923.
- [4] G. H. Hardy, “Some simple inequalities satisfied by convex functions,” *Messenger Math.*, vol. 58, pp. 145–152, 1929.
- [5] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*, ser. Cambridge Mathematical Library. Cambridge University Press, 1952. [Online]. Available: <https://books.google.ch/books?id=t1RCSP8YKt8C>
- [6] M. O. Lorenz, “Methods of measuring the concentration of wealth,” *Publications of the American Statistical Association*, vol. 9, no. 70, pp. 209–219, 1905. [Online]. Available: <http://www.jstor.org/stable/2276207>
- [7] A. Pigou, *Wealth and Welfare*. Creative Media Partners, LLC, 2017. [Online]. Available: https://books.google.co.in/books?id=1qM_swECAAJ
- [8] H. Dalton, “The measurement of the inequality of incomes,” *The Economic Journal*, vol. 30, no. 119, pp. 348–361, 1920. [Online]. Available: <http://www.jstor.org/stable/2223525>
- [9] T. van Erven and P. Harremoës, “Rényi divergence and majorization,” in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 1335–1339.
- [10] M. U. Farooq, T. Fritz, E. Haapasalo, and M. Tomamichel, “Matrix majorization in large samples,” *IEEE Transactions on Information Theory*, vol. 70, no. 5, pp. 3118–3144, 2024.
- [11] D. Elkouss, A. G. Maity, A. Nema, and S. Strelchuk, “A finite sufficient set of conditions for catalytic majorization,” *Communications Physics*, 2026. [Online]. Available: <https://doi.org/10.1038/s42005-026-02583-x>
- [12] D. P. Palomar and Y. Jiang, “Mimo transceiver design via majorization theory,” *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4–5, pp. 331–551, 06 2007. [Online]. Available: <https://doi.org/10.1561/01000000018>
- [13] M. A. Nielsen, “Conditions for a class of entanglement transformations,” *Phys. Rev. Lett.*, vol. 83, pp. 436–439, Jul 1999. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.83.436>
- [14] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, “Quantum majorization and a complete set of entropic conditions for quantum thermodynamics,” *Nature Communications*, vol. 9, no. 1, p. 5352, 2018. [Online]. Available: <https://doi.org/10.1038/s41467-018-06261-7>
- [15] V. Jorda, J. M. Sarabia, and M. Jäntti, “Inequality measurement with grouped data: Parametric and non-parametric methods,” *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 184, no. 3, pp. 964–984, 07 2021. [Online]. Available: <https://doi.org/10.1111/rssa.12702>
- [16] Y. Fan, M. Henry, B. Pass, and J. A. Rivero, “Multidimensional inequality measurement via optimal transport,” *The Review of Economics and Statistics*, pp. 1–45, 11 2024. [Online]. Available: https://doi.org/10.1162/rest_a_01539
- [17] S. I. Butt, I. Javed, P. Agarwal, and J. J. Nieto, “Newton–simpson-type inequalities via majorization,” *Journal of Inequalities and Applications*, vol. 2023, no. 1, p. 16, 2023. [Online]. Available: <https://doi.org/10.1186/s13660-023-02918-0>
- [18] C. T. Li, “Efficient approximate minimum entropy coupling of multiple probability distributions,” *IEEE Transactions on Information Theory*, vol. 67, no. 8, pp. 5259–5268, 2021.
- [19] A. Painsky, S. Rosset, and M. Feder, “Innovation representation of stochastic processes with application to causal inference,” *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1136–1154, 2020.
- [20] Y. Y. Shkel and A. Kumar Yadav, “Information spectrum converse for minimum entropy couplings and functional representations,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 66–71.
- [21] A. K. Yadav and Y. Y. Shkel, “Approximation guarantees for minimum rényi entropy functional representations,” in *2025 IEEE International Symposium on Information Theory (ISIT)*, 2025, pp. 1–6.
- [22] M. Kocaoglu, A. Dimakis, S. Vishwanath, and B. Hassibi, “Entropic causal inference,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/10674>
- [23] S. Compton, D. Katz, B. Qi, K. Greenewald, and M. Kocaoglu, “Minimum-entropy coupling approximation guarantees beyond the majorization barrier,” in *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, F. Ruiz, J. Dy, and J.-W. van de Meent, Eds., vol. 206. PMLR, 25–27 Apr 2023, pp. 10 445–10 469. [Online]. Available: <https://proceedings.mlr.press/v206/compton23a.html>
- [24] A. Rényi, “On measures of entropy and information,” *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (June 20-July 30 1960)*, pp. 547–561, Jun 1960. [Online]. Available: <http://digicoll.lib.berkeley.edu/record/112906>
- [25] L. L. Campbell, “A coding theorem and rényi’s entropy,” *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965.
- [26] C. Pfister and W. Sullivan, “Rényi entropy, guesswork moments, and large deviations,” *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2794–2800, 2004.
- [27] R. Renner and S. Wolf, “Simple and tight bounds for information reconciliation and privacy amplification,” in *Advances in Cryptology - ASIACRYPT 2005*, B. Roy, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 199–216.
- [28] M. Mosonyi and T. Ogawa, “Quantum hypothesis testing and the operational interpretation of the quantum rényi relative entropies,” *Communications in Mathematical Physics*, vol. 334, no. 3, pp. 1617–1648, 2015. [Online]. Available: <https://doi.org/10.1007/s00220-014-2248-x>
- [29] C. Tsallis, “Possible generalization of boltzmann-gibbs statistics,” *Journal of Statistical Physics*, vol. 52, no. 1, pp. 479–487, 1988. [Online]. Available: <https://doi.org/10.1007/BF01016429>
- [30] C. Tsallis, A. Plastino, and W.-M. Zheng, “Power-law sensitivity to initial conditions—new entropic representation,” *Chaos, Solitons & Fractals*, vol. 8, no. 6, pp. 885–891, 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960077996001671>
- [31] C. Tsallis, C. Anteneodo, L. Borland, and R. Osorio, “Nonextensive statistical mechanics and economics,” *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1, pp. 89–100, 2003, proceedings of the International Econophysics Conference. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378437103000426>
- [32] M. Portes de Albuquerque, I. Esquef, A. Gesualdi Mello, and M. Portes de Albuquerque, “Image thresholding using tsallis entropy,” *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1059–1065, 2004. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865504000698>
- [33] A. Jawad, K. Bamba, M. Younas, S. Qummer, and S. Rani, “Tsallis, rényi and sharma-mittal holographic dark energy models in loop quantum cosmology,” *Symmetry*, vol. 10, no. 11, 2018. [Online]. Available: <https://www.mdpi.com/2073-8994/10/11/635>
- [34] A. Squicciarini, T. Trigano, and D. Luengo, “Jensen–tsallis divergence for supervised classification under data imbalance,” *Machine Learning*, vol. 114, no. 7, p. 162, 2025. [Online]. Available: <https://doi.org/10.1007/s10994-025-06791-4>
- [35] R. Bapat, “Majorization and singular values. iii,” *Linear Algebra and its Applications*, vol. 145, pp. 59–70, 1991. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0024379591902877>
- [36] L. Lovász, “Submodular functions and convexity,” pp. 235–257, March 1983. [Online]. Available: https://ideas.repec.org/h/spr/sprchp/978-3-642-68874-4_10.html
- [37] S. Fujishige, *Submodular Functions and Optimization*, ser. Annals of Discrete Mathematics. North Holland, 1991. [Online]. Available: <https://books.google.ch/books?id=a5PEEQPLpHYC>
- [38] —, “Polymatroidal dependence structure of a set of random variables,” *Information and Control*, vol. 39, no. 1, pp. 55–72, 1978. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S00199587891063X>

- [39] L. Csirmaz, "The size of a share must be large," *Journal of Cryptology*, vol. 10, no. 4, pp. 223–231, 1997. [Online]. Available: <https://doi.org/10.1007/s001459900029>
- [40] F. Cicalese and U. Vaccaro, "Supermodularity and subadditivity properties of the entropy on the majorization lattice," *IEEE Transactions on Information Theory*, vol. 48, no. 4, pp. 933–938, 2002.
- [41] D. S. Parker and P. Ram, "The construction of huffman codes is a submodular ("convex") optimization problem over a lattice of binary trees," *SIAM Journal on Computing*, vol. 28, no. 5, pp. 1875–1905, 1999. [Online]. Available: <https://doi.org/10.1137/S0097539796311077>
- [42] C. Massri, G. Bellomo, F. Holik, and G. M. Bosyk, "Extremal elements of a sublattice of the majorization lattice and approximate majorization," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 21, p. 215305, may 2020. [Online]. Available: <https://doi.org/10.1088/1751-8121/ab8674>
- [43] Q. Guo, Y. Hu, and R. Zhang, "The geometry of coalition power: Majorization, lattices, and displacement in multiwinner elections," *arXiv preprint arXiv:2601.16723*, 2026.
- [44] R. Bruno and U. Vaccaro, "The sharma-mittal entropy is subadditive and supermodular on the majorization lattice," *arXiv preprint arXiv:2605.18600*, 2026.
- [45] B. D. Sharma and D. P. Mittal, "New non-additive measures of entropy for discrete probability distributions," *J. Math. Sci.*, vol. 10, pp. 28–40, 1975.
- [46] A. Stévin, M. G. Jabbour, S. Deside, and N. J. Cerf, "Majorization precursors to supermodularity and subadditivity on the majorization lattice," *arXiv preprint arXiv:2605.30331*, 2026.
- [47] T. Hariharan, P. Krishnan, and A. Deshpande, "On measures of dissimilarity on the majorization simplex," 2026. [Online]. Available: <https://arxiv.org/abs/2610.16687>
- [48] P. Bhatia, S. Singh, and V. Kumar, "On some properties of tsallis entropy on majorization lattice," *Acta Mathematica Academiae Paedagogicae Nyregyháziensis*, vol. 31, no. 2, pp. 331–340, 2015.
- [49] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, pp. 149–162, 1973.
- [50] G. Peyré and M. Cuturi, "Computational optimal transport with applications to data sciences," *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 02 2019.
- [51] 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.
- [52] M. Kovačević, I. Stanojević, and V. Šenk, "On the entropy of couplings," *Information and Computation*, vol. 242, pp. 369–382, 2015.
- [53] F. Cicalese, L. Gargano, and U. Vaccaro, "Information theoretic measures of distances and their econometric applications," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 409–413.
- [54] E. C. Pielou, "The measurement of diversity in different types of biological collections," *Journal of Theoretical Biology*, vol. 13, pp. 131–144, 1966. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0022519366900130>
- [55] A. F. Shorrocks, "The class of additively decomposable inequality measures," *Econometrica*, vol. 48, no. 3, pp. 613–625, 1980. [Online]. Available: <http://www.jstor.org/stable/1913126>
- [56] C. N. Haddad, D. G. Mahler, C. Diaz-Bonilla, R. Hill, C. Lakner, and G. Lara Ibarra, "The world bank's new inequality indicator: The number of countries with high inequality," Washington, DC, Jun. 2024. [Online]. Available: <https://hdl.handle.net/10986/41687>
- [57] United Nations Development Programme, *Human Development Report 2019: Beyond Income, Beyond Averages, Beyond Today: Inequalities in Human Development in the 21st Century*. New York: United Nations Development Programme, 2019. [Online]. Available: <https://hdr.undp.org/system/files/documents/hdr2019.pdf>
- [58] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf
- [59] A. J. Hoffman, "On simple linear programming problems," in *Convexity*, ser. Proceedings of Symposia in Pure Mathematics, V. L. Klee, Ed. Providence, RI: American Mathematical Society, 1963, vol. 7, pp. 317–327.
- [60] S. Salamatian, A. Cohen, and M. Médard, "Efficient coding for multi-source networks using gács-körner common information," in *2016 International Symposium on Information Theory and Its Applications (ISITA)*, 2016, pp. 166–170.

