Fractional Subadditivity of Submodular Functions: Equality Conditions and Their Applications

Gunank Jakhar, Gowtham R. Kurri, Suryajith Chillara
International Institute of Information Technology, Hyderabad
Hyderabad, India.
Email: {gunank.jakhar@research., gowtham.kurri@, suryajith.chillara@}iiit.ac.in

Vinod M. Prabhakaran
Tata Institute of Fundamental Research
Mumbai, India.
Email: vinodmp@tifr.res.in

***Abstract*—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. Submodular functions are known to satisfy various forms of fractional subadditivity. This work investigates the conditions for equality to hold exactly or approximately in the fractional subadditivity of submodular functions. We establish that a small gap in the inequality implies that the function is close to being modular, and that the gap is zero if and only if the function is modular. We then present natural implications of these results for special cases of submodular functions, such as entropy, relative entropy, and matroid rank. As a consequence, we characterize the necessary and sufficient conditions for equality to hold in Shearer's lemma, recovering a result of Ellis *et al.* (2016) as a special case. We leverage our results to propose a new multivariate mutual information, which generalizes Watanabe's total correlation (1960), Han's dual total correlation (1978), and Csiszár and Narayan's shared information (2004), and analyze its properties. Among these properties, we extend Watanabe's characterization of total correlation as the maximum correlation over partitions to fractional partitions. When applied to matrix determinantal inequalities for positive definite matrices, our results recover the equality conditions of the classical determinantal inequalities of Hadamard, Szász, and Fischer as special cases.**

I. Introduction

A submodular function is a set function that exhibits the property of *diminishing returns*, i.e., the additional value gained by adding an element to a set decreases as the set grows larger [1]. Some examples of submodular functions include entropy, matroid rank function, maximum cut in a graph. Submodular functions naturally arise in diverse areas such as machine learning [2], combinatorial optimization [3], algorithmic game theory [4], social networks [5], and statistical physics [6].

It follows from the definition of submodular functions that they satisfy a property of subadditivity, provided that the function takes the value zero on the empty set. Specifically, the value of a submodular function on the ground set is less than or equal to the sum of the function values over any partition of the ground set. This property naturally generalizes to fractional partitions, where the ground set is covered by a family of overlapping subsets, with each element assigned to these subsets with some probability. This generalization, known as fractional subadditivity [7], [8], has many applications, including algorithmic game theory [4], [8] and statistical physics [6]. Entropy, being an important example of a submodular function, satisfies fractional subadditivity [9], which, along with its special case known as Shearer's lemma [10], [11], has found applications in many areas such as graph theory, combinatorics, matrix inequalities, etc. [9], [11].

In this work, we investigate the conditions for equality to hold exactly or approximately in the fractional subadditivity of submodular functions and explore their applications. Our main contributions are as follows:

- We show that a small inequality gap in the fractional subadditivity of submodular functions implies that the function is *approximately* modular (Theorem 3).
- We establish that equality in the fractional subadditivity of a submodular function holds if and only if the function is modular (Theorem 4). This result enables the determination of whether a given submodular function f is modular, with only a minimal additional knowledge about f (see Remark 3, and Examples 1 and 2).
- We explore the implications of these results for specific submodular functions such as entropy, relative entropy, and matroid rank functions (Corollaries 1-4). Notably, we provide the necessary and sufficient conditions for equality to hold in Shearer's lemma [11, Lemma 1] as a special case and this extends a result by Ellis *et al.* [12, Lemma 9]. To the best of our knowledge, these conditions, in this level of generality, do not appear explicitly in the literature.
- As an application, we propose a new multivariate mutual information measure, which generalizes total correlation [13], dual total correlation [14], and shared information [15], [16] (Proposition 1). The latter has operational significance as secret-key capacity for multiple terminals [15]. We analyze the properties of the proposed measure and extend Watanabe's characterization of total correlation as the maximum correlation over partitions [13] to fractional partitions (Theorem 5).
- Additionally, we study applications of our results in matrix determinantal inequalities (Proposition 2), recovering equality conditions for classical determinantal inequalities, including those of Hadamard, Szász, and Fischer [17], [18].

II. Preliminaries

Notation: We use $[i : i+k]$ to represent the set $\{i, i+1, \ldots, i+k-1, i+k\}$, where $i, k \in \mathbb{N}$. The power set of a set A is denoted by 2^A. A^c denotes the complement of a set

A. We use $\mathcal{F}$ to denote a family of subsets of $[1:n]$ allowing for repetitions, represented as $\{\!\{\cdot\}\!\}$.

Definition 1 (Submodular, Supermodular, and Modular Functions [1])**.** *A set function $f: 2^{[1:n]} \to \mathbb{R}$ is called submodular if*

$$f(S) + f(T) \geq f(S \cup T) + f(S \cap T), \quad \forall S, T \subseteq [1:n].$$

A function $f: 2^{[1:n]} \to \mathbb{R}$ is called supermodular if $-f$ is submodular, and modular if it is both submodular and supermodular. Moreover, $f: 2^{[1:n]} \to \mathbb{R}$ is modular if and only if $f(A) = \sum_{i \in A} f(\{i\})$, $\forall A \subseteq [1:n]$.

Madiman and Tetali [9] demonstrated that submodular functions satisfy a conditioning property and a chain rule under an appropriately defined notion of conditioning. Specifically, for $S, T \subseteq [1:n]$, the conditional version of submodular function is defined as $f(S|T) = f(S \cup T) - f(T)$. Let S, T, U be disjoint subsets of $[1:n]$, then [9, Lemma IV] states that

$$f(S|T, U) \leq f(S|T), \tag{1}$$

$$f([1:n]) = \sum_{i=1}^{n} f(i|[1:i-1]), \tag{2}$$

where $f(S|T, U) := f(S|T \cup U)$. We now recall the following related concepts from fractional graph theory [19], [9, Definition II] as we would need them for our results.

Definition 2 (Fractional Covering, Packing, and Partition)**.**

1) *Given a family $\mathcal{F}$ of subsets of $[1:n]$, a function $\alpha: \mathcal{F} \to \mathbb{R}_+$ is called a fractional covering, if for all $i \in [1:n]$, $\sum_{S \in \mathcal{F}: i \in S} \alpha(S) \geq 1$.*
2) *Given a family $\mathcal{F}$ of subsets of $[1:n]$, a function $\beta: \mathcal{F} \to \mathbb{R}_+$ is called a fractional packing, if for all $i \in [1:n]$, $\sum_{S \in \mathcal{F}: i \in S} \beta(S) \leq 1$.*
3) *If $\gamma: \mathcal{F} \to \mathbb{R}_+$ is both a fractional covering and a fractional packing, then it is called a fractional partition.*

In this paper, we investigate the exact and approximate equality conditions of inequalities stated in the following theorem.

Theorem 1 ([9][1])**.** *Let $f: 2^{[1:n]} \to \mathbb{R}$ be any submodular function with $f(\phi) = 0$. Let $\gamma: \mathcal{F} \to \mathbb{R}_+$ be any fractional partition with respect to a family $\mathcal{F}$ of subsets of $[1:n]$. Then*

$$\sum_{S \in \mathcal{F}} \gamma(S) f(S|S^{\mathrm{c}}) \leq f([1:n]) \leq \sum_{S \in \mathcal{F}} \gamma(S) f(S). \tag{3}$$

The fractional partition γ in the lower and upper bounds can be replaced by fractional fractional packing β and fractional covering α, respectively, if the submodular function f is such that $f([1:j])$ is non-decreasing in j for $j \in [1:n]$.

The upper bound in (3), with fractional covering α, has been referred to as fractional subadditivity of submodular functions in the literature [8]. There is a duality between the upper and lower bounds in (3), relating the gaps in the inequalities. For any family $\mathcal{F}$, its complimentary family is defined as $\bar{\mathcal{F}} = \{\!\{S^{\mathrm{c}} : S \in \mathcal{F}\}\!\}$. For a fractional partition γ, its dual fractional partition $\bar{\gamma}$ is defined $\bar{\gamma}(S^{\mathrm{c}}) = \frac{\gamma(S)}{w(\gamma)-1}, \quad \forall S \in \mathcal{F}$, where $w(\gamma)$ denotes the weight of γ, given by $w(\gamma) = \sum_{S \in \mathcal{F}} \gamma(S)$.

Theorem 2 ([9, Theorem IV])**.** *Let $f: 2^{[1:n]} \to \mathbb{R}$ be any submodular function with $f(\phi) = 0$. Let $\gamma: \mathcal{F} \to \mathbb{R}_+$ be any fractional partition with respect to a family $\mathcal{F}$ of subsets of $[1:n]$. Then,*

$$\frac{\mathrm{Gap_U}(f, \mathcal{F}, \gamma)}{w(\gamma)} = \frac{\mathrm{Gap_L}(f, \bar{\mathcal{F}}, \bar{\gamma})}{w(\bar{\gamma})}, \tag{4}$$

where

$$\mathrm{Gap_L}(f, \mathcal{F}, \gamma) = f([1:n]) - \sum_{S \in \mathcal{F}} \gamma(S) f(S|S^{\mathrm{c}}), \text{ and} \tag{5}$$

$$\mathrm{Gap_U}(f, \mathcal{F}, \gamma) = \sum_{S \in \mathcal{F}} \gamma(S) f(S) - f([1:n]). \tag{6}$$

III. Equality Conditions in Fractional Subadditivity

We begin by outlining the assumptions on the family $\mathcal{F}$ of subsets of $[1:n]$ and the fractional partition γ with respect to $\mathcal{F}$, which hold throughout this paper:

1) No two indices $i, j \in [1:n]$ always appear together in the members of $\mathcal{F}$.
2) The family $\mathcal{F}$ includes only proper subsets of $[1:n]$, i.e., $[1:n] \notin \mathcal{F}$.
3) For all $S \in \mathcal{F}$, we have $\gamma(S) > 0$, $\alpha(S) > 0$, and $\beta(S) > 0$.

In Appendix A, we argue that all the above assumptions hold without loss of generality in the context of the fractional subadditivity of submodular functions.

The following theorem provides approximate equality conditions for fractional subadditivity, showing that for a submodular function f, if $\mathrm{Gap_L}(f, \mathcal{F}, \gamma)$ or $\mathrm{Gap_U}(f, \mathcal{F}, \gamma)$ is *small*, then f is *approximately* modular.

