On Characterization of Entropy Function via Information Inequalities

Zhen Zhang, *Senior Member, IEEE*, and Raymond W. Yeung, *Senior Member, IEEE*

***Abstract*— Given n discrete random variables $\Omega = \{X_1, \cdots, X_n\}$, associated with any subset α of $\{1, 2, \cdots, n\}$, there is a joint entropy $H(X_\alpha)$ where $X_\alpha = \{X_i: i \in \alpha\}$. This can be viewed as a function defined on $2^{\{1, 2, \cdots, n\}}$ taking values in $[0, +\infty)$. We call this function the entropy function of Ω. The nonnegativity of the joint entropies implies that this function is nonnegative; the nonnegativity of the conditional joint entropies implies that this function is nondecreasing; and the nonnegativity of the conditional mutual informations implies that this function has the following property: for any two subsets α and β of $\{1, 2, \cdots, n\}$**

$$H_\Omega(\alpha) + H_\Omega(\beta) \geq H_\Omega(\alpha \cup \beta) + H_\Omega(\alpha \cap \beta).$$

These properties are the so-called basic information inequalities of Shannon's information measures. Do these properties fully characterize the entropy function? To make this question more precise, we view an entropy function as a $2^n - 1$-dimensional vector where the coordinates are indexed by the nonempty subsets of the ground set $\{1, 2, \cdots, n\}$. Let Γ_n be the cone in R^{2^n-1} consisting of all vectors which have these three properties when they are viewed as functions defined on $2^{\{1, 2, \cdots, n\}}$. Let Γ_n^* be the set of all $2^n - 1$-dimensional vectors which correspond to the entropy functions of some sets of n discrete random variables. The question can be restated as: is it true that for any n, $\overline{\Gamma}_n^* = \Gamma_n$? Here $\overline{\Gamma}_n^*$ stands for the closure of the set Γ_n^*. The answer is "yes" when $n = 2$ and 3 as proved in our previous work. Based on intuition, one may tend to believe that the answer should be "yes" for any n. The main discovery of this paper is a new information-theoretic inequality involving four discrete random variables which gives a negative answer to this fundamental problem in information theory: $\overline{\Gamma}_n^*$ is strictly smaller than Γ_n whenever $n > 3$. While this new inequality gives a nontrivial outer bound to the cone $\overline{\Gamma}_4^*$, an inner bound for $\overline{\Gamma}_4^*$ is also given. The inequality is also extended to any number of random variables.

***Index Terms*— Entropy, inequality, information measure, mutual information.**

I. Introduction

Let $\Omega_n = \{X_i: i = 1, \cdots, n\}$ be n jointly distributed discrete random variables. The basic Shannon's information measures associated with these random variables include all joint entropies, conditional entropies, mutual informations, and conditional mutual informations involving some of these random variables. For any subset α of $\mathcal{N}_n = \{1, \cdots, n\}$, let

$$X_\alpha = \{X_i: i \in \alpha\}. \tag{1}$$

Let X_ϕ, where ϕ is the empty set, be a random variable taking a fixed value with probability 1. Let

$$I_\Omega(\alpha, \beta|\gamma) = I(X_\alpha; X_\beta|X_\gamma) \tag{2}$$

be the conditional mutual information and let

$$H_\Omega(\alpha) = H(X_\alpha) \tag{3}$$

be the joint entropy. Sometimes, we drop the subscript Ω in these notations, when no confusion may occur. It is well known that Shannon's information measures satisfy the following inequalities.

Proposition 1: For any three subsets α, β, and γ of $\mathcal{N}_n$, any set of n jointly distributed discrete random variables $\Omega = \{X_i, i = 1, \cdots, n\}$

$$I(\alpha, \beta|\gamma) \geq 0. \tag{4}$$

We call these inequalities the basic inequalities of Shannon's information measures, or simply the basic inequalities. By means of the chain rule for conditional mutual informations, we can see that these inequalities are implied by a subset of these inequalities of the form

$$I(\{i\}, \{j\}|\gamma) \geq 0. \tag{5}$$

That is, the subset of inequalities (4) in which the cardinalities of α and β are both 1. This subset of basic information inequalities is referred to as elemental information inequalities [35].

For any set of n jointly distributed discrete random variables $\Omega = \{X_i, i = 1, \cdots, n\}$, the associated joint entropies $H_\Omega(\alpha)$ can be viewed as a function defined on $2^{\mathcal{N}_n}$

$$H_\Omega: 2^{\mathcal{N}_n} \to [0, \infty). \tag{6}$$

The goal of this paper is to study this function for all possible sets Ω of n discrete random variables.