APPENDIX

A. Proof of Lemma 1

Proof. Given $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. Let U be a uniformly distributed continuous random variable on $[0, 1]$ i.e., $U \sim \text{Unif}([0, 1])$. Now, define random variables X and Y on the support set $\{1, 2, \dots, n\}$ as

$$X \triangleq \min \left\{ i : \sum_{j=1}^i p_j \geq U \right\}$$

$$Y \triangleq \min \left\{ i : \sum_{j=1}^i q_j \geq U \right\}$$

Thus, we have that $X \sim \mathbf{p}$, $Y \sim \mathbf{q}$ and they are coupled through the random variable U . Thus, the joint distribution of (X, Y) is the comonotone coupling of $\mathbf{p}$ and $\mathbf{q}$.

Now, let us define $Z \triangleq \max(X, Y)$ and $W \triangleq \min(X, Y)$. We will show that $Z \sim \mathbf{p} \wedge \mathbf{q}$ and explain how the PMF of W determines $\mathbf{p} \vee \mathbf{q}$. We have,

$$\mathbb{P}(Z \leq i) = \mathbb{P}(X \leq i, Y \leq i)$$

$$= \mathbb{P} \left(U \leq \sum_{j=1}^i p_j, U \leq \sum_{j=1}^i q_j \right) \quad (66)$$

$$= \mathbb{P} \left(U \leq \min \left\{ \sum_{j=1}^i p_j, \sum_{j=1}^i q_j \right\} \right) \quad (67)$$

$$= \min \left\{ \sum_{j=1}^i p_j, \sum_{j=1}^i q_j \right\} \quad (68)$$

Therefore,

$$\mathbb{P}(Z = i) = \min \left\{ \sum_{j=1}^i p_j, \sum_{j=1}^i q_j \right\}$$

$$- \min \left\{ \sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i-1} q_j \right\} \quad (69)$$

From (8), we have that $Z \sim \mathbf{p} \wedge \mathbf{q}$. Consequently, for all $k \in [n]$,

$$(\mathbf{p} \wedge \mathbf{q})(k) = \sum_{(i,j): \max(i,j)=k} \Pi_c(\mathbf{p}, \mathbf{q})(i, j) \quad (70)$$

Similarly, we have that

$$\mathbb{P}(W \leq i) = \mathbb{P}(X \leq i \text{ or } Y \leq i)$$

$$= \mathbb{P} \left(U \leq \sum_{j=1}^i p_j \text{ or } U \leq \sum_{j=1}^i q_j \right) \quad (71)$$

$$= \mathbb{P} \left(U \leq \max \left\{ \sum_{j=1}^i p_j, \sum_{j=1}^i q_j \right\} \right) \quad (72)$$

$$= \max \left\{ \sum_{j=1}^i p_j, \sum_{j=1}^i q_j \right\} \quad (73)$$

Therefore,

$$\mathbb{P}(W = i) = \max \left\{ \sum_{j=1}^i p_j, \sum_{j=1}^i q_j \right\}$$

$$- \max \left\{ \sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i-1} q_j \right\} \quad (74)$$

From (9), we have that $W \sim \beta_{\mathbf{p}, \mathbf{q}}$. Consequently, for all $k \in [n]$,

$$\beta_{\mathbf{p}, \mathbf{q}}(k) = \sum_{(i,j): \min(i,j)=k} \Pi_c(\mathbf{p}, \mathbf{q})(i, j) \quad (75)$$

The lub $\mathbf{p} \vee \mathbf{q}$ is obtained by taking successive differences of the LCM of the cumulative sum curve of $\beta_{\mathbf{p}, \mathbf{q}}$. $\square$

B. Proof of Theorem 1

Proof. Given $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. Let $\Pi_l(\mathbf{p}, \mathbf{q})$ and $\Pi_c(\mathbf{p}, \mathbf{q})$ denote the independent coupling and the comonotone coupling, respectively. Further, we use $\Pi_l^\downarrow(\mathbf{p}, \mathbf{q})$ and $\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})$ to denote the independent and the comonotone couplings as vectors (PMFs) arranged in the non-increasing order.

Fix a $k \in [n^2]$. Consider the top- k largest masses of the independent coupling $\Pi_l(\mathbf{p}, \mathbf{q})$. Since $\mathbf{p}$ and $\mathbf{q}$ are sorted in the non-increasing order, for $i' \leq i, j' \leq j$ we have that $p_{i'} q_{j'} \geq p_i q_j$. Since $\mathbf{p}$ and $\mathbf{q}$ are arranged in non-increasing order, the k largest masses can be selected so that their index set D is a Young's diagram, breaking ties accordingly. A Young's diagram is a left-justified arrangement of points in rows whose lengths form a non-increasing sequence.

Let $\mathcal{D}$ denote the optimal set of top- k masses i.e., $|\mathcal{D}| = k$, which is a Young's diagram. Now, let's say $\mathcal{D}$ uses 'a' rows and 'b' columns of the independent coupling matrix, then it completely lies inside the rectangular block $[a] \times [b]$. Thus,

$$\sum_{i=1}^k \Pi_l^\downarrow(\mathbf{p}, \mathbf{q})(i) = \sum_{(i,j) \in \mathcal{D}} p_i q_j$$

$$\leq \sum_{(i,j) \in [a] \times [b]} p_i q_j \quad (76)$$

$$= \left(\sum_{i \leq a} p_i \right) \left(\sum_{j \leq b} q_j \right) \quad (77)$$

$$\leq \min \left\{ \sum_{i \leq a} p_i, \sum_{j \leq b} q_j \right\} \quad (78)$$

Now, we consider the same rectangular block $[a] \times [b]$ within the comonotone coupling matrix and show that it contains exactly the mass in the RHS of (78).

Recall from the proof of Lemma 1 that the comonotone coupling is the coupling between the random variables (X, Y) where $X \sim \mathbf{p}$ and $Y \sim \mathbf{q}$, generated in the following manner

$$U \sim \text{Unif}([0, 1])$$

$$X \triangleq \min \left\{ i : \sum_{j=1}^i p_j \geq U \right\}$$

$$Y \triangleq \min \left\{ i : \sum_{j=1}^i q_j \geq U \right\}$$

Therefore, we have that

$$\begin{aligned} \Pi_c(\mathbf{p}, \mathbf{q})([a] \times [b]) &= \mathbb{P}(X \leq a, Y \leq b) \\ &= \mathbb{P} \left(U \leq \sum_{j=1}^a p_j, U \leq \sum_{j=1}^b q_j \right) \end{aligned} \quad (79)$$

$$= \min \left\{ \sum_{i \leq a} p_i, \sum_{j \leq b} q_j \right\} \quad (80)$$

To complete our proof, we will now show that the rectangular block $[a] \times [b]$ in the comonotone coupling has at most k non-zero elements.

First, note that the comonotone coupling is supported on a monotone path through the matrix i.e., any north-east element or the south-west element around a non-zero element is zero. Thus, it starts from the north-west corner (top-left) of the matrix as the first nonzero cell, the next nonzero cell is obtained by moving $(i, j) \rightarrow (i+1, j)$ or $(i, j) \rightarrow (i, j+1)$. But if a row and column are exhausted simultaneously, the next nonzero can be $(i, j) \rightarrow (i+1, j+1)$. Consequently, the rectangular block $[a] \times [b]$ aligned in the north-west corner has at most $a+b-1$ non-zero elements.