Theorem 3. *Let $f: 2^{[1:n]} \to \mathbb{R}$ be any submodular function with $f(\phi) = 0$, and let $\gamma: \mathcal{F} \to \mathbb{R}_+$ be any fractional partition with respect to a family $\mathcal{F}$ of subsets of $[1:n]$. For $\mathrm{Gap_L}(f, \mathcal{F}, \gamma)$ and $\mathrm{Gap_U}(f, \mathcal{F}, \gamma)$ as defined in (5) and (6), respectively, and any $\varepsilon \geq 0$, the following holds:*

If $\mathrm{Gap_L}(f, \mathcal{F}, \gamma) \leq \varepsilon$ or $\mathrm{Gap_U}(f, \mathcal{F}, \gamma) \leq \varepsilon$, then

$$f(\{i\}) + f([1:n] \setminus \{i\}) - f([1:n]) \leq \frac{\varepsilon}{\sigma}, \ \forall i \in [1:n], \tag{7}$$

where $\sigma = \min_{\substack{i,j \in [1:n]: \\ i \neq j}} \sum_{\substack{S \in \mathcal{F}: \\ i \in S, j \notin S}} \gamma(S) > 0$.

Remark 1. Although Theorem 3 is stated for a fractional partition γ, if the submodular function f is such that $f([1:j])$ is non-decreasing in j for $j \in [1:n]$, the theorem's assertions can be generalized to fractional covering α and fractional packing β for $\mathrm{Gap_L}(f, \mathcal{F}, \beta)$ and $\mathrm{Gap_U}(f, \mathcal{F}, \alpha)$, respectively, in a straightforward manner. These details are

[1]The inequality (3) in Theorem 1 is the weak-fractional form of [9, Theorem I], as presented in [9, Section VII].

presented towards the end of Appendix B. We have chosen γ to state Theorem 3 because the quantity σ, as defined in the theorem, may not always be strictly positive when using α or β instead.

The proof of Theorem 3 refines the approach used in the proof of Shearer's lemma by Llewellyn and Radhakrishnan [11], and incorporates insights from its stability version by Ellis *et al.* [12]. A detailed proof is provided in Appendix B.

In the next theorem, we present the necessary and sufficient conditions for the equality to hold in the fractional subadditivity of submodular functions, showing that equality holds if and only if f is modular.

Theorem 4. *Let $f: 2^{[1:n]} \to \mathbb{R}$ be any submodular function with $f(\phi) = 0$. Let $\gamma: \mathcal{F} \to \mathbb{R}_+$, $\alpha: \mathcal{F} \to \mathbb{Q}_+$, and $\beta: \mathcal{F} \to \mathbb{Q}_+$ be any fractional partition, fractional covering, and fractional packing with respect to a family $\mathcal{F}$ of subsets of $[1:n]$, respectively. Then the following hold:*

1) $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma) = 0$ *(similarly,* $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \gamma) = 0$*) if and only if f is modular.*
2) *If f is non-decreasing, i.e., $f(S) \le f(T)$ for $S \subseteq T$, then* $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \alpha) = 0$ *(resp.,* $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \beta) = 0$*) if and only if f is modular and $f(Z) = 0$, for all $Z \subseteq \{i \in [1:n] : \sum_{S \in \mathcal{F}: i \in S} \alpha(S) > 1\}$ (resp., for all $Z \subseteq \{i \in [1:n] : \sum_{S \in \mathcal{F}: i \in S} \beta(S) < 1\}$).*

A detailed proof of Theorem 4 is given in Appendix C. Part 1) of Theorem 4 can be viewed as a corollary of Theorem 3, but part 2) requires that f is a non-decreasing set function. This condition is stronger than the condition for the inequality $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \alpha) \ge 0$ to hold, where $f([1:j])$ is assumed to be non-decreasing in j, for $j \in [1:n]$ (see Theorem 1). Interestingly, this stronger condition is necessary[2] for the proof of part 2) to go through.

Remark 2. While duality, i.e., Theorem 2, establishes a relationship between $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma)$ and $\mathrm{Gap}_{\mathrm{L}}(f, \bar{\mathcal{F}}, \bar{\gamma})$, it does not give a direct connection between $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \gamma)$ and $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma)$ as noted in [9, Section VII]. Nevertheless, from Theorem 4, we can infer that $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \gamma) = 0$ if and only if $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma) = 0$.

Remark 3. Theorem 4 can help determine whether a submodular function f is in fact a modular function with minimal additional knowledge about f. Let us suppose that the values of a submodular function f are known for a family $\mathcal{F}$ of subsets of $[1:n]$, which admits a fractional partition, and for $[1:n]$. For instance, if the family $\mathcal{F} = \{\!\{S \subseteq [1:n] : |S| = k\}\!\}$, for some $k \in [1:n]$, then $\gamma(S) = 1/\binom{n-1}{k-1}$, $\forall S \in \mathcal{F}$ defines a fractional partition. Even without the knowledge of the values of f on the remaining subsets of $[1:n]$, i.e., $2^{[1:n]} \setminus (\mathcal{F} \cup [1:n])$, we can conclude that f is modular if and only if $\sum_{S \in \mathcal{F}} \gamma(S) f(S) = f([1:n])$. This is illustrated in the examples that follow.

[2]Appendix D provides an example of a submodular function that is not modular, where $f([1:j])$ is non-decreasing in j, $f(S) > f(T)$ for some $S \subseteq T$, and $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \alpha) \ge 0$.

Example 1 (Modularity with a symmetric[3] $\mathcal{F}$)**.** Suppose $f: 2^{[1:4]} \to \mathbb{R}$ is a submodular function and the values of f on the members of $\mathcal{F} = \{\!\{\{i\} : i \in [1:4]\}\!\}$ and for $[1:4]$ are given as $f(\{i\}) = i \cdot 2^i$, for $i \in [1:4]$, and $f([1:4]) = 98$. Note that $\gamma: \mathcal{F} \to \mathbb{R}_+$, defined by $\gamma(S) = 1$, $\forall S \in \mathcal{F}$, forms a fractional partition. Since $\sum_{i=1}^4 f(\{i\}) = \sum_{i=1}^4 i \cdot 2^i = 98 = f([1:4])$, Theorem 4 implies that f is a modular function. This conclusion holds irrespective of the values of f on $2^{[1:4]} \setminus (\mathcal{F} \cup [1:4])$.

Example 2 (Modularity with a non-symmetric $\mathcal{F}$)**.** Let $f: 2^{[1:4]} \to \mathbb{R}$ be a submodular function, and the values of f on members of a family $\mathcal{F}$ and on $[1:4]$ are given as: $f(\{2\}) = 0.3$, $f(\{4\}) = 5$, $f(\{1,2\}) = 3$, $f(\{3,4\}) = 0.6$, $f(\{2,4\}) = -2$, $f(\{1,2,3\}) = -1$, $f(\{1,3,4\}) = 3$. Note that $\gamma: \mathcal{F} \to \mathbb{R}_+$, defined by $\gamma(\{2\}) = \frac{1}{6}$, $\gamma(\{4\}) = \frac{5}{12}$, $\gamma(\{1,2\}) = \frac{1}{12}$, $\gamma(\{3,4\}) = \frac{1}{12}$, $\gamma(\{2,4\}) = \frac{1}{6}$, $\gamma(\{1,2,3\}) = \frac{7}{12}$, $\gamma(\{1,3,4\}) = \frac{1}{3}$, forms a fractional partition. Since $\sum_{S \in \mathcal{F}} \gamma(S) f(S) = \frac{151}{60} = f([1:4])$, Theorem 4 implies that f is a modular function. Note that this conclusion holds even though f takes negative values for some subsets.

Identifying the modularity of a submodular function can significantly improve the performance guarantees of optimization algorithms. For example, for certain combinatorial optimization problems, the greedy algorithm is guaranteed to find an exact optimal solution if the submodular function is, in fact, modular [3]. In contrast, for general submodular functions, it typically achieves only an approximately optimal solution [3], [20].

IV. Implications for Entropy, Relative Entropy, and Matroid Rank

In this section, we present the implications of our results from Section III for specific submodular functions: entropy, relative entropy, and matroid rank function.

A. Implications for Entropy

Consider jointly distributed random variables $X_1, X_2, \ldots, X_n$. It is well-known that the function $\mathsf{e}(F) = H(X_F)$, where $H(\cdot)$ denotes entropy and $F \subseteq [1:n]$, is submodular [1].

Corollary 1. *Let $\gamma: \mathcal{F} \to \mathbb{R}_+$ be any fractional partition with respect to a family $\mathcal{F}$ of subsets of $[1:n]$. For jointly distributed random variables $X_1, \ldots, X_n$ with $\mathsf{e}(F) = H(X_F)$, $F \subseteq [1:n]$ and any $\varepsilon \ge 0$, the following holds:*

If $\mathrm{Gap}_{\mathrm{L}}(\mathsf{e}, \mathcal{F}, \gamma) \le \varepsilon$ *or* $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e}, \mathcal{F}, \gamma) \le \varepsilon$*, then*

$$I(X_i; X_{[1:n]\setminus\{i\}}) \le \frac{\varepsilon}{\sigma}, \ \forall i \in [1:n], \tag{8}$$

where $\sigma = \min_{\substack{i,j \in [1:n]: \\ i \ne j}} \sum_{\substack{F \in \mathcal{F}: \\ i \in F, j \notin F}} \gamma(F) > 0.$

The proof of Corollary 1 is given in Appendix E.

[3]We call a family $\mathcal{F}$ symmetric if it remains invariant under every permutation on $[1:n]$.

Remark 4. The assertion of Corollary 1 for $\mathrm{Gap}_{\mathrm{U}}$, in conjunction with Remark 1, recovers [12, Lemma 9] as a special case when $\alpha(F) = \frac{1}{k(\mathcal{F})}$, where $k(\mathcal{F})$ denotes the maximum integer k such that every $i \in [1:n]$ belongs to at least k members of $\mathcal{F}$. This general assertion, involving α, is also implicitly hinted at in [12, Prior to Theorem 4], albeit without a proof.

Corollary 2. *Let $\gamma : \mathcal{F} \to \mathbb{R}_+$ and $\alpha : \mathcal{F} \to \mathbb{Q}_+$ be any fractional partition and any fractional covering with respect to a family $\mathcal{F}$ of subsets of $[1:n]$. For jointly distributed random variables $X_1, \ldots, X_n$ with $\mathsf{e}(F) = H(X_F)$, $F \subseteq [1:n]$, the following hold:*

1) $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e}, \mathcal{F}, \gamma) = 0$ *(similarly,* $\mathrm{Gap}_{\mathrm{L}}(\mathsf{e}, \mathcal{F}, \gamma) = 0$*) if and only if X_i, $i \in [1:n]$ are mutually independent.*
2) $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e}, \mathcal{F}, \alpha) = 0$ *(resp.,* $\mathrm{Gap}_{\mathrm{L}}(\mathsf{e}, \mathcal{F}, \beta) = 0$*) if and only if X_i for i such that $\sum_{F \in \mathcal{F}: i \in F} \alpha(F) = 1$ are mutually independent, and X_i for i such that $\sum_{F \in \mathcal{F}: i \in F} \alpha(F) > 1$ (resp. $\sum_{F \in \mathcal{F}: i \in F} \beta(F) < 1$) are constants.*

Remark 5. (i) The assertion of Corollary 2 on $\mathrm{Gap}_{\mathrm{L}}(\mathsf{e}, \mathcal{F}, \gamma)$ provides equality conditions for the inequality stating that the joint entropy upper bounds the erasure entropy [21, Theorem 1] as a special case when $\mathcal{F} = \{\!\{\{i\} : i \in [1:n]\}\!\}$ and $\gamma(F) = 1, \forall F \in \mathcal{F}$.
(ii) The assertion on $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e}, \mathcal{F}, \alpha)$ recovers the equality conditions for Shearer's lemma [11] as a special case when $\alpha(F) = \frac{1}{k(\mathcal{F})}$, where $k(\mathcal{F})$ denotes the maximum integer k such that each $i \in [1:n]$ belongs to at least k members of $\mathcal{F}$. To the best of our knowledge, these conditions do not appear explicitly in the literature. While the independence of the random variables can be inferred from [12, Proof of Lemma 9], the assertion that some of the random variables are constants does not follow from there.