All basic Shannon's information measures can be expressed as linear functions of the joint entropies. Actually, we have

$$\begin{aligned} I(\alpha, \beta|\gamma) = & H(\alpha \cup \gamma) + H(\beta \cup \gamma) \\ & - H(\alpha \cup \beta \cup \gamma) - H(\gamma). \end{aligned} \tag{7}$$

The basic inequalities can be interpreted as a set of inequalities for the entropy function as follows:

Manuscript received December 12, 1996; revised November 15, 1997. The work of Z. Zhang was supported in part by the National Science Foundation under Grant NCR-9502828. The work of R. W. Yeung was supported in part by The Research Grant Council of Hong Kong under Earmarked Grant CUHK 332/96E.

Z. Zhang is with the Department of Electrical Engineering–Systems, Communication Sciences Institute, University of Southern California, Los Angeles, CA 90089-2565 USA (e-mail: zzhang@milly.usc.edu).

R. W. Yeung is with the Department of Information Engineering, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China (e-mail: whyeung@ie.cuhk.edu.hk).

Publisher Item Identifier S 0018-9448(98)03630-X.

Proposition 2: For any set of n jointly distributed discrete random variables $\Omega = \{X_i, i = 1, \cdots, n\}$, the entropy function H associated with these random variables has the following properties.

- For any two subsets α and β of $\mathcal{N}_n$
$$H(\alpha \cup \beta) + H(\alpha \cap \beta) \leq H(\alpha) + H(\beta). \tag{8}$$
- $\alpha \subset \beta$ implies
$$H(\alpha) \leq H(\beta). \tag{9}$$
and
-
$$H(\phi) = 0. \tag{10}$$

Let $\mathcal{F}_n$ be the set of all functions defined on $2^{\mathcal{N}_n}$ taking values in $[0, \infty)$. Define

$$\begin{aligned}\Gamma_n \overset{\text{def}}{=} \{&F \in \mathcal{F}_n : F(\phi) = 0; \alpha \subset \beta \Rightarrow F(\alpha) \leq F(\beta);\\ &\forall \alpha, \beta \in 2^{\mathcal{N}_n}, F(\alpha) + F(\beta) \geq F(\alpha + F(\alpha \cap \beta)\}.\end{aligned} \tag{11}$$

Apparently, for any $\Omega = \{X_i : i = 1, \cdots, n\}$, $H_\Omega \in \Gamma_n$. This means that the set Γ_n characterizes some of the properties of the entropy function. A natural question to ask is whether or not this set "fully" characterizes the entropy function. To make the question more precise, we have introduced in [39] the following definitions.

Definition 1: A function $F \in \mathcal{F}_n$ is called *constructible* if and only if there exists a set of n jointly distributed discrete random variables Ω such that the joint entropy function H_Ω associated with these random variables satisfies $H_\Omega = F$. Define

$$\Gamma_n^* = \{F \in \mathcal{F}_n : F \text{ is constructible}\}. \tag{12}$$

In [39], we have seen that the structure of this set could be very complicated and we mentioned that the following concept is more manageable:

Definition 2: A function $F \in \mathcal{F}_n$ is called *asymptotically constructible* if and only if there exist a sequence of sets of n discrete random variables Ω^k for $k = 1, \cdots$ such that the joint entropy functions H_{Ω^k} associated with Ω^k satisfy $\lim_{k \to \infty} H_{\Omega^k} = F$.

Obviously, a function F is asymptotically constructible if and only if $F \in \overline{\Gamma}_n^*$, the closure of the set Γ_n^*.

In [39], we proved the following results.

Theorem 1:

$$\Gamma_2^* = \Gamma_2 \tag{13}$$

and

$$\overline{\Gamma}_3^* = \Gamma_3. \tag{14}$$

Up to this work, it was not known whether or not this result can be generalized. That is, we did not know whether for $n > 3$

$$\overline{\Gamma}_n^* = \Gamma_n. \tag{15}$$

This is a fundamental problem in information theory. In [39], we proved a conditional inequality of Shannon's information measures.