Recall the Young's diagram formed by the top- k elements of the independent coupling matrix, which is tightly enclosed by the rectangular block $[a] \times [b]$. Therefore, $\mathcal{D}$ entirely contains the first row of length b and the first column of length a , which implies that $k = |\mathcal{D}| \geq a+b-1$. As a result, using (78) and (80) we have

$$\sum_{i=1}^k \Pi_i^\downarrow(\mathbf{p}, \mathbf{q})(i) \leq \min \left\{ \sum_{i \leq a} p_i, \sum_{j \leq b} q_j \right\} \quad (81)$$

$$\leq \sum_{i=1}^{a+b-1} \Pi_c^\downarrow(\mathbf{p}, \mathbf{q})(i) \quad (82)$$

$$\leq \sum_{i=1}^k \Pi_c^\downarrow(\mathbf{p}, \mathbf{q})(i) \quad (83)$$

Therefore, the partial sum of top- k masses of the comonotone coupling is at least as large as the partial sums of top- k masses of the independent coupling. Since, the result holds for every $k \in [n^2]$, we conclude

$$\Pi_i^\downarrow(\mathbf{p}, \mathbf{q}) \preceq \Pi_c^\downarrow(\mathbf{p}, \mathbf{q}). \quad (84)$$

This completes our proof. $\square$

C. Proof of Lemma 2

Proof. Consider a PMF $\mathbf{p} \in \mathcal{P}_n$. Then, the Tsallis entropy of $\mathbf{p}$ can be written as

$$T_\alpha(\mathbf{p}) = \sum_{i=1}^n \frac{p_i - p_i^\alpha}{\alpha - 1} = \sum_{i=1}^n h_\alpha(p_i) \quad (85)$$

where $h_\alpha(x) := \frac{x - x^\alpha}{\alpha - 1}$ and $h_1(x) := -x \log x$.

Let $B_{\mathbf{p}}$ and $B_{\mathbf{q}}$ denote the sets of cumulative (partial)-sum breakpoints of the PMFs $\mathbf{p}$ and $\mathbf{q}$, respectively. Then, the common refinement of both the interval partitions is generated by the set $B_{\mathbf{p}} \cup B_{\mathbf{q}}$, and its interval lengths are the probability masses of the comonotone coupling $\Pi_c(\mathbf{p}, \mathbf{q})$.

Let $B = \{b_0, \dots, b_m\}$ be a finite set such that $B \subseteq [0, 1]$, and $0 = b_0 < b_1 < \dots < b_m = 1$ be the ordered elements of B . Now, define $g(B) := (b_1 - b_0, b_2 - b_1, \dots, b_m - b_{m-1})$. Thus, $g(B)$ is a vector (PMF) of interval lengths induced by the breakpoints in B . Now, let us define the function $F_\alpha(B) := T_\alpha(g(B))$. Then, we note that

$$F_\alpha(B_{\mathbf{p}}) = T_\alpha(\mathbf{p}), \quad F_\alpha(B_{\mathbf{q}}) = T_\alpha(\mathbf{q}), \quad (86)$$

and from the definition of the comonotone coupling (cf. Section II), we have

$$F_\alpha(B_{\mathbf{p}} \cup B_{\mathbf{q}}) = T_\alpha(\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})). \quad (87)$$

Therefore, we will now show that

$$F_\alpha(B_{\mathbf{p}} \cup B_{\mathbf{q}}) + F_\alpha(B_{\mathbf{p}} \cap B_{\mathbf{q}}) \leq F_\alpha(B_{\mathbf{p}}) + F_\alpha(B_{\mathbf{q}}). \quad (88)$$

Assume that $\alpha \in (0, \infty)$. If an interval of length $a+b$ is split into two intervals of lengths a and b , then the increase in $F_\alpha(\cdot)$ is

$$\Delta_\alpha(a, b) = h_\alpha(a) + h_\alpha(b) - h_\alpha(a+b) \quad (89)$$

$$= \frac{(a+b)^\alpha - a^\alpha - b^\alpha}{\alpha - 1} \quad (90)$$

For $a, b > 0$, we have that

$$\frac{\partial}{\partial a} \Delta_\alpha(a, b) = \frac{\alpha((a+b)^{\alpha-1} - a^{\alpha-1})}{\alpha - 1} > 0, \quad (91)$$

and similarly $\frac{\partial}{\partial b} \Delta_\alpha(a, b) > 0$. Consequently, $\Delta_\alpha(a, b)$ is increasing in both (a, b) for every $\alpha \in (0, \infty)$, including the limiting case of $\alpha \rightarrow 1$. For $\alpha = 0$, we have $h_0(0) = 0$ and $h_0(x) = 1 - x$ for $x > 0$. Thus, $\Delta_0(a, b) = 1$ whenever $a, b > 0$. Now, let the cumulative-sum breakpoint set of $\mathbf{q}$ be $B_{\mathbf{q}} = \{b_0^{\mathbf{q}} = 0, b_1^{\mathbf{q}}, \dots, b_m^{\mathbf{q}} = 1\}$. Let $B_q \setminus B_p = \{x_1, \dots, x_r\}$, where $r = |B_q \setminus B_p|$. Set $A_0 = B_p \cap B_q$ and define $A_t = A_{t-1} \cup \{x_t\}$ for $t = 1, \dots, r$. Furthermore, let $C_t = A_t \cup B_p$ for $t = 0, \dots, r$. Then $A_r = B_q$, $C_0 = B_p$, and $C_r = B_p \cup B_q$. Now, when a breakpoint x_t is added to A_{t-1} , it splits some existing interval (of length $a_t + b_t$) into two intervals of lengths a_t and b_t . When the same breakpoint x_t is added to C_{t-1} , we note that the interval being split is contained inside the former interval. Thus, it is split into two intervals of lengths, say a'_t and b'_t , where

$$0 < a'_t \leq a_t, \quad 0 < b'_t \leq b_t. \quad (92)$$

Since Δ_α is non-decreasing in both the arguments for every $\alpha \in [0, \infty)$, we get

$$F_\alpha(C_t) - F_\alpha(C_{t-1}) \leq F_\alpha(A_t) - F_\alpha(A_{t-1}). \quad (93)$$

Summing over all iterations, $t \in [r]$, we obtain $F_\alpha(B_{\mathbf{p}} \cup B_{\mathbf{q}}) - F_\alpha(B_{\mathbf{p}}) \leq F_\alpha(B_{\mathbf{q}}) - F_\alpha(B_{\mathbf{p}} \cap B_{\mathbf{q}})$. Therefore,

$$F_\alpha(B_{\mathbf{p}} \cup B_{\mathbf{q}}) + F_\alpha(B_{\mathbf{p}} \cap B_{\mathbf{q}}) \leq F_\alpha(B_{\mathbf{p}}) + F_\alpha(B_{\mathbf{q}}). \quad (94)$$

Note that $g(B_{\mathbf{p}} \cup B_{\mathbf{q}})$ consists of the positive elements of

the comonotone coupling vector $\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})$, up to permutation. Under the quantile construction of $\Pi_c(\mathbf{p}, \mathbf{q})$, the points in $(B_{\mathbf{p}} \cap B_{\mathbf{q}})$ are exactly the cumulative levels at which both X and Y change levels simultaneously. Therefore, these points define the maximal random variable that can be recovered from both X and Y [49], [60]. Consequently, $g(B_{\mathbf{p}} \cap B_{\mathbf{q}})$ is its PMF which achieves the Gács-Körner common information. Equivalently, we have

$$T_\alpha(\Pi_c^\downarrow(\mathbf{p}, \mathbf{q})) + T_\alpha(\mathbf{z}(\mathbf{p}, \mathbf{q})) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}). \quad (95)$$

where $\mathbf{z}(\mathbf{p}, \mathbf{q}) = g(B_{\mathbf{p}} \cap B_{\mathbf{q}})$ and equality holds for $\alpha = 0$.

(Equality cases). First, note that if $\alpha = 0$, then for every $\mathbf{p}$ and $\mathbf{q}$ we have that (93) and (94) are satisfied with equality. Consequently, the equality holds in (95).

For any $\alpha \in [0, \infty)$, if either $\mathbf{p}$ or $\mathbf{q}$ is deterministic, say $\mathbf{p} = (1, 0, \dots, 0)$, then we have that $B_{\mathbf{p}} = \{0, 1\}$, and $B_{\mathbf{p}} \subseteq B_{\mathbf{q}}$. Consequently, $\Pi_c^\downarrow(\mathbf{p}, \mathbf{q}) = \mathbf{q}$ and $\mathbf{z}(\mathbf{p}, \mathbf{q}) = (1)$. Thus, the equality holds in (95). This completes our proof. $\square$

D. Proof of Proposition 1

Proof. Let G be the bipartite support graph of the forest-supported coupling Π , after removing any zero mass atoms. Let us assign weight Π_{ij} to the edge joining the atoms x_i and y_j , of $X \sim \mathbf{p}$ and $Y \sim \mathbf{q}$ respectively. The connected component label of G is the Gács-Körner common part under the coupling Π [60]. Thus, the total edge masses of its components are exactly the entries of $\mathbf{z}_\Pi$.

Fix $\alpha \in (0, \infty)$, and let h_α and Δ_α be as we defined previously in the proof of Lemma 2. Recall that

$$\Delta_\alpha(a, b) = h_\alpha(a) + h_\alpha(b) - h_\alpha(a + b)$$

is non-decreasing in each argument. Consider a tree component T of the bipartite graph G . For each edge e in T , let w_e denote its weight, and for each vertex v , let d_v denote the sum of the weights of its incident edges. Let $M = \sum_{e \in E(T)} w_e$, and define

$$D_\alpha(T) := \sum_{v \in V(T)} h_\alpha(d_v) - \sum_{e \in E(T)} h_\alpha(w_e) - h_\alpha(M),$$

By applying induction on the number of edges, we will show that $D_\alpha(T) \geq 0$. If T consists of a single edge, both incident vertex masses equal the edge weight M , and therefore we have $D_\alpha(T) = 0$. Otherwise, choose a leaf of the tree T , namely a vertex incident to exactly one edge. Remove this vertex and its incident edge, whose weight we denote by a . Let b be the total incident mass at the neighboring vertex before removal.