A detailed proof of Corollary 2 is given in Appendix F. While Corollaries 1 and 2 are stated for discrete entropy, we note that Theorem 1, Corollary 1, and part 1) of Corollary 2 also hold for differential entropy, as it satisfies the submodularity property [1]. However, part 2) of Corollary 2 does not hold because differential entropy is not generally non-decreasing.

B. Implications for Relative Entropy

Let $P_{X_{[1:n]}}$ be any joint probability distribution and $Q_{X_{[1:n]}}$ be a product probability distribution on $\mathcal{X}^n$, i.e., $Q_{X_{[1:n]}}(x_{[1:n]}) = \prod_{i=1}^n Q_{X_i}(x_i)$. For $F \subseteq [1:n]$, let

$$d(F) = -D(P_{X_F} \| Q_{X_F}), \tag{9}$$

where $D(P_{X_F} \| Q_{X_F})$ denotes the relative entropy between the probability distributions P_{X_F} and Q_{X_F}. In [9, Theorem V], it is shown that d is submodular.

Corollary 3. *Let $P_{X_{[1:n]}}$ be any joint probability distribution and $Q_{X_{[1:n]}}$ be a product probability distribution on $\mathcal{X}^n$, and $d(F) = -D(P_{X_F} \| Q_{X_F})$, $F \subseteq [1:n]$. Let $\gamma : \mathcal{F} \to \mathbb{R}_+$ and $\alpha : \mathcal{F} \to \mathbb{Q}_+$ be any fractional partition and fractional covering with respect to a family $\mathcal{F}$ of subsets of $[1:n]$. Then the following hold:*

1) $\mathrm{Gap}_{\mathrm{U}}(d, \mathcal{F}, \gamma) = 0$ *if and only if $P_{X_{[1:n]}}$ is a product probability distribution.*
2) $\mathrm{Gap}_{\mathrm{U}}(d, \mathcal{F}, \alpha) = 0$ *if and only if $P_{X_{[1:n]}}$ is a product probability distribution and $P_{X_Z} = Q_{X_Z}$, for $Z = \{i \in [1:n] : \sum_{F \in \mathcal{F}: i \in F} \alpha(F) > 1\}$.*

The proof of Corollary 3 is given in Appendix G.

C. Implications for Matroid Rank Function

Definition 3 (Matroid and Rank Funciton [1])**.** *A set system $(\mathcal{E}, \mathcal{I})$ where $\mathcal{I} \subseteq 2^{\mathcal{E}}$ is a matroid if*

1) *$\phi \in \mathcal{I}$.*
2) *$\forall I_2 \in \mathcal{I},\ I_1 \subseteq I_2 \implies I_1 \in \mathcal{I}$.*
3) *$\forall I_1, I_2 \in \mathcal{I}$, with $|I_1| < |I_2|$, there exists $e \in I_2 \setminus I_1$ such that $I_1 \cup \{e\} \in \mathcal{I}$.*

Given a matroid $\mathcal{M} = (\mathcal{E}, \mathcal{I})$, the rank function $r : 2^{\mathcal{E}} \to \mathbb{Z}_{\geq 0}$ is defined as $r(S) = \max_{I \in \mathcal{I}: I \subseteq S} |I|$, for all $S \subseteq \mathcal{E}$.

Some examples of a matroid are as follows:

(i) $\mathcal{M} = (\mathcal{E}, \mathcal{I})$, where $\mathcal{E}$ is the set of column indices corresponding to a matrix $\mathcal{A}$ and $\mathcal{I}$ is the set of those subsets of $\mathcal{E}$ which correspond to linearly independent columns in $\mathcal{A}$.
(ii) $\mathcal{M} = (\mathcal{E}, \mathcal{I})$, where $\mathcal{E}$ is the set of edges of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ on the set of vertices $\mathcal{V}$, and $\mathcal{I}$ is the set of subsets of edges which do not form a cycle in $\mathcal{G}$,

As the rank function of a matroid is submodular [1], we have the following.

Corollary 4. *Let $\mathcal{M} = (\mathcal{E}, \mathcal{I})$ be any matroid with the rank function $r : 2^{\mathcal{E}} \to \mathbb{Z}_{\geq 0}$. Let $\gamma : \mathcal{F} \to \mathbb{R}_+$ be any fractional partition with respect to any family $\mathcal{F}$ of subsets of $\mathcal{E}$. Then,*

$$\sum_{F \in \mathcal{F}} \gamma(F) r(F) = r(\mathcal{E})$$

if and only if $\mathcal{I} = 2^A$, for some $A \subseteq \mathcal{E}$.

A notable example of a matroid for which Corollary 4 holds is the free matroid, defined as $\mathcal{M} = (\mathcal{E}, 2^{\mathcal{E}})$ [1]. The proof of Corollary 4 is given in Appendix H.

V. Applications

A. A New Multivariate Mutual Information

We propose a new multivariate mutual information, which is particularly motivated by part 1) of Corollary 2.

Definition 4 ($(\mathcal{F}, \gamma)$-Mutual Information)**.** *Given a family $\mathcal{F}$ of subsets of $[1:n]$, a fractional partition $\gamma : \mathcal{F} \to \mathbb{R}_+$, and jointly distributed random variables $X_1, X_2, \ldots, X_n$, the $(\mathcal{F}, \gamma)$-mutual information (denoted by $(\mathcal{F}, \gamma)$-MI) of $X_1, X_2, \ldots, X_n$ is defined as*

$$\begin{aligned} (\mathcal{F}, \gamma)\text{-}MI(X_1; \cdots; X_n) &= \mathrm{Gap}_{\mathrm{U}}(\mathsf{e}, \mathcal{F}, \gamma) \\ &= \sum_{F \in \mathcal{F}} \gamma(F) H(X_F) - H(X_{[1:n]}). \end{aligned} \tag{10}$$

It follows from Theorems 1 and Corollary 2 that the $(\mathcal{F},\gamma)$-MI satisfies the usual criteria expected of a mutual information - namely, *non-negativity* and *independence property*. In addition, $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n)$ recovers total correlation [13] $\mathrm{TC}(X_1;\cdots;X_n) = \sum_{i=1}^n H(X_i) - H(X_{[1:n]})$, dual total correlation [14] $\mathrm{DTC}(X_1;\cdots;X_n) = H(X_{[1:n]}) - \sum_{i=1}^n H(X_i|X_{[1:n]\setminus\{i\}})$, and shared information [15], [16] $\mathrm{SI}(X_1;\cdots;X_n) = H(X_{[1:n]}) - \max_{\gamma:\mathcal{B}\to\mathbb{R}_+} \sum_{F\in\mathcal{B}} \gamma(F)H(X_F|X_{F^c})$, where $\mathcal{B} = 2^{[1:n]}\setminus\{\phi,[1:n]\}$ and γ denotes a fractional partition with respect to $\mathcal{B}$, as special cases. These are explicitly stated in the following proposition.

Proposition 1. *1) $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n) \geq 0$ with equality if and only if $X_1,\ldots,X_n$ are mutually independent.*
2) If $\mathcal{F} = \{\!\{\{i\} : i\in[1:n]\}\!\}$ and $\gamma(F)=1,\ \forall F\in\mathcal{F}$, then $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n) = \mathrm{TC}(X_1;\cdots;X_n)$.
3) If $\mathcal{F} = \{\!\{[1:n]\setminus\{i\} : i\in[1:n]\}\!\}$ and $\gamma(F) = 1/(n-1),\ \forall F\in\mathcal{F}$, then $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n) = \frac{1}{n-1}\mathrm{DTC}(X_1;\cdots;X_n)$.
4) For $\mathcal{F} = 2^{[1:n]}\setminus\{\phi,[1:n]\}$,

$$\min_{\gamma:\mathcal{F}\to\mathbb{R}_+} \frac{(\mathcal{F},\gamma)\text{-}MI(X_1;\cdots;X_n)}{w(\gamma)-1} = \mathrm{SI}(X_1;\cdots;X_n). \quad (11)$$

A detailed proof of Proposition 1 is given in Appendix I. The reader might wonder why only the upper gap, $\mathrm{Gap_U}$, is considered as the mutual information measure, rather than the lower gap, $\mathrm{Gap_L}$. Due to the one-one correspondence between the upper and lower gaps, achieved through appropriately chosen $\mathcal{F}$ and γ via duality (Theorem 2), it suffices to focus on either of the gaps. We now present some properties of $(\mathcal{F},\gamma)$-MI.

Theorem 5. *1) $\max_{\mathcal{F},\gamma}(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n) = \mathrm{TC}(X_1;\cdots;X_n)$.*
2) For any random variables $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$, we have

$$\begin{aligned}&(\mathcal{F},\gamma)\text{-}MI(Y_1;\cdots;Y_n)\\ &\leq (\mathcal{F},\gamma)\text{-}MI(X_1;\cdots;X_n) + \sum_{i=1}^n H(Y_i|X_i). \quad (12)\end{aligned}$$

3) Given a family $\mathcal{F}$ and a fractional partition $\gamma:\mathcal{F}\to\mathbb{R}_+$, let $\tilde{\mathcal{F}} = \{\!\{F\cap[1:n-1] : F\in\mathcal{F}\}\!\}$, and for each $\tilde{F}\in\tilde{\mathcal{F}}$, let $\tilde{\gamma}(\tilde{F}) = \gamma(F)$, where $F\in\mathcal{F}$ is the set corresponding to $\tilde{F}$. Then the following hold.

$$\begin{aligned}&(\mathcal{F},\gamma)\text{-}MI(X_1;\cdots;X_n)\\ &= (\tilde{\mathcal{F}},\tilde{\gamma})\text{-}MI(X_1;\cdots;X_{n-1})\\ &+ \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F) I(X_n; X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F}), \quad (13)\end{aligned}$$

and

$$\begin{aligned}&(\tilde{\mathcal{F}},\tilde{\gamma})\text{-}MI(X_1;\cdots;X_{n-1})\\ &\leq (\mathcal{F},\gamma)\text{-}MI(X_1;\cdots;X_n)\\ &\leq (\tilde{\mathcal{F}},\tilde{\gamma})\text{-}MI(X_1;\cdots;X_{n-1}) + I(X_n;X_{[1:n-1]}). \quad (14)\end{aligned}$$

4) $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n)$ is symmetric[4] if and only if it is of the form $\sum_{i=1}^{n-1}\gamma_i\left(\sum_{|F|=i} H(X_F)\right) - H(X_{[1:n]})$, where $\sum_{i=1}^{n-1}\gamma_i\binom{n-1}{i-1} = 1$.