Theorem 2: For any four discrete random variables $\Omega_4 = \{X, Y, Z, U\}$

$$I(X; Y) = I(X; Y|Z) = 0 \tag{16}$$

implies

$$I(X; Y|Z, U) \leq I(Z; U|X, Y) + I(X; Y|U). \tag{17}$$

We also proved in [39] that this result implies

$$\Gamma_n^* \neq \Gamma_n. \tag{18}$$

Therefore, it lends some evidence for the following conjecture.

Conjecture 1: For $n \geq 4$

$$\overline{\Gamma}_n^* \neq \Gamma_n. \tag{19}$$

To give an affirmative answer to this problem is the goal of the current paper.

The paper is organized as follows: in the next section, we state the main results and introduce some definitions and notations; Sections III and IV are devoted to the proofs of the results; in Section V, we summarize the findings of the paper and raise some problems for future study.

Before closing this section, we would like to give a brief account of the works we found in the literature which are relevant to the subject matter of the current paper. As Shannon's information measures are the most important measures in information theory, researchers in this area have been investigating their structural properties since the 1950's. The early works on this subject have been done along various directions by Campbell [2], Hu [10], McGill [21], Watanabe [31], [32].

McGill [21] has proposed a multiple mutual information for any number of random variables, which is a generalization of Shannon's mutual information for two random variables. Properties of the multiple mutual information have been investigated in the subsequent works of Kawabata and Yeung [6], Tsujishita [30], and Yeung [37].

The work of Hu [10] was the first attempt to establish an analogy between information theory and set theory. Toward this end, he defined the following formal substitution of symbols:

$$\begin{array}{rcl} H/I & \leftrightarrow & \mu \\ , & \leftrightarrow & \cup \\ ; & \leftrightarrow & \cap \\ | & \leftrightarrow & - \end{array}$$

where μ is any set-additive function. In the above substitution, on the left are symbols in information theory, while on the right are symbols in set theory. He showed that a linear information-theoretic identity holds for all distributions if and only if the corresponding set-theoretic identity obtained via the formal substitution of symbols holds for all additive function μ. For example, the information-theoretic identity

$$H(X) + H(Y) = H(X, Y) + I(X; Y)$$

corresponds to the set-theoretic identity

$$\mu(X) + \mu(Y) = \mu(X \cup Y) + \mu(X \cap Y).$$

Hu's work was originally published in Russian, and it was not widely known in the west until it was reported in Csiszár and Körner [3].

Important progress has been made in the mid 1970's and the early 1980's, mainly by Han [7], [9]. Let us point out that any linear information expression can be expressed as a linear combination of unconditional joint entropies by repeated applications (if necessary) of the following identity:

$$I(X; Y|Z) = H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z).$$

In [7], Han proved the fundamental result that a linear combination of unconditional joint entropies is always equal to zero if and only if all the coefficients are zero. This result was also obtained independently by Csiszár and Körner [3]. Han further established a necessary and sufficient condition for a symmetrical linear information expression to be always nonnegative, and a necessary and sufficient condition for a linear information expression involving three random variables to be always nonnegative [9]. In [9], he raised the important question of what linear combinations of unconditional joint entropies are always nonnegative. In his work, Han viewed a linear combination of unconditional joint entropies as a vector space, and he developed a lattice-theoretic description of Shannon's information measures with which some notations can be greatly simplified. During this time, Fujishige [5] found that the entropy function is a polymatroid [33].

In the 1990's, Yeung [34] further developed Hu's work into an explicit set-theoretic formulation of Shannon's information measures. Specifically, he showed that Shannon's information measures uniquely define a signed measure μ^*, called the I-measure, on a properly defined field. With this formulation, Shannon's information measures can formally be viewed as a signed measure, and McGill's multiple mutual information is naturally included. As a consequence, all set-theoretic techniques can now be used for the manipulation of information expressions. Furthermore, the use of information diagrams to represent the structure of Shannon's information measures becomes justified. We note that information diagrams have previously been used informally to illustrate the structure of Shannon's information measures [1], [22], [24]. Subsequently, Kawabata and Yeung [6] studied the structure of μ^* when the random variables form a Markov chain. Recently, Yeung *et al.* [36] have extended the study in [6] to random variables forming a Markov random field.