The remaining graph is a nonempty tree T' , with total edge mass $M - a$. The neighboring vertex has incident mass $b - a$, while all other remaining vertex masses are unchanged. Consequently, we have

$$\begin{aligned} D_\alpha(T) - D_\alpha(T') &= h_\alpha(b) - h_\alpha(b - a) \\ &\quad - h_\alpha(M) + h_\alpha(M - a) \\ &= \Delta_\alpha(a, M - a) - \Delta_\alpha(a, b - a) \\ &\geq 0, \end{aligned}$$

where the last inequality follows from the fact that $b \leq M$ and the monotonicity of Δ_α . By the induction hypothesis, we have $D_\alpha(T') \geq 0$, which therefore implies $D_\alpha(T) \geq 0$.

Summing this inequality over all tree components of the bipartite graph G , the vertex masses give the positive entries of $\mathbf{p}$ and $\mathbf{q}$, the edge weights give the positive entries of Π , and the component masses give $\mathbf{z}_\Pi$. Thus,

$$T_\alpha(\Pi^\downarrow) + T_\alpha(\mathbf{z}_\Pi) \leq T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}).$$

(Equality case). Consider $\alpha = 0$. Let N_E , N_V , and c denote the numbers of edges, vertices, and connected components of G , respectively. Since G is a forest, we have $N_E + c = N_V$. Using $T_0(\mathbf{r}) = |\text{supp}(\mathbf{r})| - 1$, we obtain

$$\begin{aligned} T_0(\Pi^\downarrow) + T_0(\mathbf{z}_\Pi) &= (N_E - 1) + (c - 1) \\ &= N_V - 2 \\ &= T_0(\mathbf{p}) + T_0(\mathbf{q}). \end{aligned}$$

This proves our desired inequality for every $\alpha \in [0, \infty)$, with equality at $\alpha = 0$. $\square$

E. Proof of Theorem 5

Proof. We prove our result in three steps. First, we provide a fixed pair of PMFs (Example E.1) for which submodularity fails for every $\alpha \in (0, 1)$. Then, we give a perturbation-based construction showing that, for every fixed $\alpha \in (0, 1)$, there exists a pair of PMFs for which supermodularity fails. Both constructions use PMFs of support size three. Finally, we provide a fixed four-coordinate PMF pair (Example E.2) and numerical values illustrating the failure of supermodularity for $\alpha \in (0, 1)$.

Given $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, let us define,

$$\Delta_\alpha(\mathbf{p}, \mathbf{q}) \triangleq H_\alpha(\mathbf{p} \wedge \mathbf{q}) + H_\alpha(\mathbf{p} \vee \mathbf{q}) - H_\alpha(\mathbf{p}) - H_\alpha(\mathbf{q}).$$

Now, given an order α , we note that supermodularity is equivalent to $\Delta_\alpha(\mathbf{p}, \mathbf{q}) \geq 0$ for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, while submodularity is equivalent to $\Delta_\alpha(\mathbf{p}, \mathbf{q}) \leq 0$ for all $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$.

Example E.1 – Pair with $\Delta_\alpha(\mathbf{p}, \mathbf{q}) > 0$. Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_3$ be

$$\mathbf{p} = (0.60, 0.20, 0.20),$$

$$\mathbf{q} = (0.45, 0.40, 0.15).$$

Consequently, the glb $(\mathbf{p} \wedge \mathbf{q})$ and lub $(\mathbf{p} \vee \mathbf{q})$ of $\mathbf{p}$ and $\mathbf{q}$ are

$$\mathbf{w} \triangleq \mathbf{p} \wedge \mathbf{q} = (0.45, 0.35, 0.20),$$

$$\mathbf{z} \triangleq \mathbf{p} \vee \mathbf{q} = (0.60, 0.25, 0.15).$$

For illustration, the corresponding numerical values of $\Delta_\alpha(\mathbf{p}, \mathbf{q})$, computed using natural logarithms are

α	$\Delta_\alpha(p, q)$
0	0
0.2	0.00580417
0.5	0.01387056
0.7	0.01882329
0.9	0.02343356
1	0.02560746
2	0.04204643
∞	0

(96)

Now, we prove that this fixed pair $(\mathbf{p}, \mathbf{q})$ satisfies $\Delta_\alpha(\mathbf{p}, \mathbf{q}) > 0$ for every $\alpha \in (0, 1)$. Fix an $\alpha \in (0, 1)$ and let us define

$$\Phi_\alpha(u, v) \triangleq H_\alpha(u, v - u, 1 - v)$$

for $0 < u < v < 1$. Further, let us write $a = u$, $b = v - u$, $c = 1 - v$, and $s = a^\alpha + b^\alpha + c^\alpha$. Then

$$\Phi_\alpha(u, v) = \frac{1}{1 - \alpha} \log s.$$

Consider the rectangular area $\mathcal{R} = [0.45, 0.60] \times [0.80, 0.85]$. Throughout this rectangle, we have that $a \geq b \geq c > 0$, while its four corners correspond to the PMFs $\mathbf{w}, \mathbf{z}, \mathbf{p}, \mathbf{q}$, giving

$$\begin{aligned} \Delta_\alpha(\mathbf{p}, \mathbf{q}) &= \Phi_\alpha(0.45, 0.80) + \Phi_\alpha(0.60, 0.85) \\ &\quad - \Phi_\alpha(0.60, 0.80) - \Phi_\alpha(0.45, 0.85) \\ &= \int_{0.45}^{0.60} \int_{0.80}^{0.85} \frac{\partial^2 \Phi_\alpha}{\partial u \partial v}(u, v) dv du. \end{aligned} \quad (97)$$

To determine the sign of the integrand, let us further define $x = b^{\alpha-1} - a^{\alpha-1}$, and $y = c^{\alpha-1} - b^{\alpha-1}$. Since $a \geq b \geq c > 0$ and $\alpha - 1 < 0$, we have $x, y \geq 0$. On differentiation, we obtain

$$\begin{aligned} \frac{\partial^2 \Phi_\alpha}{\partial u \partial v} &= \frac{\alpha b^{\alpha-2}}{s} - \frac{\alpha^2 xy}{(1 - \alpha)s^2} \\ &= \frac{\alpha b^{\alpha-2}}{s} (1 - R), \end{aligned} \quad (98)$$

where

$$R \triangleq \frac{\alpha xy}{(1 - \alpha)b^{\alpha-2}s}.$$

Then, by the mean value theorem we have $x \leq (1 - \alpha)(a - b)b^{\alpha-2}$ and $y \leq (1 - \alpha)(b - c)c^{\alpha-2}$. Moreover, throughout the rectangle $\mathcal{R}$, we have $a - b \leq 0.40$, $b - c \leq 0.25$, and $c \geq 0.15$. Using $s \geq 2c^\alpha$ and $\alpha(1 - \alpha) \leq 1/4$, we obtain the following

$$\begin{aligned} 0 \leq R &\leq \frac{\alpha(1 - \alpha)(a - b)(b - c)c^{\alpha-2}}{s} \\ &\leq \frac{\alpha(1 - \alpha)(a - b)(b - c)}{2c^2} \\ &\leq \frac{(1/4)(0.40)(0.25)}{2(0.15)^2} = \frac{5}{9} < 1. \end{aligned}$$

Therefore, we have

$$\frac{\partial^2 \Phi_\alpha}{\partial u \partial v} \geq \frac{4}{9} \frac{\alpha b^{\alpha-2}}{s} > 0$$

throughout the rectangular set $\mathcal{R}$. It follows from (97) that $\Delta_\alpha(\mathbf{p}, \mathbf{q}) > 0$, for every $\alpha \in (0, 1)$. Hence, the Rényi entropy H_α is not submodular on $\mathcal{P}_3$ for any $\alpha \in (0, 1)$.

Now, we show that for every fixed $\alpha \in (0, 1)$, there exists a pair of PMFs supported on a set of cardinality three such that $\Delta_\alpha(\mathbf{p}, \mathbf{q}) < 0$. Fix an $\alpha \in (0, 1)$ and consider

$$(a, b, c) = (0.60, 0.40 - t, t),$$

where $t \in (0, 0.20)$. This vector is strictly decreasing and strictly positive, with cumulative sum coordinates $u = 0.60$, and $v = 1 - t$. Using the same notation of s, x, y as introduced above, as $t \rightarrow 0$ we have that $s \rightarrow 0.60^\alpha + 0.40^\alpha > 0$, $x \rightarrow 0.40^{\alpha-1} - 0.60^{\alpha-1} > 0$, $y \rightarrow +\infty$. Consequently, the

term $\alpha b^{\alpha-2}/s$ is bounded in (98), while $-\alpha^2 xy/((1 - \alpha)s^2)$ tends to $-\infty$. Thus

$$\frac{\partial^2 \Phi_\alpha}{\partial u \partial v}(0.60, 1 - t) \rightarrow -\infty \quad \text{as } t \rightarrow 0. \quad (99)$$

Now, choose $t_0 \in (0, 0.20)$ sufficiently small that the mixed derivative is strictly negative at $(0.60, 1 - t_0)$. Set $\mathbf{x} = (a, b, c) = (0.60, 0.40 - t_0, t_0)$, $A = a$, $B = a + b$. Since $a > b > c > 0$, the point (A, B) lies in the open region

$$\mathcal{D} \triangleq \{(u, v) : u > v - u > 1 - v > 0\}.$$

By continuity, there exist $\varepsilon, \eta > 0$ such that the rectangle $[A, A + \varepsilon] \times [B, B + \eta]$ is contained in $\mathcal{D}$ and the mixed derivative is strictly negative throughout this rectangle. Further, let us define

$$\begin{aligned} \mathbf{p}_\varepsilon &= (a + \varepsilon, b - \varepsilon, c), \\ \mathbf{q}_\eta &= (a, b + \eta, c - \eta). \end{aligned}$$