Remark 6. Property 1) in Theorem 5 generalizes Watanabe's observation [13], which states that $\max_{\mathcal{P}}\sum_{P\in\mathcal{P}} H(X_P) - H(X_{[1:n]}) = \mathrm{TC}(X_1;\cdots;X_n)$, where $\mathcal{P}$ denotes a partition of $[1:n]$, to fractional partitions defined via $\mathcal{F}$ and γ. Special cases of properties 2) and 3) for TC and DTC appear in [22, Lemmas 4.5, 4.8 and 4.9], [23, Equation (13)].

Property 2) states that the mutual information among $Y_1,\ldots,Y_n$ cannot exceed the mutual information among $X_1,\ldots,X_n$, plus the equivocation of Y_i given X_i across all $i\in[1:n]$. In particular, if X_i almost surely determines Y_i for each $i\in[1:n]$, then $(\mathcal{F},\gamma)$-$MI(Y_1;\cdots;Y_n) \leq (\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n)$. Property 3) provides a recursive formula for $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n)$ in terms of that of fewer random variables with appropriately defined $\tilde{\mathcal{F}}$ and $\tilde{\gamma}$, and also shows that $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n)$ is non-decreasing in the number of random variables. A detailed proof of Theorem 5 is given in Appendix J.

B. Matrix Determinantal Inequalities

Using information-theoretic inequalities to prove matrix determinantal inequalities for positive semidefinite matrices has been well-studied in the literature [18]. The following proposition presents the equality conditions for the determinantal inequalities proved in [9, Corollary III] using the fractional subadditivity of differential entropy.

Proposition 2. *Let K be a positive definite $n\times n$ matrix, and let $\mathcal{F}$ be a family of subsets on $[1:n]$. For $F\in\mathcal{F}$, let $K(F)$ denote the submatrix of K corresponding to the rows and columns indexed by the elements of F. Then, using $|M|$ to denote the determinant of a matrix M, we have that, for any fractional partition γ with respect to $\mathcal{F}$,*

$$\prod_{F\in\mathcal{F}} |K(F)|^{\gamma(F)} = |K|$$

if and only if K is a diagonal matrix, i.e., $K_{ij}=0$ for all $i\neq j$.

Remark 7. Proposition 2 recovers the equality conditions of the classical determinantal inequalities of Hadamard, Szász, and Fischer [17], [18] by choosing $\mathcal{F} = \{\!\{\{i\}: i\in[1:n]\}\!\}$ with $\gamma(F)=1,\ \forall F\in\mathcal{F}$; $\mathcal{F} = \{\!\{[1:n]\setminus\{i\} : i\in[1:n]\}\!\}$ with $\gamma(F) = \frac{1}{n-1},\ \forall F\in\mathcal{F}$; $\mathcal{F} = \{\!\{F, F^c\}\!\}$, for any arbitrary $F\subset[1:n]$, with $\gamma(F)=\gamma(F^c)=1$, respectively.

The proof of Proposition 2 is given in Appendix K.

[4] $(\mathcal{F},\gamma)$-$MI(X_1;\cdots;X_n)$ is said to be symmetric if its value is invariant for every permutation among $X_1, X_2,\ldots,X_n$.

REFERENCES

[1] S. Fujishige, *Submodular functions and optimization*. Elsevier, 2005.

[2] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *AAAI*, vol. 7, 2007, pp. 1650–1654.

[3] M. Conforti and G. Cornuéjols, "Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the rado-edmonds theorem," *Discrete Applied Mathematics*, vol. 7, no. 3, pp. 251–274, 1984.

[4] B. Lehmann, D. Lehmann, and N. Nisan, "Combinatorial auctions with decreasing marginal utilities," in *Proceedings of the 3rd ACM conference on Electronic Commerce*, 2001, pp. 18–28.

[5] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.

[6] A. C. Van Enter, R. Fernández, and A. D. Sokal, "Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of gibbsian theory," *Journal of Statistical Physics*, vol. 72, pp. 879–1167, 1993.

[7] J. M. Ollagnier and D. Pinchon, "Filtre moyennant et valeurs moyennes des capacités invariantes," *Bulletin de la Société mathématique de France*, vol. 110, pp. 259–277, 1982.

[8] U. Feige, "On maximizing welfare when utility functions are subadditive," in *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, ser. STOC '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 41–50.

[9] M. Madiman and P. Tetali, "Information inequalities for joint distributions, with interpretations and applications," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2699–2713, 2010.

[10] F. Chung, R. Graham, P. Frankl, and J. Shearer, "Some intersection theorems for ordered sets and graphs," *Journal of Combinatorial Theory, Series A*, vol. 43, no. 1, pp. 23–37, 1986.

[11] J. Radhakrishnan, "Entropy and counting," *Computational mathematics, modelling and algorithms*, vol. 146, 2003.

[12] D. Ellis, G. Kindler, E. Friedgut, and A. Yehudayoff, "Geometric stability via information theory," *Discrete Analysis*, 2016.

[13] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM Journal of Research and Development*, vol. 4, no. 1, pp. 66–82, 1960.

[14] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Information and Control*, vol. 36, no. 2, pp. 133–156, 1978.

[15] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3047–3061, 2004.

[16] P. Narayan, H. Tyagi *et al.*, "Multiterminal secrecy by public discussion," *Foundations and Trends® in Communications and Information Theory*, vol. 13, no. 2-3, pp. 129–275, 2016.

[17] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[18] A. Dembo, T. Cover, and J. Thomas, "Information theoretic inequalities," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1501–1518, 1991.

[19] E. R. Scheinerman and D. H. Ullman, *Fractional Graph Theory: A Rational Approach to the Theory of Graphs*. New York: Wiley, 1997.

[20] M. L. Fisher, G. L. Nemhauser, and L. A. Wolsey, *An analysis of approximations for maximizing submodular set functions—II*. Springer, 1978.

[21] S. Verdu and T. Weissman, "The information lost in erasures," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5030–5058, 2008.

[22] T. Austin, "Multi-variate correlation and mixtures of product measures," *Kybernetika*, vol. 56, no. 3, pp. 459–499, 2020.

[23] N. Ay, E. Olbrich, N. Bertschinger, and J. Jost, "A unifying framework for complexity measures of finite systems," in *Proceedings of ECCS*, vol. 6. Citeseer, 2006.

[24] M. Thomas and A. T. Joy, *Elements of information theory*. Wiley-Interscience, 2006.

Appendix A
Justification of Assumptions

When two indices always occur together in members of $\mathcal{F}$, they can be treated as a single index that embodies this pair of indices. Under this treatment, structure of $\mathcal{F}$ and the corresponding γ-values remain unchanged. This clarifies Assumption 1.

For Assumption 2, consider any family $\mathcal{F}$ and a fractional partition $\gamma : \mathcal{F} \to \mathbb{R}_+$ such that $[1:n] \in \mathcal{F}$, and let $\delta = \gamma([1:n])$. Then from Theorem 1 we get that - for any submodular function $f : 2^{[1:n]} \to \mathbb{R}$ with $f(\phi) = 0$, we have

$$f([1:n]) \leq \sum_{S \in \mathcal{F}} \gamma(S) f(S) \tag{15}$$

$$= \delta f([1:n]) + \sum_{\substack{S \in \mathcal{F}: \\ i \in S, \\ S \neq [1:n]}} \gamma(S) f(S), \tag{16}$$

which implies that

$$f([1:n]) = \frac{1}{1-\delta} \sum_{\substack{S \in \mathcal{F}: \\ i \in S, \\ S \neq [1:n]}} \gamma(S) f(S). \tag{17}$$

Let us now define a new family $\mathcal{F}' = \{\!\{ S \in \mathcal{F} : S \neq [1:n] \}\!\}$ and a fractional partition $\gamma' : \mathcal{F}' \to \mathbb{R}_+$ such that $\gamma'(S) = \frac{\gamma(S)}{1-\delta},\ \forall S \in \mathcal{F}'$. Note that it is a valid fractional partition since for each $i \in [1:n]$,

$$\sum_{\substack{S \in \mathcal{F}': \\ i \in S}} \gamma'(S) = \sum_{\substack{S \in \mathcal{F}: \\ i \in S, \\ S \neq [1:n]}} \frac{\gamma(S)}{1-\delta} \tag{18}$$

$$= \frac{1}{1-\delta} \sum_{\substack{S \in \mathcal{F}: \\ i \in S, \\ S \neq [1:n]}} \gamma(S) \tag{19}$$

$$= \frac{1}{1-\delta} \left(\sum_{\substack{S \in \mathcal{F}: \\ i \in S}} \gamma(S) - \delta \right) \tag{20}$$

$$= \frac{1-\delta}{1-\delta} = 1, \tag{21}$$

where (20) and (21) hold because $\gamma([1:n]) = \delta$ and $\sum_{\substack{S \in \mathcal{F}: \\ i \in S}} \gamma(S) = 1$, respectively. Then, (17) can be re-expressed as follows.

$$f([1:n]) \leq \sum_{\substack{S \in \mathcal{F}': \\ i \in S}} \gamma'(S) f(S). \tag{22}$$

Thus, removing $[1:n]$ from the family $\mathcal{F}$ is permissible, as we can always readjust the fractional partition in a way that preserves the integrity of the inequality. Although we focused just on the upper bound in Theorem 1, the same reasoning also holds for the lower bound. Further, the arguments remain valid even when we consider fractional covering (or packing) instead of fractional partition.

For Assumption 3, consider any family $\mathcal{F}$ and a fractional partition $\gamma : \mathcal{F} \to \mathbb{R}_+$ such that $\exists S' \in \mathcal{F}$ such that $\gamma(S') = 0$. Here we argue that we can always remove that S' from the family $\mathcal{F}$ to get $\mathcal{F}' = \mathcal{F} \setminus S'$ and define another $\gamma' : \mathcal{F}' \to \mathbb{R}_+$ such that $\gamma'(S) = \gamma(S),\ \forall S \in \mathcal{F}'$, and the inequality concerning the fractional subadditivity remains unchanged.

To restate, the arguments for Assumption 3 go through for the lower bound in Theorem 1, and for any fractional covering (or packing).