Recently, Yeung [35] defined the set Γ_n^* of all constructible entropy functions and observed that whether an information inequality (linear or nonlinear) always holds is completely characterized by Γ_n^*. This geometrical framework enables him to develop a unified description of all information inequalities (unconstrained or constrained) which are implied by the nonnegativity of Shannon's information measures, called the basic inequalities. This gives a partial answer to the question raised by Han in [9], and it directly leads to the question of whether all information inequalities which always hold are implied by the basic inequalities for the same set of random variables, or equivalently, whether $\Gamma_n^* = \Gamma_n$ is true. In fact, the same question was raised in [26] and [34], although at that time Γ_n^* had not been defined and the intimate relation between Γ_n^* and information inequalities was not known. This question is the starting point of the work by the authors in [39] and in the current paper. With the result in [39] that $\overline{\Gamma}_n^*$ is a convex cone, answering the question raised by Han in [9] is equivalent to determining $\overline{\Gamma}_n^*$. In a recent paper [37], we have used the region Γ_n^* to study the so-called distributed source-coding problem.

As a consequence of the work in [35], a software called ITIP (Information Theoretic Inequality Prover) [38] has been developed which can verify all linear information inequalities involving a definite number of random variables that are implied by the basic inequalities for the same set of random variables.

Along another line, motivated by the study of the logic of integrity constraints from databases, researchers in the area of probabilistic reasoning have spent much effort in characterizing the compatibility of conditional independence relation among random variables. This effort was launched by a seminal paper by Dawid [4], in which he proposed four axioms as heuristical properties of conditional independence. In information-theoretic terms, these four axioms can be summarized by the following statement:

$$I(X; Y, Z|U) = 0 \Leftrightarrow I(X; Y|U) = 0$$

and

$$I(X; Z|Y, U) = 0.$$

Subsequent work on this subject has been done by Pearl and his collaborators in the 1980's, and their work is summarized in the book by Pearl [23]. Pearl conjectured that Dawid's four axioms completely characterize the conditional independence structure of any joint distribution. This conjecture, however, was refuted by the work of Studený [25]. Since then, Matúš and Studený have written a series of papers on this problem [13]–[29]. They have solved the problem for up to four random variables. It has been shown in [35] that the problem of characterizing the compatibility of conditional independence relations among random variables is a subproblem of the determination of Γ_n^*.

II. Statement of Main Results

The key result of this paper is the following theorem:

Theorem 3: For any four discrete random variables X, Y, Z, U, let

$$\Delta(Z, U|X, Y) = I(Z; U) - I(Z; U|X) - I(Z; U|Y). \quad (20)$$

Then the following inequality holds:

$$\Delta(Z, U|X, Y) \le \tfrac{1}{2}\,[I(X; Y) + I(X; ZU) + I(Z; U|X) - I(Z; U|Y)]. \quad (21)$$

Note that the right-hand side of (21) is not symmetric in X and Y, whereas the left-hand side is. Therefore, we also have the following inequality:

$$\Delta(Z, U|X, Y) \le \tfrac{1}{2}\,[I(X; Y) + I(Y; ZU) - I(Z; U|X) + I(Z; U|Y)]. \quad (22)$$

Averaging inequalities (21) and (22) gives

$$\Delta(Z, U|X, Y) \le \tfrac{1}{2} I(X; Y) + \tfrac{1}{4} [I(X; ZU) + I(Y; ZU)]. \tag{23}$$

This theorem will be proved in the next section.

Let $F \in \mathcal{F}_n$. Define for three subsets α, β, and γ of $\mathcal{N}_n$

$$I_F(\alpha; \beta|\gamma) = F(\alpha \cup \gamma) + F(\beta \cup \gamma) - F(\alpha \cup \beta \cup \gamma) - F(\gamma).$$

When γ is the empty set, we simply write $I_F(\alpha; \beta)$ in place of $I_F(\alpha; \beta|\gamma)$. Let

$$\begin{aligned}\Delta_F(i, j|k, l) = {}& I_F(\{i\}; \{j\}) - I_F(\{i\}; \{j\}|\{k\}) \\ & - I_F(\{i\}; \{j\}|\{l\}).\end{aligned} \tag{24}$$

Define

$$\begin{aligned}\tilde{\Gamma}_4 = \{F \in \Gamma_4\text{: for any permutation } \pi \text{ of } \{1, 2, 3, 4\} \\ \Delta_F(\pi(1), \pi(2)|\pi(3), \pi(4)) \\ \le \tfrac{1}{2} [I_F(\pi(3); \pi(4)) + I_F(\pi(1); \pi(2)|\pi(3)) \\ - I_F(\pi(1); \pi(2)|\pi(4)) \\ + I_F(\pi(3); \pi(1)\pi(2))]\}.\end{aligned} \tag{25}$$

(Notice that, if we replace $\pi(i)$ for $i = 1, 2, 3, 4$ by Z, U, X, Y, respectively, then the inequality in (25) is just (21).) Theorem 3 says that

$$\overline{\Gamma}^*_4 \subset \tilde{\Gamma}_4.$$

Theorem 3 implies the following result.