Their respective cumulative breakpoint sets are

$$\begin{aligned} B_{\mathbf{p}_\varepsilon} &= \{0, A + \varepsilon, B, 1\}, \\ B_{\mathbf{q}_\eta} &= \{0, A, B + \eta, 1\}. \end{aligned}$$

Since the rectangle is contained in $\mathcal{D}$, the pointwise minimum and maximum of the cumulative sums both have strictly decreasing increments. Therefore, on taking the pointwise minimum and maximum of the cumulative sums we get

$$\begin{aligned} \mathbf{p}_\varepsilon \wedge \mathbf{q}_\eta &= \mathbf{x}, \\ \mathbf{p}_\varepsilon \vee \mathbf{q}_\eta &= (a + \varepsilon, b + \eta - \varepsilon, c - \eta). \end{aligned}$$

Then, it follows that

$$\begin{aligned} \Delta_\alpha(\mathbf{p}_\varepsilon, \mathbf{q}_\eta) &= \Phi_\alpha(A + \varepsilon, B + \eta) + \Phi_\alpha(A, B) \\ &\quad - \Phi_\alpha(A + \varepsilon, B) - \Phi_\alpha(A, B + \eta) \\ &= \int_0^\varepsilon \int_0^\eta \frac{\partial^2 \Phi_\alpha}{\partial u \partial v}(A + \sigma, B + \tau) d\tau d\sigma \\ &< 0. \end{aligned} \quad (100)$$

Consequently, we see that the supermodularity property is violated. Since $\alpha \in (0, 1)$ was arbitrarily chosen, this construction proves that H_α is not supermodular on $\mathcal{P}_3$ for any $\alpha \in (0, 1)$. The choices of t_0, ε, η , and hence of the pair $(\mathbf{p}_\varepsilon, \mathbf{q}_\eta)$ may depend on α . For illustration, the following fixed four coordinate pair has a negative gap at selected orders $\alpha \in (0, 1)$.

Example E.2 – Pair with $\Delta_\alpha(\mathbf{p}, \mathbf{q}) < 0$. Let $\mathbf{p}, \mathbf{q} \in \mathcal{P}_4$ be

$$\begin{aligned} \mathbf{p} &= (0.42, 0.28, 0.20, 0.10), \\ \mathbf{q} &= (0.40, 0.30, 0.22, 0.08). \end{aligned}$$

Their first three cumulative sums are $(0.42, 0.70, 0.90)$ and $(0.40, 0.70, 0.92)$, respectively. The pointwise minimum and maximum again have decreasing increments. Consequently, the glb and lub are

$$\begin{aligned} \mathbf{w} &\triangleq \mathbf{p} \wedge \mathbf{q} = (0.40, 0.30, 0.20, 0.10), \\ \mathbf{z} &\triangleq \mathbf{p} \vee \mathbf{q} = (0.42, 0.28, 0.22, 0.08). \end{aligned}$$

For illustration, the corresponding numerical values of

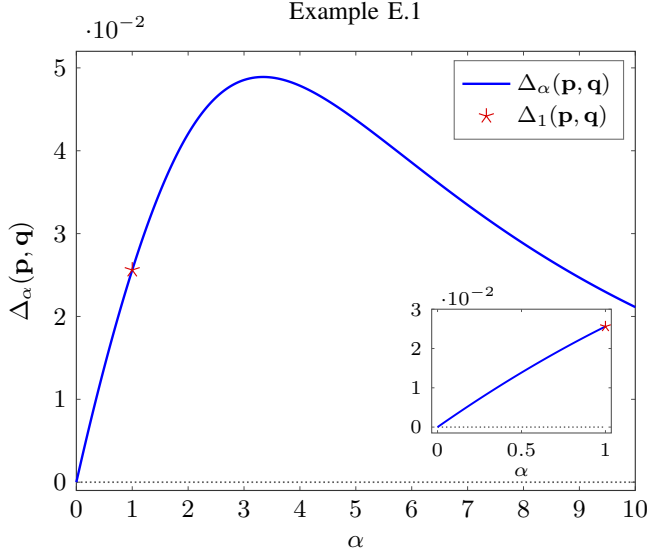


Fig. 1. Modularity gap for Rényi entropy, $\Delta_\alpha(\mathbf{p}, \mathbf{q})$, in Example E.1.

$\Delta_\alpha(\mathbf{p}, \mathbf{q})$ are

α	$\Delta_\alpha(\mathbf{p}, \mathbf{q})$
0	0
0.2	$-5.0348184 \times 10^{-6}$
0.5	$-1.8208862 \times 10^{-5}$
0.7	$-1.9885069 \times 10^{-5}$
0.9	$-1.0017001 \times 10^{-5}$
1	0
2	2.4803125×10^{-4}
∞	0

(101)

Therefore, through Example E.1 we showed that $\Delta_\alpha(\mathbf{p}, \mathbf{q}) > 0$ for every $\alpha \in (0, 1)$, disproving submodularity on $\mathcal{P}_3$. The three-coordinate PMF perturbation construction gives, for every fixed $\alpha \in (0, 1)$, a pair with $\Delta_\alpha(\mathbf{p}, \mathbf{q}) < 0$, disproving supermodularity on $\mathcal{P}_3$. Example E.2 further illustrates this failure of supermodularity through a fixed four-coordinate pair whose modularity gap is evaluated at selected orders in $(0, 1)$. Embedding these three-coordinate examples in $\mathcal{P}_n$ helps us to establish that H_α is neither supermodular nor submodular for every $n \geq 3$ and every $\alpha \in (0, 1)$. This completes the proof. $\square$

E. Proof of Theorem 6

Proof. Our proof method is based on studying the Lorenz curves of $\mathbf{p}$, $\mathbf{q}$, $\mathbf{p} \wedge \mathbf{q}$, and $\mathbf{p} \vee \mathbf{q}$. We handle the cases of $\alpha < 1$ and $\alpha \geq 1$ separately below.

(Proof for $\alpha \in (0, 1)$). Let $S_\alpha(\mathbf{p}) := \sum_{i=1}^n p_i^\alpha$, as defined in the proof of Theorem 4. Further, let $L_{\mathbf{p}}$, $L_{\mathbf{q}}$, $L_{\mathbf{w}}$, $L_{\mathbf{v}}$, m , and M be also as defined in the proof of Theorem 4. Then we have that,

$$\int_0^n (m'(t))^\alpha dt + \int_0^n (M'(t))^\alpha dt = S_\alpha(\mathbf{p}) + S_\alpha(\mathbf{q}). \quad (102)$$

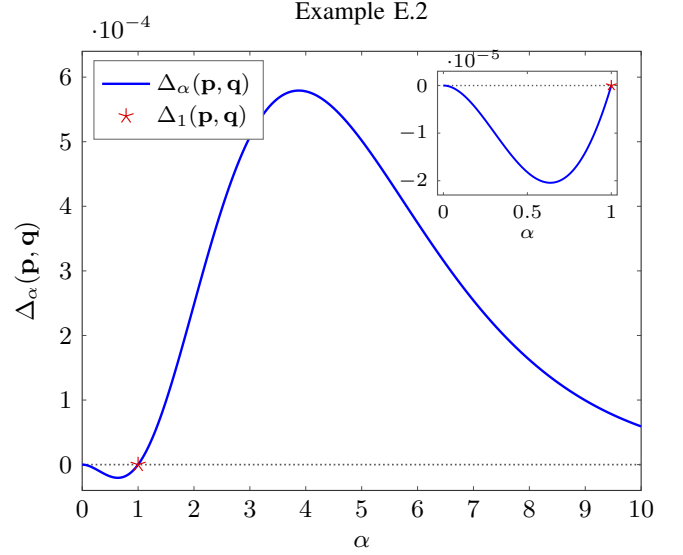


Fig. 2. Modularity gap for Rényi entropy, $\Delta_\alpha(\mathbf{p}, \mathbf{q})$, in Example E.2.

Now, fix an interval $I_k \triangleq (k-1, k)$. Then, m is a piece-wise linear concave function within I_k while $L_{\mathbf{w}}$ is a straight line joining $m(k-1)$ to $m(k)$. Thus,

$$L'_{\mathbf{w}}(t) = \frac{m(k) - m(k-1)}{1} = \int_{k-1}^k m'(t) dt$$

Since $u \rightarrow u^\alpha$ is a concave map for $\alpha < 1$, using Jensen's inequality we obtain

$$(L'_{\mathbf{w}}(t))^\alpha \geq \int_{k-1}^k m'(t)^\alpha dt \quad (103)$$

Now, we consider the function M in the same interval I_k . Unlike m , M is a piece-wise linear (not necessarily concave) function while $F_{\mathbf{v}}$ is a straight line joining $M(k-1)$ to $M(k)$. If I_k has no non-integer crossings, then $F'_{\mathbf{v}}(t) = M'(t)$.

However, if there are crossings then M is convex and we need to concavify it to get $F_{\mathbf{v}}$. Thus,

$$F'_{\mathbf{v}}(t) = \frac{M(k) - M(k-1)}{1} = \int_{k-1}^k M'(t) dt$$

Since $u \rightarrow u^\alpha$ is a concave map for $\alpha < 1$, using Jensen's inequality we have

$$(F'_{\mathbf{v}}(t))^\alpha \geq \int_{k-1}^k (M'(t))^\alpha dt \quad (104)$$

Now, if $F_{\mathbf{v}}$ is a piecewise-linear concave function, then it is the lub $L_{\mathbf{v}}$. Otherwise, it needs to be concavified and consequently the lub $L_{\mathbf{v}}$ is the least concave majorant (LCM) of $F_{\mathbf{v}}$.