Appendix B
Proof of Theorem 3

We first derive a lower bound for $\mathrm{Gap_U}(f, \mathcal{F}, \gamma)$ as follows.

$$\mathrm{Gap_U}(f, \mathcal{F}, \gamma)$$

$$= \sum_{S \in \mathcal{F}} \gamma(S) f(S) - f([1:n]) \tag{23}$$

$$= \sum_{\substack{S \in \mathcal{F}: \\ n \in S}} \gamma(S) f(S) + \sum_{\substack{S \in \mathcal{F}: \\ n \notin S}} \gamma(S) f(S) - f([1:n]) \tag{24}$$

$$\begin{aligned} &= \sum_{\substack{S \in \mathcal{F}: \\ n \in S}} \gamma(S) \left[f(\{n\}) + f(S \setminus \{n\} \mid \{n\}) \right] \\ &\quad + \sum_{\substack{S \in \mathcal{F}: \\ n \notin s}} \gamma(S) f(S) - f(\{n\}) - f([1:n-1] \mid \{n\}) \end{aligned} \tag{25}$$

$$\begin{aligned} &= \sum_{\substack{S \in \mathcal{F}: \\ n \in S}} \gamma(S) f(S \setminus \{n\} \mid \{n\}) + \sum_{\substack{S \in \mathcal{F}: \\ n \notin s}} \gamma(S) f(S) \\ &\quad - f([1:n-1] \mid \{n\}) \end{aligned} \tag{26}$$

$$\begin{aligned} &= \sum_{\substack{S \in \mathcal{F}: \\ n \in S}} \gamma(S) \sum_{j \in S \setminus \{n\}} f(\{j\} \mid [1:j-1] \cap S \setminus \{n\}, \{n\}) \\ &\quad + \sum_{\substack{S \in \mathcal{F}: \\ n \notin S}} \gamma(S) \sum_{j \in S} f(\{j\} \mid [1:j-1] \cap S) \\ &\quad - f([1:n-1] \mid \{n\}) \end{aligned} \tag{27}$$

$$\begin{aligned} &= \sum_{j=1}^{n-1} \sum_{\substack{S \in \mathcal{F}: \\ n \in S, \\ j \in S}} \gamma(S) f(\{j\} \mid [1:j-1] \cap S \setminus \{n\}, \{n\}) \\ &\quad + \sum_{j=1}^{n-1} \sum_{\substack{S \in \mathcal{F}: \\ n \notin S, \\ j \in S}} \gamma(S) f(\{j\} \mid [1:j-1] \cap S) \\ &\quad - f([1:n-1] \mid \{n\}) \end{aligned} \tag{28}$$

$$\begin{aligned} &\geq \sum_{j=1}^{n-1} \sum_{\substack{S \in \mathcal{F}: \\ n \in S, \\ j \in S}} \gamma(S) f(\{j\} \mid [1:j-1], \{n\}) \\ &\quad + \sum_{j=1}^{n-1} \sum_{\substack{S \in \mathcal{F}: \\ n \notin S, \\ j \in S}} \gamma(S) f(\{j\} \mid [1:j-1]) \\ &\quad - f([1:n-1] \mid \{n\}) \end{aligned} \tag{29}$$

$$= \sum_{j=1}^{n-1} f(\{j\} \mid [1:j-1], \{n\}) \left(1 - \sum_{\substack{S\in\mathcal{F}:\\ n\notin S,\\ j\in S}} \gamma(S) \right)$$

$$+ \sum_{j=1}^{n-1} f(\{j\} \mid [1:j-1]) \left(\sum_{\substack{S\in\mathcal{F}:\\ n\notin S,\\ j\in S}} \gamma(S) \right)$$

$$- f([1:n-1] \mid \{n\}) \tag{30}$$

$$\geq \sum_{j=1}^{n-1} f(\{j\} \mid [1:j-1], \{n\})$$

$$+ \sum_{j=1}^{n-1} \sigma \left(f(\{j\} \mid [1:j-1]) - f(\{j\} \mid [1:j-1], \{n\}) \right)$$

$$- f([1:n-1] \mid \{n\}) \tag{31}$$

$$= \sigma \sum_{j=1}^{n-1} f(\{j\} \mid [1:j-1])$$

$$- \sigma \sum_{j=1}^{n-1} f(\{j\} \mid [1:j-1], \{n\}) \tag{32}$$

$$= \sigma \left[f([1:n-1]) - f([1:n] \mid \{n\}) \right] \tag{33}$$

$$= \sigma \left[f([1:n-1]) + f(\{n\}) - f([1:n]) \right]. \tag{34}$$

In the above math block, (26) holds because $\sum_{\substack{S\in\mathcal{F}:\\ n\in S}} \gamma(S) = 1$, (29) holds because $f(\{j\} \mid S_1, S_2) \leq f(\{j\} \mid S_1)$ $\forall S_1, S_2 \subseteq [1:n]$, (28) follows by interchanging the summations, (30) follows from the fact that

$$\sum_{\substack{S\in\mathcal{F}:\\ n\in S,\\ j\in S}} \gamma(S) = \sum_{\substack{S\in\mathcal{F}:\\ j\in S}} \gamma(S) - \sum_{\substack{S\in\mathcal{F}:\\ n\notin S,\\ j\in S}} \gamma(S) \tag{35}$$

$$= 1 - \sum_{\substack{S\in\mathcal{F}:\\ n\notin S,\\ j\in S}} \gamma(S), \tag{36}$$

(31) follows from the definition of $\sigma = \min_{\substack{i,j\in[1:n]:\\ i\neq j}} \sum_{\substack{S\in\mathcal{F}:\\ i\in S, j\notin S}} \gamma(S)$, and (27), (32), (33), (34) all follow from the chain rule. By putting together (34) with the given condition that $\mathrm{Gap_U}(f, \mathcal{F}, \gamma) \leq \varepsilon$, we get

$$f(n) + f([1:n-1]) - f([1:n]) \leq \frac{\varepsilon}{\sigma}.$$

Reworking the proof as described above with an arbitrary $i \in [1:n]$ instead of n by considering an arbitrary permutation of $[1:n]$ that maps n to i, we obtain the following.

$$f(\{i\}) + f([1:n] \setminus \{i\}) - f([1:n]) \leq \frac{\varepsilon}{\sigma}, \forall i \in [1:n]. \tag{37}$$

We now argue that $\sigma > 0$. It suffices to prove the following claim as $\gamma(S) > 0$, $\forall S \in \mathcal{F}$ by our *initial assumptions*.

Claim 1. *For $i, j \in [1:n]$ such that $i \neq j$, $\exists S \in \mathcal{F}$ such that $i \in S, j \notin S$.*

Proof of Claim 1. Assume, for the sake of contradiction, that $\forall S \in \mathcal{F}$, $i \in S$ implies $j \in S$.

Case (i): i and j always occur together

Then, they should have been clubbed together, according to our *initial assumptions*.

Case (ii): i always occurs with j, and j occurs separately. i.e. $|\{\!\{S' \in \mathcal{F} : j \in S'\}\!\}| > |\{\!\{S' \in \mathcal{F} : i \in S'\}\!\}|$.

Then

$$\sum_{\substack{S'\in\mathcal{F}:\\ j\in S'}} \gamma(S') = \sum_{\substack{S'\in\mathcal{F}:\\ i\in S'\\ j\in S'}} \gamma(S') + \sum_{\substack{S'\in\mathcal{F}:\\ j\in S'\\ i\notin S'}} \gamma(S') \tag{38}$$

$$= 1 + \sum_{\substack{S'\in\mathcal{F}:\\ j\in S'\\ i\notin S'}} \gamma(S') \tag{39}$$

$$> 1, \tag{40}$$

where (39) follows from the definition of γ and the fact that i always appears with j, and (40) follows because j occurs separately and $\gamma(S') > 0, \forall S' \in \mathcal{F}$. Thus, we arrive at a contradiction. □

We now prove that (37) continues to hold when $\mathrm{Gap_L}(f, \mathcal{F}, \gamma) \leq \varepsilon$, instead of $\mathrm{Gap_U}(f, \mathcal{F}, \gamma) \leq \varepsilon$. Using duality (Theorem 2), we have

$$\mathrm{Gap_U}(f, \bar{\mathcal{F}}, \bar{\gamma}) = \frac{w(\bar{\gamma})}{w(\gamma)} \mathrm{Gap_L}(f, \mathcal{F}, \gamma) \tag{41}$$

$$\leq \frac{\epsilon}{w(\gamma) - 1}, \tag{42}$$

where (42) follows because $\mathrm{Gap_L}(f, \mathcal{F}, \gamma) \leq \varepsilon$ and

$$w(\bar{\gamma}) = \sum_{S^c \in \bar{\mathcal{F}}} \bar{\gamma}(S^c) \tag{43}$$

$$= \sum_{S^c \in \bar{\mathcal{F}}} \frac{\gamma(S)}{w(\gamma) - 1} \tag{44}$$

$$= \frac{1}{w(\gamma) - 1} \sum_{S\in\mathcal{F}} \gamma(S) \tag{45}$$

$$= \frac{w(\gamma)}{w(\gamma) - 1}. \tag{46}$$

Applying our conclusion for the upper bound (i.e., (37)) to (42), we get, $\forall i \in [1:n]$,

$$f(\{i\}) + f([1:n] \setminus \{i\}) - f([1:n]) \leq \frac{\varepsilon}{\sigma'(w(\gamma) - 1)} \tag{47}$$

$$= \frac{\varepsilon}{\sigma}, \tag{48}$$

where (48) follows because

$$\sigma' = \min_{\substack{i,j\in[1:n]:\\ i\neq j}} \sum_{\substack{S^{c}\in\bar{\mathcal{F}}:\\ i\in S^{c}, j\notin S^{c}}} \bar{\gamma}(S^{c})$$
$$= \min_{\substack{i,j\in[1:n]:\\ i\neq j}} \sum_{\substack{S\in\mathcal{F}:\\ i\notin S, j\in S}} \frac{\gamma(S)}{w(\gamma)-1}$$
$$= \frac{1}{w(\gamma)-1} \min_{\substack{i,j\in[1:n]:\\ i\neq j}} \sum_{\substack{S\in\mathcal{F}:\\ i\in S, j\notin S}} \gamma(S)$$
$$= \frac{\sigma}{w(\gamma)-1}.$$

This completes the proof of Theorem 3.

Details Omitted from Remark 1: Firstly, it is important to mention that the duality, while stated for γ in Theorem 2, also holds for α (fractional covering) and β (fractional packing), as discussed in [9, Discussion after Theorem 4]. Remark 1 states that Theorem 3 can be extended to incorporate fractional covering and packing in place of fractional partition. Formally, we get the following corollary.