Theorem 4: For $n \ge 4$

$$\overline{\Gamma}^*_n \ne \Gamma_n. \tag{26}$$

Proof: Apparently, we need to prove the theorem only for $n = 4$. This will imply the conclusion of the theorem for any $n \ge 4$. Define a function F by letting

$$\begin{aligned}F(\phi) &= 0 \\ F(X) &= F(Y) = F(Z) = F(U) = 2a > 0 \\ F(X, Y) &= 4a,\ F(X, U) = F(X, Z) = F(Y, U) \\ &= F(Y, Z) = F(Z, U) = 3a \\ F(X, Y, Z) &= F(X, Y, U) = F(X, Z, U) \\ &= F(Y, Z, U) = F(X, Y, Z, U) = 4a.\end{aligned}$$

Then Theorem 4 is proved by checking that $F \in \Gamma_4$ and $F \notin \tilde{\Gamma}_4$. □

From Theorem 4, we have

$$\tilde{\Gamma}_4 \ne \Gamma_4 \qquad \tilde{\Gamma}_4 \subset \Gamma_4.$$

That is, the set $\tilde{\Gamma}_4$ is a nontrivial outer bound of the set $\overline{\Gamma}^*_4$.

Theorem 3 can be generalized to the following information inequalities for $n + 2$ random variables where $n \ge 2$.

Theorem 5: For any set of $n + 2$ discrete random variables U, Z, X_i: $i = 1, 2, \cdots, n$ and any $i \in \{1, 2, \cdots, n\}$

$$\begin{aligned}nI(U; Z) - \sum_{j=1}^{n} I(U; Z|X_j) - nI(U; Z|X_i) \\ \le I(X_i; UZ) + \sum_{j=1}^{n} H(X_j) - H(X_1 X_2 \cdots X_n).\end{aligned} \tag{27}$$

Furthermore, by averaging (27) over i, we obtain

$$\begin{aligned}nI(U; Z) - 2\sum_{j=1}^{n} I(U; Z|X_j) \\ \le \frac{1}{n} \sum_{i=1}^{n} I(X_i; UZ) + \sum_{j=1}^{n} H(X_j) - H(X_1 X_2 \cdots X_n).\end{aligned} \tag{28}$$

The proof of this theorem is omitted because it can be proved using exactly the same idea used in the proof of Theorem 3 and an inductive argument.

So far, when we study the entropy function, it is viewed as a function defined on $2^{\mathcal{N}_n}$. That is, we use the subsets of $\mathcal{N}_n$ as coordinates. $H(\alpha)$ is simply the joint entropy $H(X_\alpha)$. It is more convenient to use another coordinate system when we study the inner bound of the set $\overline{\Gamma}^*_n$. To introduce this new coordinate system, we employ the concept of atoms. The atoms are also indexed by the subsets of $\mathcal{N}_n$, or the elements of $2^{\mathcal{N}_n}$. To motivate the definitions we are going to introduce, we first check the definitions of the conditional mutual informations of more than two random variables. Let $X_1, \cdots, X_n$ be n discrete random variables, then the conditional mutual information of k random variables given X_α [37] is defined as

$$I(X_{i1}; \cdots; X_{ik}|X_\alpha) = \sum_{\gamma \subset \{i_1, \cdots, i_k\}} (-1)^{1+|\gamma|} H(X_{\gamma \cup \alpha}). \tag{29}$$

Consider an arbitrary function F in $\mathcal{F}_n$. We define a function $F[\alpha|\beta]$ for any pair of subsets α, β of the ground set $\mathcal{N}_n$ where α is nonempty. The values of the original function are denoted by $F(.)$, while for the values of the function $F[\alpha|\beta]$, we use $[.]$ to replace $(.)$ to indicate that these are different from the values of the original function.

$$F[\alpha|\beta] \overset{\text{def}}{=} \sum_{\gamma \subset \alpha} (-1)^{1+|\gamma|} F(\gamma \cup \beta). \tag{30}$$

$$F[\alpha] \overset{\text{def}}{=} F[\alpha|\alpha^c] \tag{31}$$

where α^c stands for the complement of α with respect to the ground set $\{1, 2, \cdots, n\}$. As we said, this concept is parallel to the concept of conditional mutual informations of more than two random variables. We have, for instance, when $\alpha = \{1, 2, 3\}$, $n = 5$, F is the entropy function of five random variables $X_1, \cdots, X_5$

$$F[\alpha] = I(X_1; X_2; X_3|X_4 X_5).$$

We say $F[\alpha]$ is the value of the function F at the atom α. The atoms are also indexed by the subsets of $\mathcal{N}_n$. The weight of an atom is defined as the cardinality of its index set.