Therefore applying Jensen again over the pooled intervals used by the LCM operation, we get

$$S_\alpha(\mathbf{v}) = \int_0^n (L'_{\mathbf{v}}(t))^\alpha dt \quad (105)$$

$$\geq \int_0^n (F'_{\mathbf{v}}(t))^\alpha dt \quad (106)$$

$$\geq \int_0^n (M'(t))^\alpha dt. \quad (107)$$

On summing (103) over all intervals I_k for $k \in [n]$ and using (107), we obtain

$$S_\alpha(\mathbf{w}) + S_\alpha(\mathbf{v}) \geq \int_0^n (m'(t))^\alpha dt + \int_0^n (M'(t))^\alpha dt \quad (108)$$

Therefore, from (102) we have

$$S_\alpha(\mathbf{w}) + S_\alpha(\mathbf{v}) \geq S_\alpha(\mathbf{p}) + S_\alpha(\mathbf{q}). \quad (109)$$

Recall that $T_\alpha(\mathbf{x}) := \frac{1-S_\alpha(\mathbf{x})}{\alpha-1}$, therefore we have that

$$1 - S_\alpha(\mathbf{w}) + 1 - S_\alpha(\mathbf{v}) \leq 1 - S_\alpha(\mathbf{p}) + 1 - S_\alpha(\mathbf{q}) \quad (110)$$

Since $\alpha - 1 < 0$, on dividing both sides by $\alpha - 1$ we obtain

$$\frac{1 - S_\alpha(\mathbf{w})}{\alpha - 1} + \frac{1 - S_\alpha(\mathbf{v})}{\alpha - 1} \geq \frac{1 - S_\alpha(\mathbf{p})}{\alpha - 1} + \frac{1 - S_\alpha(\mathbf{q})}{\alpha - 1} \quad (111)$$

Therefore,

$$T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}) \leq T_\alpha(\mathbf{w}) + T_\alpha(\mathbf{v}). \quad (112)$$

This completes our proof and shows that Tsallis entropy is supermodular for the case of $\alpha < 1$.

(Proof for $\alpha \geq 1$). The proof for the case of $\alpha > 1$ directly follows from the proof of Theorem 4. Let $S_\alpha(\mathbf{p}) := \sum_{i=1}^n p_i^\alpha$, as defined in the proof of Theorem 4. Thus, from (42) we know that for $\alpha \geq 1$,

$$S_\alpha(\mathbf{w}) + S_\alpha(\mathbf{v}) \leq S_\alpha(\mathbf{p}) + S_\alpha(\mathbf{q}), \quad (113)$$

where $\mathbf{w} := \mathbf{p} \wedge \mathbf{q}$ and $\mathbf{v} := \mathbf{p} \vee \mathbf{q}$. Recall that $T_\alpha(\mathbf{x}) := \frac{1-S_\alpha(\mathbf{x})}{\alpha-1}$, therefore we have that

$$1 - S_\alpha(\mathbf{w}) + 1 - S_\alpha(\mathbf{v}) \geq 1 - S_\alpha(\mathbf{p}) + 1 - S_\alpha(\mathbf{q}) \quad (114)$$

Since $\alpha - 1 > 0$, on dividing both sides by $\alpha - 1$ we obtain

$$\frac{1 - S_\alpha(\mathbf{w})}{\alpha - 1} + \frac{1 - S_\alpha(\mathbf{v})}{\alpha - 1} \geq \frac{1 - S_\alpha(\mathbf{p})}{\alpha - 1} + \frac{1 - S_\alpha(\mathbf{q})}{\alpha - 1} \quad (115)$$

Therefore,

$$T_\alpha(\mathbf{p}) + T_\alpha(\mathbf{q}) \leq T_\alpha(\mathbf{w}) + T_\alpha(\mathbf{v}). \quad (116)$$

This completes our proof and shows that Tsallis entropy is supermodular for the case of $\alpha > 1$. The $\alpha \rightarrow 1$ case follows due to the continuity of Tsallis entropy in order α .

(Modularity for $\alpha = 0$). Let $a := |\text{supp}(\mathbf{p})|$ and $b := |\text{supp}(\mathbf{q})|$, where $a, b \leq n$. For $\alpha \rightarrow 0$, we have $|\text{supp}(\mathbf{p} \wedge \mathbf{q})| = \max(a, b)$ and $|\text{supp}(\mathbf{p} \vee \mathbf{q})| = \min(a, b)$. Therefore,

$$T_0(\mathbf{p} \wedge \mathbf{q}) + T_0(\mathbf{p} \vee \mathbf{q}) = \max(a, b) + \min(a, b) - 2 \quad (117)$$

$$= a + b - 2 \quad (118)$$

$$= T_0(\mathbf{p}) + T_0(\mathbf{q}). \quad (119)$$

This completes our proof. $\square$

G. Proof of Theorem 7

Proof. We are given the collection $\mathcal{S} := \{\mathbf{p}^{(i)}\}_{i=1}^m$, where every $\mathbf{p}^{(i)}$ has at most n states. In order to prove our result, we will show that if each element of $\bigwedge \mathcal{S}$ is split equally into m probability masses, then the resulting mn length vector is

majorized by the comonotone coupling vector i.e., if $\mathbf{u} \in \mathcal{P}_m$ such that $\mathbf{u} = (\frac{1}{m}, \dots, \frac{1}{m})$ is the uniform PMF, then

$$\bigwedge \mathcal{S} \otimes \mathbf{u} \preceq \Pi_c^\downarrow(\mathcal{S}). \quad (120)$$

For every $k \in [n]$, let

$$T_k = \min_{i \in [m]} \left\{ \sum_{j=1}^k p_j^{(i)} \right\} \quad (121)$$

Now fix a $k \in [n]$. Then, consider the interval $(0, T_k]$ in the comonotone coupling. Let i^* denote the index of the marginal PMF achieving the minimum in (121). Then, $\mathbf{p}^{(i^*)}$ contributes at most k distinct breakpoints in $(0, T_k]$, including the endpoint T_k , while each of the other $m - 1$ marginal PMFs contributes at most $k - 1$ distinct breakpoints strictly inside $(0, T_k)$. The endpoint T_k is counted only once, even if it is shared by several marginals. Thus, the interval $(0, T_k]$ has mass T_k and contains at most $m(k - 1) + 1$ distinct breakpoints i.e., at most $m(k - 1) + 1$ atoms of the comonotone coupling. Consequently, for the ‘largest’ $m(k - 1) + 1$ atoms of the comonotone coupling we have that

$$\sum_{j=1}^{mk-(m-1)} \Pi_c^\downarrow(j) \geq T_k = \sum_{j=1}^k \left(\bigwedge \mathcal{S} \right) (j) \quad (122)$$

$$= \sum_{j=1}^{mk} \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) (j) \quad (123)$$

$$\geq \sum_{j=1}^{mk-(m-1)} \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) (j) \quad (124)$$

where (122) follows from the definition of the glb $\bigwedge \mathcal{S}$, and (123) holds as each element of $\bigwedge \mathcal{S}$ is equally split into m equal parts. Furthermore, using (123), we can strengthen the inequality in (124) to obtain the following,

$$\sum_{j=1}^{mk} \Pi_c^\downarrow(j) \geq \sum_{j=1}^{mk-1} \Pi_c^\downarrow(j) \quad (125)$$

$\vdots$

$$\geq \sum_{j=1}^{mk-(m-1)} \Pi_c^\downarrow(j) \quad (126)$$

$$\geq \sum_{j=1}^{mk} \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) (j) \quad (127)$$

$$\geq \sum_{j=1}^{mk-1} \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) (j) \quad (128)$$

$\vdots$

$$\geq \sum_{j=1}^{mk-(m-1)} \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) (j) \quad (129)$$

On iterating over all $k \in [n]$, we have that the partial sum of $\Pi_c^\downarrow$ is always at least as large as the partial sum of $\bigwedge \mathcal{S} \otimes \mathbf{u}$,

i.e., for all $r \in [mn]$,

$$\sum_{j=1}^r \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) (j) \leq \sum_{j=1}^r \Pi_c^\downarrow(j) \quad (130)$$

which implies

$$\bigwedge \mathcal{S} \otimes \mathbf{u} \preceq \Pi_c^\downarrow(\mathcal{S}) \quad (131)$$

By Schur concavity of Rényi entropy we have

$$H_\alpha(\Pi_c^\downarrow(\mathcal{S})) \leq H_\alpha \left(\bigwedge \mathcal{S} \otimes \mathbf{u} \right) \quad (132)$$

$$= H_\alpha \left(\bigwedge \mathcal{S} \right) + H_\alpha(\mathbf{u}) \quad (133)$$

$$= H_\alpha \left(\bigwedge \mathcal{S} \right) + \log m \quad (134)$$

$$\leq H_\alpha^*(\text{OPT}_\mathcal{S}) + \log m. \quad (135)$$

Here (133) holds from noting that $\mathbf{u}$ is independent of $\bigwedge \mathcal{S}$ and the additivity of Rényi entropy for product distributions. (135) holds due to the fact that $\bigwedge \mathcal{S}$ yields a lower bound on the Rényi entropy of the optimal coupling i.e., $H_\alpha(\bigwedge \mathcal{S}) \leq H_\alpha(Q^*) \leq H_\alpha^*(\text{OPT}_\mathcal{S})$.