Corollary 5. *Let $f : 2^{[1:n]} \to \mathbb{R}$ be any submodular function with $f(\phi) = 0$. Let $\alpha : \mathcal{F} \to \mathbb{R}_+$ be any fractional covering with respect to a family $\mathcal{F}$ of subsets of $[1 : n]$ and $\bar{\alpha}$ be the dual fractional packing corresponding to α. For any $\varepsilon > 0$, the following holds:*

If $\sigma > 0$ and, $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \bar{\alpha}) \leq \varepsilon$ or $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \alpha) \leq \varepsilon$, then

$$f(\{i\}) + f([1 : n] \setminus \{i\}) - f([1 : n]) \leq \frac{\varepsilon}{\sigma}, \ \forall i \in [1 : n], \tag{49}$$

where $\sigma = \min_{\substack{i,j\in[1:n]:\\ i\neq j}} \sum_{\substack{S\in\mathcal{F}:\\ i\in S, j\notin S}} \alpha(S)$.

Proof: This proof closely follows that of Theorem 3, starting with a lower bound on $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \alpha)$. The only place where this proof differs from the earlier one is (26). Instead of equality, we get an inequality as $\sum_{S\in\mathcal{F}: n\in S} \alpha(S) \geq 1$ and f is non-negative. Consequently, we arrive at the same conclusion (37) for $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \alpha)$ as well.

The proof for $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \bar{\alpha})$ follows exactly the same steps as the proof for $\mathrm{Gap}_{\mathrm{L}}(f, \mathcal{F}, \gamma)$ in Theorem 3, using duality and the fact that the dual of a fractional packing is a fractional covering (Interested reader can refer to [9, After Definition VI] for details).

APPENDIX C
PROOF OF THEOREM 4

Proof of Part 1). Let $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma)$ be equal to 0. Substituting $\varepsilon = 0$ in Theorem 3 and noting that $\sigma > 0$, we get

$$f(\{i\}) + f([1 : n] \setminus \{i\}) - f([1 : n]) = 0, \ \forall i \in [1 : n], \tag{50}$$

which implies that

$$f(\{i\}) - f(\{i\} \mid [1 : n] \setminus \{i\}) = 0, \ \forall i \in [1 : n], \tag{51}$$

where $f(S|T)$ is as defined in Section II. Since conditioning reduces the values of a submodular function by (1), we get

$$f(\{i\}) \geq f(\{i\}|[1 : i-1]) \tag{52}$$
$$\geq f(\{i\}|[1 : n] \setminus \{i\}). \tag{53}$$

Now, from (51), we get

$$f(\{i\}) = f(\{i\}|[1 : i-1]). \tag{54}$$

Using the chain rule (2) for f, we have

$$f([1 : n]) = \sum_{i=1}^{n} f(\{i\} \mid [1 : i-1]) \tag{55}$$
$$= \sum_{i=1}^{n} f(\{i\}), \tag{56}$$

where deduction to (56) uses (54). Fix an arbitrary $S \subseteq [1 : n]$. From (56), we have

$$0 = f([1 : n]) - \sum_{i=1}^{n} f(\{i\}) \tag{57}$$
$$= f(S) + f(S^{c}|S) - \sum_{i\in S} f(\{i\}) - \sum_{i\in S^{c}} f(\{i\}) \tag{58}$$
$$= \underbrace{\Big(f(S) - \sum_{i\in S} f(\{i\})\Big)}_{\leq 0} + \underbrace{\Big(f(S^{c}|S) - \sum_{i\in S^{c}} f(\{i\})\Big)}_{\leq 0}, \tag{59}$$

where (58) follows from (2). Notice that both the expressions in (59) are non-positive because $f(S) \leq \sum_{i\in S} f(\{i\})$ by submodularity, and $f(S^{c}|S) \leq f(S^{c}) \leq \sum_{i\in S^{c}} f(\{i\})$ by (1) and submodularity. Since the sum of these two expressions is zero, each of them must be equal to zero. In particular, $f(S) = \sum_{i\in S} f(\{i\})$. As $S \subseteq [1 : n]$ is arbitrary, this proves that f is modular.

For the other direction, let us suppose that f is modular. Then,

$$\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma) = \sum_{S\in\mathcal{F}} \gamma(S) f(S) - f([1 : n]) \tag{60}$$
$$= \sum_{S\in\mathcal{F}} \gamma(S) \sum_{i\in S} f(\{i\}) - f([1 : n]) \tag{61}$$
$$= \sum_{i=1}^{n} f(\{i\}) \left(\sum_{\substack{S\in\mathcal{F}:\\ i\in S}} \gamma(S) \right) - f([1 : n]) \tag{62}$$
$$= \sum_{i=1}^{n} f(\{i\}) - f([1 : n]) \tag{63}$$
$$= 0, \tag{64}$$

where (61) and (64) use the fact that f is modular and (63) follows because $\sum_{\substack{S\in\mathcal{F}:\\ i\in S}} \gamma(s) = 1$. This proves that $\mathrm{Gap}_{\mathrm{U}}(f, \mathcal{F}, \gamma) = 0$ if and only if f is modular. The assertion for $\mathrm{Gap_L}(f, \mathcal{F}, \gamma)$ can be proved using this together with duality (Theorem 2), as in the proof of Theorem 3.

Proof of Part 2). We first state and prove a claim that is essential for our proof of this part.

Claim 2. *Let $f : 2^{[1:n]} \to \mathbb{R}$ be any submodular function with $f(\phi) = 0$ such that f is non-decreasing, i.e., $f(S) \leq f(T)$ for $S \subseteq T$. Let $\mathcal{F}$ be any family of subsets of $[1 : n]$ such that $\forall i \in [1 : n]$, $|\{S \in \mathcal{F} : i \in S\}| \geq k$, then*

$$kf([1 : n]) = \sum_{S\in\mathcal{F}} f(S)$$

if and only if f is modular and $f(Z) = 0$, $\forall Z \subseteq \{i \in [1 : n] : |\{S \in \mathcal{F} : i \in S\}| > k\}$

Proof of Claim 2. Let $\mathcal{F} = \{\!\{S_1, \ldots, S_r\}\!\}$. Construct a new family $\mathcal{F}' = \{\!\{S'_1, \ldots, S'_r\}\!\}$ of subsets of $[1 : n]$ such that $S'_i \subseteq S_i$, $\forall i \in [1 : n]$ and each $i \in [1 : n]$ appears exactly in k members of $\mathcal{F}'$. Now $kf([1 : n]) = \sum_{S\in\mathcal{F}} f(S)$ implies that

$$0 = \sum_{S\in\mathcal{F}} f(S) - kf([1 : n]) \tag{65}$$

$$= \underbrace{\left(\sum_{S'\in\mathcal{F}'} f(S') - kf([1 : n])\right)}_{\geq 0} + \underbrace{\sum_{i\in[1:r]} f((S_i \setminus S'_i)|S'_i)}_{\geq 0}, \tag{66}$$

where (66) follows from the chain rule (2). Notice that both the expressions in (66) are non-negative. The non-negativity of the first expression follows from (1), while the non-negativity of the second expression follows from the fact that $f(S_i) \geq f(S'_i)$, as f is non-decreasing. Since the sum of these two expressions is zero, each of them must be equal to zero. The first expression being equal to zero implies that f is modular, by part 1) of Theorem 4. The second expression being equal to zero implies that, for all $i \in [1 : r]$,

$$0 = f((S_i \setminus S'_i)|S'_i) \tag{67}$$
$$= f(S_i) - f(S'_i) \tag{68}$$
$$= f(S_i \setminus S'_i) \tag{69}$$
$$= \sum_{j\in S_i\setminus S'_i} f(\{j\}), \tag{70}$$

where (68) follows from the chain rule (2), and (69) and (70) follow from f being modular. From this we get that

$$f(\{j\}) = 0, \tag{71}$$

$\forall j \in \bigcup_{i\in[1:r]}(S_i \setminus S'_i) = \{i \in [1 : n] : |\{S \in \mathcal{F} : i \in S\}| > k\}$. Thus, by modularity of f, we have $f(Z) = 0, \forall Z \subseteq \{i \in [1 : n] : |\{S \in \mathcal{F} : i \in S\}| > k\}$. □

We are now ready to prove part 2) of Theorem 4. Suppose $\mathrm{Gap_U}(f, \mathcal{F}, \alpha) = 0$. Let $\mathcal{F} = \{\!\{S_1, \ldots, S_r\}\!\}$. Let $\alpha(F_i) = \frac{p_i}{q_i}$ as $\alpha(F_i)$ is rational, for $i \in [1 : r]$. Denoting $L = \mathbf{lcm}(q_1, q_2, \ldots, q_r)$, i.e., the least common multiple of $q_1, q_2, \ldots, q_r$, we can express $\alpha(F_i)$ as $\frac{a_i}{L}$, for some $a_i \in \mathbb{N}$, $i \in [1 : r]$. Note that

$$\sum_{S\in\mathcal{F}} \alpha(S)f(S) = \sum_{S\in\mathcal{F}} \frac{a_i}{L} f(S) \tag{72}$$
$$= \frac{1}{L} \sum_{S\in\mathcal{F}} \sum_{j=1}^{a_i} f(S) \tag{73}$$
$$= \frac{1}{L} \sum_{S\in\mathcal{F}'} f(S), \tag{74}$$

where

$$\mathcal{F}' = \{\!\{\underbrace{S_1, \ldots, S_1}_{a_1 \text{ times}}, \underbrace{S_2, \ldots, S_2}_{a_2 \text{ times}}, \ldots, \underbrace{S_r, \ldots, S_r}_{a_r \text{ times}}\}\!\}. \tag{75}$$

Also, since

$$\sum_{S\in\mathcal{F}:i\in S} \alpha(S) = \sum_{S\in\mathcal{F}:i\in S} \frac{a_i}{L} \tag{76}$$
$$= \sum_{S\in\mathcal{F}:i\in S} \sum_{j=1}^{a_i} \frac{1}{L} \tag{77}$$
$$= \sum_{S\in\mathcal{F}':i\in S} \frac{1}{L} \tag{78}$$
$$= \frac{1}{L} |\{S \in \mathcal{F}' : i \in S\}|, \tag{79}$$

we have,

$$\{i : \sum_{S\in\mathcal{F}:i\in S} \alpha(S) = 1\} = \{i : |\{S \in \mathcal{F}' : i \in S\}| = L\}, \tag{80}$$

$$\{i : \sum_{S\in\mathcal{F}:i\in S} \alpha(S) > 1\} = \{i : |\{S \in \mathcal{F}' : i \in S\}| > L\}. \tag{81}$$

Now, by putting together (74) with the fact that $\mathrm{Gap_U}(f, \mathcal{F}, \alpha) = 0$, we get that

$$Lf([1 : n]) = \sum_{S\in\mathcal{F}'} f(S), \tag{82}$$

which along with Claim 2 further implies that f is modular and satisfies $f(Z) = 0, \forall Z \subseteq \{i : \sum_{S\in\mathcal{F}:i\in S} \alpha(S) > 1\}$.