The basic information inequalities can be restated under this new coordinate system as follows: *If F is the entropy function of a set of n random variables $\Omega = \{X_1, \cdots, X_n\}$, then for any subset α of $\mathcal{N}_n$ of cardinality of 2 and any subset β of $\mathcal{N}_n - \alpha$*

$$F[\alpha|\beta] \geq 0 \tag{32}$$

and for any single element set α

$$F[\alpha] \geq 0. \tag{33}$$

This includes only a subset of the basic inequalities (called the elemental inequalities in [35]). But as we mentioned before, this subset of basic inequalities implies all other basic inequalities.

We use some simplified notations: as an example, if $\alpha = \{1, 2, 3\}$, $\beta = \{4, 5, 6\}$, we write $F[1, 2, 3|4, 5, 6]$ in place of $F[\{1, 2, 3\}|\{4, 5, 6\}]$. A useful formula for the function $F[.]$ is the following lemma.

Lemma 1:

$$F[\alpha|\beta] = \sum_{\gamma \subset (\alpha \cup \beta)^c} F[\alpha \cup \gamma] \tag{34}$$

where A^c stands for the complement of the set A.

For four random variables X_1, X_2, X_3, X_4, if F is the entropy function of the four random variables, the basic inequalities are as follows: let $\{i, j, k, l\}$ be a permutation of $\{1, 2, 3, 4\}$

$$F[i, j] \geq 0$$
$$F[i, j] + F[i, j, k] \geq 0$$
$$F[i, j] + F[i, j, k] + F[i, j, l] + F[i, j, k, l] \geq 0$$

and

$$F[i] \geq 0.$$

We have from (20)

$$\Delta(X_1, X_2|X_3, X_4) = F[1, 2, 3, 4] - F[1, 2]. \tag{35}$$

We use $\Delta_F(i, j|k, l)$ in place of $\Delta(X_i, X_j|X_k, X_l)$ when F is the entropy function. By the same formula, Δ_F can be extended to any function F which may not be an entropy function, that is,

$$\Delta_F(i, j|k, l) \overset{\text{def}}{=} F[i, j, k, l] - F[i, j]. \tag{36}$$

The following is an interesting quantity useful in the study of Γ_n^*:

$$S_F(i, j|k, l) \overset{\text{def}}{=} F[i, j]+F[i, j, k]+F[i, j, l]+F[k, l]. \tag{37}$$

Let ϕ be the empty set, we have

$$\begin{aligned} S_F(i, j|k, l) &= F[i, j|\phi] - \Delta_F(k, l|i, j) \\ &= F[i, j] + F[i, j, k] + F[i, j, l] \\ &\quad + F[i, j, k, l] - \Delta_F(k, l|i, j). \end{aligned} \tag{38}$$

When F is the entropy function of four discrete random variables $X_1, X_2, X_3,$ and X_4

$$\begin{aligned} S_F(1, 2|3, 4) = &I(X_1; X_2) + I(X_3; X_4|X_1) \\ &+ I(X_3; X_4|X_2) - I(X_3; X_4). \end{aligned}$$

This quantity may be negative. This fact will be proved at the end of this section. The importance of this information quantity will be seen in Theorem 6 stated in this section. Theorem 3 can be restated as follows.