Optimality for $\alpha = \infty$. We will now show that the comonotone coupling is optimal for the min-entropy. For any coupling $\Pi \in \mathcal{C}(\mathcal{S})$, we know that every probability mass of Π is upper bounded by the largest mass of each marginal PMF. Thus,

$$\|\Pi\|_\infty = \max_{i \in [n^m]} \Pi_i \leq \min_{i \in [m]} \max_{j \in [n]} p_j^{(i)}. \quad (136)$$

Moreover, note from Algorithm 1 that the first element (also the largest element) of the output comonotone coupling vector $\Pi_c^\downarrow(\mathcal{S})$ is

$$\Pi_c^\downarrow(\mathcal{S})(1) = \min_{i \in [m]} \max_{j \in [n]} p_j^{(i)}. \quad (137)$$

Therefore,

$$\|\Pi_c(\mathcal{S})\|_\infty = \min_{i \in [m]} \max_{j \in [n]} p_j^{(i)} = \sup_{\Pi \in \mathcal{C}(\mathcal{S})} \|\Pi\|_\infty. \quad (138)$$

Since $H_\infty(\Pi) = -\log \|\Pi\|_\infty$, thus maximizing the largest probability mass is equivalent to minimizing H_∞ . Therefore

$$H_\infty(\Pi_c(\mathcal{S})) = \inf_{\Pi \in \mathcal{C}(\mathcal{S})} H_\infty(\Pi) = H_\infty^*(\text{OPT}_\mathcal{S}). \quad (139)$$

This completes our proof of our theorem. $\square$

H. Proof of Corollary 2

Proof. Let us define the PMF $\mathbf{v}$ by splitting every atom of $\bigwedge \mathcal{S}$ into m equal parts, except the first atom,

$$\mathbf{v} = \left(w_1, \underbrace{\frac{w_2}{m}, \dots, \frac{w_2}{m}}_m, \dots, \underbrace{\frac{w_n}{m}, \dots, \frac{w_n}{m}}_m \right).$$

We will show that $\mathbf{v} \preceq \Pi_c^\downarrow(\mathcal{S}) \preceq \bigwedge \mathcal{S}$. Let us use the shorthand $\mathbf{c} = \Pi_c^\downarrow(\mathcal{S})$ and $r_k = 1 + m(k-1)$. The breakpoint set analysis in Appendix G gives us

$$\sum_{j=1}^{r_k} c_j \geq \sum_{j=1}^k w_j = \sum_{j=1}^{r_k} v_j.$$

The linearly interpolated cumulative-sum curve of $\mathbf{c}$ is concave, while that of $\mathbf{v}$ is linear between consecutive indices r_k . Consequently, the same inequality holds at every intermediate index, which allows us to prove $\mathbf{v} \preceq \mathbf{c}$. The majorization order $\mathbf{c} \preceq \bigwedge \mathcal{S}$ follows because every feasible coupling is majorized by each marginal and therefore also by their glb. Thus, by Schur concavity,

$$\begin{aligned} H_\alpha(\mathbf{c}) &\leq H_\alpha(\mathbf{v}) \\ &= H_\alpha \left(\bigwedge \mathcal{S} \right) + \varepsilon_\alpha \left(\bigwedge \mathcal{S} \right) \\ &\leq H_\alpha^*(\text{OPT}_\mathcal{S}) + \varepsilon_\alpha \left(\bigwedge \mathcal{S} \right), \end{aligned}$$

where the expressions for $\varepsilon_\alpha(\bigwedge \mathcal{S})$ can be computed directly. Furthermore, the bound $\varepsilon_\alpha(\bigwedge \mathcal{S}) \leq \log m$ and monotonicity in α can be verified from the computed expressions. $\square$

I. Optimality and Suboptimality cases for Algorithm 1

We only consider examples with two marginal PMFs $\mathbf{p}$ and $\mathbf{q}$. We compute different lower bounds (converses) and upper bounds (achievable constructions) to compare them.

• **Example 1:** In this example, we show that both the comonotone coupling as well as the greedy coupling match the information-spectrum lower bound Q^* in [20]. Therefore, both algorithms are provably optimal. Let $\mathbf{p} = (p, 1-p)$ and $\mathbf{q} = (q, 1-q)$ such that $p, q \in [1/2, 1)$ and $p > q$. Then, we have that $\mathbf{p} \wedge \mathbf{q} = \mathbf{q}$. The stronger information-spectrum based lower bound Q^* is

$$Q^* = (q, \max(p-q, 1-p), \min(p-q, 1-p))$$

Among the achievability results (coupling constructions), we compute the comonotone coupling - $\Pi_c^\downarrow$ (our Algorithm 1), greedy coupling - $\Pi_g^\downarrow$ ([22]), glb (greatest lower bound) splitting based coupling - $\Pi_{\text{CGV}}^\downarrow$ ([51]), and the geometric splitting based coupling - $\Pi_{\text{Li}}^\downarrow$ ([18]). Consequently, we get that

$$\begin{aligned} \Pi_c^\downarrow &= \Pi_g^\downarrow \\ &= \Pi_{\text{CGV}}^\downarrow \\ &= \Pi_{\text{Li}}^\downarrow \\ &= (q, \max(p-q, 1-p), \min(p-q, 1-p)) \\ &= Q^*. \end{aligned}$$

Thus, all the above existing approximation algorithms are optimal (for Rényi entropy of any order $\alpha \in [0, \infty]$) for a collection of two Bernoulli marginal PMFs. This is not true in general for a collection of more than two PMFs.

• **Example 2:** In this example, we show that the comonotone coupling matches the information-spectrum lower bound Q^* , and therefore is optimal. However, the greedy coupling algorithm is suboptimal for every finite Rényi order $\alpha \in [0, \infty)$. Consider the following PMFs,

$$\begin{aligned} \mathbf{p} &= (0.6, 0.3, 0.05, 0.05), \\ \mathbf{q} &= (0.35, 0.25, 0.2, 0.2). \end{aligned}$$

Among, the converse results (lower bounds), we have

$$\mathbf{p} \wedge \mathbf{q} = (0.35, 0.25, 0.2, 0.2),$$

$$Q^* = (0.35, 0.25, 0.2, 0.1, 0.05, 0.05)$$

Among the achievability results (coupling constructions), we compute the comonotone coupling $\Pi_c^\downarrow$ (our Algorithm 1), greedy coupling $\Pi_g^\downarrow$ (see [22]), glb (greatest lower bound) based splitting coupling $\Pi_{CGV}^\downarrow$ (see [51]), and the geometric splitting based coupling $\Pi_{Li}^\downarrow$ (see [18]).

$$\Pi_c^\downarrow = (0.35, 0.25, 0.2, 0.1, 0.05, 0.05)$$

$$\Pi_g^\downarrow = (0.35, 0.25, 0.2, 0.05, 0.05, 0.05, 0.05)$$

$$\Pi_{CGV}^\downarrow = (0.35, 0.15, 0.15, 0.15, 0.10, 0.05, 0.05)$$

$$\Pi_{Li}^\downarrow = (0.35, 0.15, 0.15, 0.15, 0.10, 0.05, 0.05)$$

In Figure 3, we plot the Rényi entropy curves $H_\alpha(\cdot)$ for the lower bounds and achievable coupling constructions to compare the performances of different algorithms.

• **Example 3:** In this example, none of the coupling constructions compared below attains the information-spectrum lower bound, for finite order Rényi entropy. Nevertheless, our Algorithm 1 produces the coupling with the smallest Rényi entropy among these constructions. However, we exhibit a feasible coupling with strictly smaller entropy than all of the four constructions for every $0 < \alpha < \infty$, thereby establishing their suboptimality in this range. Consider the following PMFs,

$$\mathbf{p} = (0.36, 0.28, 0.24, 0.12),$$

$$\mathbf{q} = (0.56, 0.16, 0.16, 0.12).$$

Among the converse results (lower bounds), we have

$$\mathbf{p} \wedge \mathbf{q} = (0.36, 0.28, 0.24, 0.12),$$

$$Q^* = (0.36, 0.20, 0.16, 0.16, 0.12)$$

Among the achievability results (coupling constructions), we compute the comonotone coupling $\Pi_c^\downarrow$ (our Algorithm 1), greedy coupling $\Pi_g^\downarrow$ (see [22]), glb (greatest lower bound) based splitting coupling $\Pi_{CGV}^\downarrow$ (see [51]), and the geometric splitting based coupling $\Pi_{Li}^\downarrow$ (see [18]),

$$\Pi_c^\downarrow = (0.36, 0.20, 0.16, 0.12, 0.08, 0.08),$$

$$\Pi_g^\downarrow = (0.36, 0.20, 0.16, 0.12, 0.08, 0.04, 0.04),$$

$$\Pi_{CGV}^\downarrow = (0.36, 0.16, 0.16, 0.12, 0.12, 0.08),$$

$$\Pi_{Li}^\downarrow = (0.36, 0.16, 0.16, 0.12, 0.12, 0.08).$$

To establish sub-optimality of the existing coupling constructions, we can consider this alternative coupling matrix

$$\Pi_{alt} = \begin{pmatrix} 0.36 & 0 & 0 & 0 \\ 0 & 0 & 0.16 & 0.12 \\ 0.20 & 0.04 & 0 & 0 \\ 0 & 0.12 & 0 & 0 \end{pmatrix}.$$

Its row sums yield $\mathbf{p}$ and its column sums yield $\mathbf{q}$, so it is a feasible coupling of the prescribed marginals. Its sorted PMF

vector is

$$\Pi_{alt}^\downarrow = (0.36, 0.20, 0.16, 0.12, 0.12, 0.04).$$

We observe that this PMF vector majorizes each of the four coupling vectors above and is distinct from each of them. In particular, its first four cumulative sums are same as those of $\Pi_c^\downarrow$, while its fifth cumulative sum is 0.96, compared with 0.92 for $\Pi_c^\downarrow$. Thus, all four constructions are sub-optimal for every finite $\alpha \in (0, \infty)$. At order infinity, all four constructions are optimal, since their largest mass is 0.36, the largest possible coupling atom for these marginals.

In Figure 4, we plot the Rényi entropy curves $H_\alpha(\cdot)$ for the lower bounds and achievable coupling constructions to compare the performances of different algorithms.