For the other direction, suppose that f is modular and satisfies $f(Z) = 0, \forall Z \subseteq \{i : \sum_{S\in\mathcal{F}:i\in S} \alpha(S) > 1\} = B$. Then, we have

$$\mathrm{Gap_U}(f, \mathcal{F}, \alpha) \tag{83}$$
$$= \sum_{S\in\mathcal{F}} \alpha(S)f(S) - f([1 : n]) \tag{84}$$
$$= \sum_{S\in\mathcal{F}} \alpha(S) \sum_{i\in S} f(\{i\}) - f([1 : n]) \tag{85}$$
$$= \sum_{i=1}^{n} f(\{i\}) \left(\sum_{\substack{S\in\mathcal{F}:\\ i\in S}} \alpha(S)\right) - f([1 : n]) \tag{86}$$
$$= \sum_{i\in B} f(\{i\}) \left(\sum_{\substack{S\in\mathcal{F}:\\ i\in S}} \alpha(S) - 1\right)$$

$$+\sum_{i\in B^{\mathrm{c}}} f(\{i\})\left(\sum_{\substack{S\in\mathcal{F}:\\ i\in S}} \alpha(S)-1\right) \tag{87}$$

$$=0, \tag{88}$$

where (88) follows from the fact that $f(\{i\})=0$, for all $i\in B$ and $\sum\limits_{S\in\mathcal{F}:i\in S}\alpha(S)=1$, for all $i\in B^{\mathrm{c}}$. This proves that $\mathrm{Gap}_{\mathrm{U}}(f,\mathcal{F},\alpha)=0$ if and only if f is modular and $f(Z)=0,\forall Z\subseteq\{i:\sum_{S\in\mathcal{F}:i\in S}\alpha(S)>1\}$. The assertion for $\mathrm{Gap}_{\mathrm{L}}(f,\mathcal{F},\beta)$ can be proved using this together with duality (Theorem 2), as in the proof of Theorem 3. Specifically, note that for every $i\in[1:n]$, we have

$$\sum_{S\in\mathcal{F}:i\in S}\beta(S)<1 \text{ if and only if } \sum_{S^{\mathrm{c}}\in\bar{\mathcal{F}}:i\in S^{\mathrm{c}}}\bar{\beta}(S^{\mathrm{c}})>1.$$

This completes the proof of Theorem 4.

Appendix D
Example Motivating a Stronger Condition on f in Part 2) of Theorem 4

Let $f:2^{[1:3]}\to\mathbb{R}$ be a set function with the following values: $f(\phi)=0,\ f(\{1\})=-100,\ f(\{2\})=0.001,\ f(\{3\})=50.0005,\ f(\{1,2\})=-100,\ f(\{2,3\})=50.0005,\ f(\{1,3\})=-50.0005,\ f(\{1,2,3\})=-100$. Consider $\mathcal{F}=\{\!\{\{1,2\},\{1,2\},\{2,3\},\{1,3\}\}\!\}$ and $\alpha(S)=1/2,\ \forall S\in\mathcal{F}$. The function f is a submodular but not modular and satisfies $\mathrm{Gap}_{\mathrm{U}}(f,\mathcal{F},\alpha)=0$. It also satisfies the condition that $f([1:j])$ is non-decreasing in j as $f(\{1\})=f(\{1,2\})=f(\{1,2,3\})=-100$. However, $f(S)>f(T)$ for some $S\subseteq T$ (specifically, for $S=\phi,$ and $T=\{1,2\}$).

Appendix E
Proof of Corollary 1

Since entropy is a submodular function, we invoke Theorem 3 with $f(S)=\mathsf{e}(S)$, $S\subseteq[1:n]$. Then $\mathrm{Gap}_{\mathrm{L}}(\mathsf{e},\mathcal{F},\gamma)\le\varepsilon$ implies that, $\forall i\in[1:n]$,

$$H(X_i)+H(X_{[1:n]\setminus\{i\}})-H(X_{[1:n]})\le\frac{\varepsilon}{\sigma}, \tag{89}$$

which further implies that

$$I(X_i;X_{[1:n]\setminus\{i\}})\le\frac{\varepsilon}{\sigma}. \tag{90}$$

Similarly, we can show the above result for $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e},\mathcal{F},\gamma)$.

Appendix F
Proof of Corollary 2

Proof of Part 1). we invoke part 1) of Theorem 4 with $f(S)=\mathsf{e}(S)$, $S\subseteq[1:n]$. This implies that $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e},\mathcal{F},\gamma)=0$ if and only if e is modular. Notice that the modularity of e is equivalent to the condition $H(X_{[1:n]})=\sum\limits_{i=1}^{n}H(X_i)$, as this equality holds if and only if $X_i,\ i\in[1:n]$ are mutually independent [24, Theorem 2.6.6].

Proof of Part 2). Since entropy is a non-decreasing submodular function, we invoke part 2) of Theorem 4 with it. This implies that $\mathrm{Gap}_{\mathrm{U}}(\mathsf{e},\mathcal{F},\alpha)=0$ if and only if e is modular and $\mathsf{e}(Z)=0$, for all $Z\subseteq\{i\in[1:n]:\sum_{F\in\mathcal{F}:i\in F}\alpha(F)>1\}$. From the arguments in the proof of part 1) of this corollary, we know that e is modular if and only if $X_i,\ i\in[1:n]$ are mutually independent. Moreover, consider an arbitrary i such that $\sum\limits_{F\in\mathcal{F}:i\in F}\alpha(F)>1$. For this i, $\mathsf{e}(\{i\})=H(X_i)=0$ which occurs if and only if X_i is a constant.

Appendix G
Proof of Corollary 3

Proof of Part 1). From Theorem 4, we have that

$$\sum_{F\in\mathcal{F}}\gamma(F)D(P_{X_F}||Q_{X_F}) \;=\; D(P_{X_{[1:n]}}||Q_{X_{[1:n]}}) \tag{91}$$

if and only if d is modular, i.e., $\forall F\subseteq[1:n]$, $D(P_{X_F}||Q_{X_F})=\sum\limits_{i\in F}D(P_{X_i}||Q_{X_i})$. Consider $F=[1:n]$, we have,

$$D(P_{X_{[1:n]}}||Q_{X_{[1:n]}})=\sum_{i=1}^{n}D(P_{X_i}||Q_{X_i}) \tag{92}$$

which implies that

$$H_P(X_{[1:n]})-\sum_{x_{[1:n]}}P_{X_{[1:n]}}(x_{[1:n]})\log Q_{X_{[1:n]}}(x_{[1:n]})$$
$$=-\sum_{i=1}^{n}H_P(X_i)-\sum_{i=1}^{n}\sum_{x_i}P_{X_i}(x_i)\log Q_{X_i}(x_i), \tag{93}$$

where $H_P(.)$ denotes entropy with respect to the distribution $P_{X_{[1:n]}}$ and which further implies that

$$H_P(X_{[1:n]})=\sum_{i=1}^{n}H_P(X_i) \tag{94}$$

because $P_{X_{[1:n]}}(x_{[1:n]})\log Q_{X_{[1:n]}}(x_{[1:n]})=\sum\limits_{i=1}^{n}\sum\limits_{x_i}P_{X_i}(x_i)\log Q_{X_i}(x_i)$ as $Q_{X_{[1:n]}}$ is a product probability distribution. Now, (94) holds if and only if $P_{X_{[1:n]}}$ is a product probability distribution [24, Theorem 2.6.6].

Proof of Part 2). Since relative entropy is a non-decreasing submodular function, we invoke part 2) of Theorem 4 with it. This implies that $\mathrm{Gap}_{\mathrm{U}}(d,\mathcal{F},\alpha)=0$ if and only if d is modular and $d(Z)=0$, for all $Z\subseteq\{i\in[1:n]:\sum_{F\in\mathcal{F}:i\in F}\alpha(F)>1\}$. From the arguments in the proof of part 1) of this corollary, we know that d is modular if and only if $P_{X_{[1:n]}}$ is a product probability distribution. Moreover, consider $Z=\{i\in[1:n]:\sum\limits_{F\in\mathcal{F}:i\in F}\alpha(F)>1\}$. We have that $d(Z)=0$ which occurs if and only if the distributions P_{X_Z} and Q_{X_Z} are equal.

Appendix H
Proof of Corollary 4

From part 1) of Theorem 4, we have that

$$\sum_{F\in\mathcal{F}}\gamma(F)r(F)=r(\mathcal{E}) \tag{95}$$

if and only if r is modular i.e. $\forall F\subseteq\mathcal{E},\ r(F)=\sum_{i\in F}r(\{i\})$. Let $\mathcal{M}=(\mathcal{E},\mathcal{I})$ be any matroid with the modular rank function r. We show that $\mathcal{I} = 2^{B^c}$, where $B = \{j : \{j\} \cap I = \phi,\ \forall I \in \mathcal{I}\}$. We first prove that if $F \subseteq B^c$, then $F \in \mathcal{I}$. Note that for any arbitrary $F \subseteq \mathcal{E}$, we have

$$r(F \cap B^c) = \sum_{\substack{i \in F, \\ i \notin B}} r(i) \tag{96}$$

$$= \sum_{\substack{i \in F, \\ i \notin B}} 1 \tag{97}$$

$$= |F \cap B^c|, \tag{98}$$

where (98) follows because $\forall i \in B^c$, $r(\{i\}) = 1$ from the definitions of matroid and rank function. Now (98) implies that $F \cap B^c \in \mathcal{I}, \forall F \subseteq \mathcal{E}$ from the property of matroid that $r(T) = |T| \implies T \in \mathcal{I}$ for any $T \subseteq \mathcal{E}$. This further implies that $F \in \mathcal{I},\ \forall F \subseteq B^c$.

We conclude by showing that if $F \in \mathcal{I}$, then $F \subseteq B^c$. Assume $F \in \mathcal{I}$. From the definition of matroid rank function, we have that $r(F) = |F|$. Also, we have

$$|F| = r(F) \tag{99}$$

$$= \sum_{\substack{i \in F, \\ i \notin B}} r(i) + \sum_{\substack{i \in F, \\ i \in B}} r(i) \tag{100}$$

$$= |F \cap B^c|, \tag{101}$$

where (101) follows from the fact that $r(i) = 0,\ \forall i \in B$. Now (101) implies that $F \subseteq B^c$. This proves that $\mathcal{I} = 2^{B^c}$.

Appendix I
Proof of Proposition 1

Proof of Part 1). The non-negativity follows directly from Theorem 1, noting that entropy is a submodular function. From Part 1) of Corollary 2, we have $(\mathcal{F}, \gamma)\text{-}MI(X_1, \cdots, X_n) = 0$ if and only if $X_1, \ldots, X_n$ are mutually independent.