• **Example 4:** Through this example, we show that the greedy coupling algorithm may not be optimal under the conditions of Corollary 3. Consider the marginal PMFs

$$\mathbf{p} = (0.30, 0.25, 0.20, 0.15, 0.10)$$

$$\mathbf{q} = (0.55, 0.45),$$

with zero-padding for the smaller PMF vector. Their interior breakpoint sets satisfy

$$B_q = \{0.55\} \subset B_p = \{0.30, 0.55, 0.75, 0.90\}.$$

Thus, the condition in Corollary 3 holds and Algorithm 1 produces the output coupling vector

$$\Pi_c^\downarrow = (0.30, 0.25, 0.20, 0.15, 0.10) = \mathbf{p}.$$

This coupling attains the marginal entropy lower bound and is therefore optimal for every order $\alpha \in [0, \infty]$. In contrast, the greedy coupling algorithm outputs

$$\Pi_g^\downarrow = (0.30, 0.25, 0.20, 0.15, 0.05, 0.05).$$

Indeed, we observe that the first four allocations have masses 0.30, 0.25, 0.20, and 0.15. The remaining mass of $\mathbf{p}$ is 0.10, whereas the two remaining masses of $\mathbf{q}$ are both 0.05. Therefore, the greedy algorithm splits the final atom into two equal parts. Since $\Pi_g^\downarrow$ is obtained from $\Pi_c^\downarrow$ by splitting a positive atom, the Schur concavity gives us

$$H_\alpha(\Pi_g^\downarrow) > H_\alpha(\Pi_c^\downarrow)$$

for every $0 < \alpha < \infty$. The inequality also holds at $\alpha = 0$. At $\alpha = \infty$, both couplings are optimal, since their largest mass is 0.30.

These examples illustrate optimality and suboptimality of Algorithm 1 and compare it with the other coupling constructions.

J. Proof of Theorem 8

Proof. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{P}_n$. A distance function $d_\alpha(\mathbf{x}, \mathbf{y})$ on $\mathcal{P}_n$ is said to be a metric if it satisfies the following conditions:

- 1) Symmetry, i.e., $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
- 2) Positivity, i.e., $d(\mathbf{x}, \mathbf{x}) = 0$ and $d(\mathbf{x}, \mathbf{y}) > 0$, for every $\mathbf{x} \neq \mathbf{y}$.
- 3) Triangle inequality, i.e., $d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \geq d(\mathbf{x}, \mathbf{z})$.

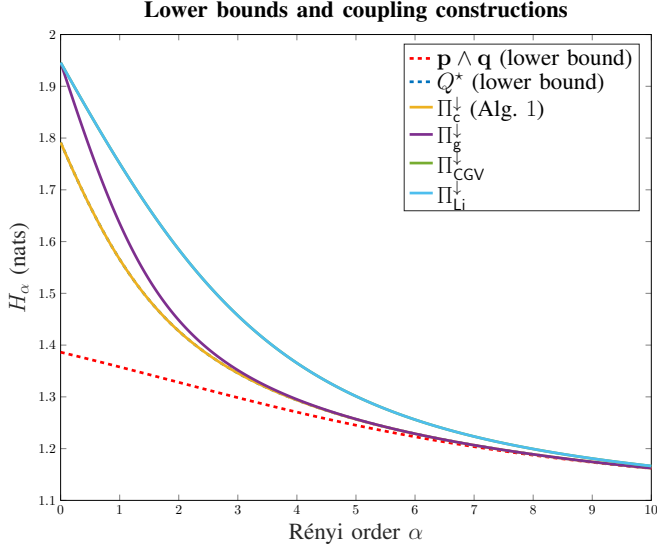


Fig. 3. Example 2: Comparison of lower bounds (converses) and upper bounds (achievable coupling constructions) for minimum Rényi entropy couplings. We observe that the lower bound Q^* and the comonotone coupling $\Pi_c^\downarrow$ coincide with each other, and $\Pi_{CGV}^\downarrow = \Pi_{Li}^\downarrow$.

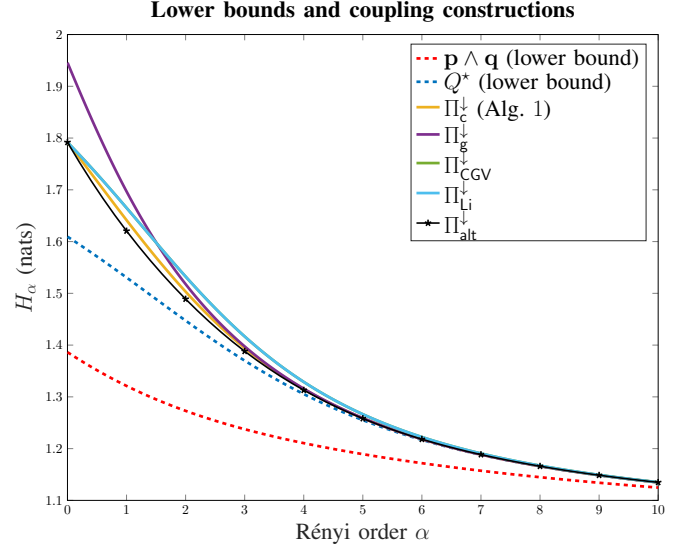


Fig. 4. Example 3: Comparison of lower bounds (converses) and upper bounds (achievable coupling constructions) for minimum Rényi entropy couplings. We observe that $\Pi_{CGV}^\downarrow = \Pi_{Li}^\downarrow$.

We show that the following properties hold for $d_\alpha^R(\cdot, \cdot)$ in detail whereas the analysis for $d_\alpha^T(\cdot, \cdot)$ follows similarly. We have

$$d_\alpha^R(\mathbf{x}, \mathbf{y}) \triangleq H_\alpha(\mathbf{x}) + H_\alpha(\mathbf{y}) - 2H_\alpha(\mathbf{x} \vee \mathbf{y}). \quad (140)$$

Since $\mathbf{x}, \mathbf{y} \preceq \mathbf{x} \vee \mathbf{y}$, both $H_\alpha(\mathbf{x}) - H_\alpha(\mathbf{x} \vee \mathbf{y})$ and $H_\alpha(\mathbf{y}) - H_\alpha(\mathbf{x} \vee \mathbf{y})$ are nonnegative. For $0 < \alpha < \infty$, strict Schur concavity implies that their sum vanishes if and only if $\mathbf{x} = \mathbf{y}$. The symmetry property holds trivially for every $\alpha \in [0, \infty]$. However, the positivity holds only for $\alpha \in (0, \infty)$, as the Rényi entropy of order zero only depends on the support size of the PMFs, and the Rényi entropy of order infinity only depends on the largest probability mass of the PMF. Next, we will use Schur concavity and supermodularity (Theorem 4) to validate the triangle inequality. We have,

$$\begin{aligned} d_\alpha^R(\mathbf{x}, \mathbf{y}) + d_\alpha^R(\mathbf{y}, \mathbf{z}) - d_\alpha^R(\mathbf{x}, \mathbf{z}) \\ = 2 \left[H_\alpha(\mathbf{y}) + H_\alpha(\mathbf{x} \vee \mathbf{z}) \right. \\ \left. - H_\alpha(\mathbf{x} \vee \mathbf{y}) - H_\alpha(\mathbf{y} \vee \mathbf{z}) \right] \end{aligned} \quad (141)$$

$$\begin{aligned} \geq 2 \left[H_\alpha(\mathbf{y}) + H_\alpha(\mathbf{x} \vee \mathbf{z}) \right. \\ \left. - H_\alpha((\mathbf{x} \vee \mathbf{y}) \wedge (\mathbf{y} \vee \mathbf{z})) \right. \\ \left. - H_\alpha((\mathbf{x} \vee \mathbf{y}) \vee (\mathbf{y} \vee \mathbf{z})) \right] \end{aligned} \quad (142)$$

$$\begin{aligned} = 2 \left[\left(H_\alpha(\mathbf{y}) - H_\alpha((\mathbf{x} \vee \mathbf{y}) \wedge (\mathbf{y} \vee \mathbf{z})) \right) \right. \\ \left. + \left(H_\alpha(\mathbf{x} \vee \mathbf{z}) - H_\alpha(\mathbf{x} \vee \mathbf{y} \vee \mathbf{z}) \right) \right] \end{aligned} \quad (143)$$

$$= 2 [A(\mathbf{x}, \mathbf{y}, \mathbf{z}) + B(\mathbf{x}, \mathbf{y}, \mathbf{z})] \quad (144)$$

$$\geq 0. \quad (145)$$

Here (142) holds for every $\alpha \in \{0\} \cup [1, \infty]$ due to the supermodularity of Rényi entropy on the majorization lattice (applied on $\mathbf{x} \vee \mathbf{y}$ and $\mathbf{y} \vee \mathbf{z}$). Furthermore, note that $\mathbf{x} \vee \mathbf{z} \preceq \mathbf{x} \vee \mathbf{y} \vee \mathbf{z}$, thus $B(\mathbf{x}, \mathbf{y}, \mathbf{z}) \geq 0$ due to Schur concavity. Similarly, $\mathbf{y} \preceq \mathbf{x} \vee \mathbf{y}$ as well as $\mathbf{y} \preceq \mathbf{y} \vee \mathbf{z}$. Consequently, $\mathbf{y} \preceq (\mathbf{x} \vee \mathbf{y}) \wedge (\mathbf{y} \vee \mathbf{z})$ and $A(\mathbf{x}, \mathbf{y}, \mathbf{z}) \geq 0$, which implies (145).

We have shown that $d_\alpha^R(\mathbf{x}, \mathbf{y})$ satisfies positivity, symmetry, and triangle inequality for every $\alpha \in [1, \infty)$. Therefore, it is a distance metric on $\mathcal{P}_n$ for all $\alpha \in [1, \infty)$. This completes our proof. $\square$