Proof of Part 2). For $\mathcal{F} = \{\!\{\{i\} : i \in [1:n]\}\!\}$ and $\gamma(F) = 1,\ \forall F \in \mathcal{F}$, we have

$$\sum_{F \in \mathcal{F}} \gamma(F) H(X_F) - H(X_{[1:n]}) = \sum_{i=1}^{n} H(X_i) - H(X_{[1:n]})$$
$$= \mathrm{TC}(X_1; \cdots; X_n).$$

Proof of Part 3). For $\mathcal{F} = \{\!\{[1:n] \setminus \{i\} : i \in [1:n]\}\!\}$ and $\gamma(F) = 1/(n-1),\ \forall F \in \mathcal{F}$, we have

$$\sum_{F \in \mathcal{F}} \gamma(F) H(X_F) - H(X_{[1:n]}) \tag{102}$$

$$= \sum_{i=1}^{n} \frac{1}{n-1} H(X_{[1:n]\setminus\{i\}}) - H(X_{[1:n]}) \tag{103}$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \big(H(X_{[1:n]}) - H(X_{\{i\}} | X_{[1:n]\setminus\{i\}})\big) - H(X_{[1:n]}) \tag{104}$$

$$= \frac{1}{n-1} \big(H(X_{[1:n]}) - H(X_{\{i\}} | X_{[1:n]\setminus\{i\}})\big) \tag{105}$$

$$= \frac{1}{n-1} \mathrm{DTC}(X_1; \cdots; X_n), \tag{106}$$

where (104) follows from chain rule for entropy.

Proof of Part 4).

$$\min_{\gamma : \mathcal{F} \to \mathbb{R}_+} \frac{(\mathcal{F}, \gamma)\text{-}MI(X_1; \cdots; X_n)}{w(\gamma) - 1} \tag{107}$$

$$= \min_{\gamma : \mathcal{F} \to \mathbb{R}_+} \mathrm{Gap_L}(\mathsf{e}, \mathcal{F}, \bar{\gamma}) \tag{108}$$

$$= \min_{\gamma : \mathcal{F} \to \mathbb{R}_+} H(X_{[1:n]}) - \sum_{F \in \mathcal{F}} \bar{\gamma}(F) H(X_F | X_{F^c}) \tag{109}$$

$$= \min_{\gamma : \mathcal{F} \to \mathbb{R}_+} H(X_{[1:n]}) - \sum_{F \in \mathcal{F}} \gamma(F) H(X_F | X_{F^c}) \tag{110}$$

$$= H(X_{[1:n]}) - \max_{\gamma : \mathcal{F} \to \mathbb{R}_+} \sum_{F \in \mathcal{F}} \gamma(F) H(X_F | X_{F^c}) \tag{111}$$

$$= \mathrm{SI}(X_1; \cdots; X_n), \tag{112}$$

where (108) follows from duality and the fact that $\bar{\mathcal{F}} = \mathcal{F}$ for $\mathcal{F} = 2^{[1:n]} \setminus \{\phi, [1:n]\}$, and (110) follows from one-to-one correspondence between the fractional partitions γ and $\bar{\gamma}$, i.e.,

$$\bar{\gamma}(F^c) = \frac{\gamma(F)}{w(\gamma) - 1},$$
$$\text{and} \quad \gamma(F) = \frac{\bar{\gamma}(F^c)}{w(\bar{\gamma}) - 1}.$$

Appendix J
Proof of Theorem 5

We first present the proof of Property 1). Note that

$$(\mathcal{F}, \gamma)\text{-}MI(X_1, \cdots, X_n)$$

$$= \sum_{F \in \mathcal{F}} \gamma(F) H(X_F) - H(X_{[1:n]}) \tag{113}$$

$$\leq \sum_{F \in \mathcal{F}} \gamma(F) \sum_{i \in F} H(X_i) - \sum_{i=1}^{n} H(X_i | X_{[1:i-1]}) \tag{114}$$

$$= \sum_{i=1}^{n} H(X_i) \left(\sum_{\substack{F \in \mathcal{F}: \\ i \in F}} \gamma(F) \right) - \sum_{i=1}^{n} H(X_i | X_{[1:i-1]}) \tag{115}$$

$$= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i | X_{[1:i-1]}) \tag{116}$$

$$= \sum_{i=1}^{n} H(X_i) - H(X_{[1:n]}) \tag{117}$$

$$= \mathrm{TC}(X_1, \cdots, X_n), \tag{118}$$

where (114) follows from the subadditivity of entropy, (115) follows from the interchange of summations, (116) follows from the fact that $\sum_{\substack{F \in \mathcal{F}: \\ i \in F}} \gamma(F) = 1$ and the chain rule for entropy. Finally, this bound is achieved when $\mathcal{F} = \{\!\{\{i\} : i \in [1:n]\}\!\}$ and $\gamma(F) = 1,\ \forall F \in \mathcal{F}$.

We now present the proof of Property 3) and we use it to prove Property 2) later. Let $\mathcal{F} = \{\!\{F_1, \ldots, F_p\}\!\}$ and correspondingly let $\tilde{\mathcal{F}} = \{\!\{F'_1, \ldots, F'_p\}\!\}$. From the statement of

Property 3) we also get that $\gamma(F_i) = \tilde{\gamma}(F_i')$ for all $i \in [1:p]$. Then,

$$(\mathcal{F},\gamma)\text{-}MI(X_1,\cdots,X_n) - (\tilde{\mathcal{F}},\tilde{\gamma})\text{-}MI(X_1,\cdots,X_{n-1}) \tag{119}$$

$$= \sum_{F\in\mathcal{F}} \gamma(F)H(X_F) - H(X_{[1:n]}) - \sum_{F\in\tilde{\mathcal{F}}} \tilde{\gamma}(\tilde{F})H(X_{\tilde{F}}) + H(X_{[1:n-1]}) \tag{120}$$

$$= \sum_{i=1}^{p} \left[\gamma(F_i)H(X_{F_i}) - \tilde{\gamma}(\tilde{F}_i)H(X_{\tilde{F}_i})\right] - H(X_n|X_{[1:n-1]}) \tag{121}$$

$$= \sum_{i=1}^{p} \gamma(F_i)\left[H(X_{F_i}) - H(X_{\tilde{F}_i})\right] - H(X_n|X_{[1:n-1]}) \tag{122}$$

$$= \sum_{i=1}^{p} \gamma(F_i)H(X_{F_i\setminus\tilde{F}_i}|X_{\tilde{F}_i}) - H(X_n|X_{[1:n-1]}) \tag{123}$$

$$= \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)H(X_n|X_{F\setminus\{n\}}) - H(X_n|X_{[1:n-1]}) \tag{124}$$

$$= \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)H(X_n|X_{F\setminus\{n\}}) - \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)H(X_n|X_{[1:n-1]}) \tag{125}$$

$$= \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)\left[H(X_n|X_{F\setminus\{n\}}) - H(X_n|X_{[1:n-1]})\right] \tag{126}$$

$$= \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)I(X_n;X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F}), \tag{127}$$

where (121) and (123) follow from chain rule for entropy, and (124) follows from the fact that $F_i\setminus\tilde{F}_i$ can only have two possible values i.e. ϕ (when $F_i = F_i'$) or $\{n\}$ (when $F_i \neq F_i'$) as $\tilde{\mathcal{F}}$ is constructed from $\mathcal{F}$ by truncating n in all $F \in \mathcal{F}$ where $n \in F$. We use the fact that $\sum_{F\in\mathcal{F}:n\in F}\gamma(F) = 1$ to write (125). (127) follows from the fact that $F \setminus \{n\} = [1:n-1]\cap F$ when $n \in F$. Finally, we can observe that if we lower bound $H(X_n|X_{F\setminus\{n\}})$ in the first term of each summand in (126) by $H(X_n|X_{[1:n-1]})$ then we would get that $(\mathcal{F},\gamma)\text{-}MI(X_1,\cdots,X_n) \geq (\tilde{\mathcal{F}},\tilde{\gamma})\text{-}MI(X_1,\cdots,X_{n-1})$. On the contrary, if we were to upper bound $H(X_n|X_{F\setminus\{n\}})$ in the first term of each summand in (126) by $H(X_n)$, then we would get that $(\mathcal{F},\gamma)\text{-}MI(X_1,\cdots,X_n) \leq (\tilde{\mathcal{F}},\tilde{\gamma})\text{-}MI(X_1,\cdots,X_{n-1}) + I(X_n;X_{[1:n-1]})$.

With this information, we will now prove Property 2). We first prove the inequality for the case when $Y_i = X_i,\ \forall i \in [1:n-1]$. Proof for the general inequality then follows from induction and symmetry arguments.

$$(\mathcal{F},\gamma)\text{-}MI(X_1,\cdots,X_{n-1},Y_n) - (\mathcal{F},\gamma)\text{-}MI(X_1,\cdots,X_{n-1},X_n) \tag{128}$$

$$= \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)I(Y_n;X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F}) - \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)I(X_n;X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F}) \tag{129}$$

$$\leq \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)I(Y_n,X_n;X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F}) - \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)I(X_n;X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F}) \tag{130}$$

$$= \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)I(Y_n;X_{[1:n-1]\cap F^c}|X_{[1:n-1]\cap F},X_n) \tag{131}$$

$$\leq \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)H(Y_n|X_{[1:n-1]\cap F},X_n) \tag{132}$$

$$\leq \sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F)H(Y_n|X_n) \tag{133}$$

$$= H(Y_n|X_n), \tag{134}$$

where (129) follows from (127), and (130) and (131) follow from the application of chain rule for mutual information, (133) follows from the fact that conditioning only reduces entropy, and (134) uses the fact that $\sum_{\substack{F\in\mathcal{F}:\\ n\in F}} \gamma(F) = 1$.

Proof for Property 4) is obtained by invoking the result of Han [14, Lemma 3.1], on the symmetry property of entropy vectors in the entropy vector space and by noticing that $\gamma_n = -1$.

Appendix K
Proof of Corollary 2

Let $X_{[1:n]}$ have a multivariate Gaussian distribution with mean 0 and covariance matrix K, then

$$h(X_{[1:n]}) = \frac{1}{2}\log\left[(2\pi e)^n|K|\right], \tag{135}$$

$$\text{and} \quad h(X_F) = \frac{1}{2}\log\left[(2\pi e)^{|F|}|K(F)|\right]. \tag{136}$$

Now, [9, Proof of Corollary III], in conjunction with part 1) of Corollary 2 (applied to differential entropy) implies that $\prod_{F\in\mathcal{F}} |K(F)|^{\gamma(F)} = |K|$ if and only if $X_i,\ i \in [1:n]$ are mutually independent, which is equivalent to K being a diagonal matrix, i.e., $K_{ij} = 0$, for all $i \neq j$.