Theorem 3: For four random variables X_1, X_2, X_3, X_4, if F is the entropy function, then

$$S_F(1, 2|3, 4) + F[1, 3|4] + F[1, 4|3] + F[3, 4|1] \geq 0. \tag{39}$$

Inequalities (22) and (23) are now

$$S_F(1, 2|3, 4) + F[2, 3|4] + F[2, 4|3] + F[3, 4|2] \geq 0 \tag{40}$$

and

$$\begin{aligned} 2S_F(1, 2|3, 4) &+ F[1, 3|4] + F[1, 4|3] + F[2, 3|4] \\ &+ F[2, 4|3] + F[3, 4|1] + F[3, 4|2] \geq 0. \end{aligned} \tag{41}$$

That is, although the quantity S_F can be negative for some entropy functions F, but it is bounded from below by the maximum of the following two quantities:

$$-(F[1, 3|4] + F[1, 4|3] + F[3, 4|1])$$
$$-(F[2, 3|4] + F[2, 4|3] + F[3, 4|2]).$$

We notice that $F[i, j|k]$ is the conditional mutual information of the two random variables X_i and X_j given X_k which is always nonnegative.

Define

$$\begin{aligned} \hat{\Gamma}_4 = \{F \in \Gamma_4\text{: for any permutation } \pi \text{ of } \{1, 2, 3, 4\} \\ S_F(\pi(1), \pi(2)|\pi(3), \pi(4)) \geq 0\}. \end{aligned} \tag{42}$$

The last theorem of the paper is

Theorem 6:

$$\hat{\Gamma}_4 \subset \overline{\Gamma}_4^*. \tag{43}$$

Theorem 6 has been previously proven in [20]. The example in [22, Sec. V and Lemma 4.1] imply this result. But their proof is very abstract and [22, Lemma 4.1] is based on further references. To make this result more understandable, we give a direct proof of Theorem 6 in Section IV. This theorem provides an inner bound to the cone $\overline{\Gamma}_4^*$.

The following nontrivial example shows that this inner bound is not tight. In other words, this example shows that the information quantity S_F can be negative for some entropy functions F.

A Counterexample for the Positivity of S_F: A projective plane (for p) is a collection of $p^2 - p + 1$ subsets of $\{0, 1, 2, \cdots, p^2 - p\}$ of cardinality p such that the intersection of any pair of subsets from the collection has cardinality exactly 1 (see, for instance, [14, Appendix B]). For instance, 012, 034, 056, 135, 146, 236, 245 is a projective plane for $p = 3$ on the ground set $\{0, 1, 2, 3, 4, 5, 6\}$.

$$0123,\ 0456,\ 0789,\ 0abc,\ 147a,\ 158b,\ 169c,\ 248c,\ 259a,\ 267b,\ 349b,\ 357c,\ 368a$$

is another example of a projective plane on the ground set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c\}$ for $p = 4$. We are going to construct four discrete random variables for which the function S_F is negative. In our construction, we use a pair of projective planes of the same size satisfying an additional property that, if we take one subset from each projective plane, the intersection of the two subsets taken always has cardinality at most 2. An example for such a pair is the previous projective plane for $p = 4$ and the following projective plane of the same cardinality and on the same ground set:

$$0148,\ 025c,\ 037a,\ 069b,\ 1279,\ 156a,\ 13bc,\ 2346,\ 457b,\ 49ac,\ 28ab,\ 3589,\ 678c.$$

Let the ground set of the projective plane be $\mathcal{P}$, let $\mathcal{D} = \cup_{a \in \mathcal{P}}(a, a)$ be the diagonal of $\mathcal{P} \times \mathcal{P}$. Any projective plane $\mathcal{A}$ has the property that

$$\cup_{A \in \mathcal{A}}(A \times A - \mathcal{D}) = \mathcal{P} \times \mathcal{P} - \mathcal{D}$$

where the sets $A \times A - \mathcal{D}$ are disjoint. Let the two projective planes we constructed above be $\mathcal{A}_i$ for $i = 1, 2$. For $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$, if the intersection of the two sets has cardinality 2, then $A \times A - \mathcal{D}$ and $B \times B - \mathcal{D}$ intersect at a set of cardinality 2. Otherwise, they are disjoint.

We define the following random variables X_1, X_2, X_3, X_4: let both X_1 and X_2 take values in $\mathcal{P}$ and the pair X_1, X_2 has the uniform joint distribution over $\mathcal{P} \times \mathcal{P} - \mathcal{D}$. Let the first projective plane be $\mathcal{A}_1 = \{A_1, \cdots, A_{13}\}$ and the second projective plane be $\mathcal{A}_2 = \{B_1, \cdots, B_{13}\}$. Let X_3 take value i on $A_i \times A_i - \mathcal{D}$ for $i \in \{1, 2, \cdots, 13\}$. Let X_4 take value i on $B_i \times B_i - \mathcal{D}$ for $i \in \{1, 2, \cdots, 13\}$. We have $I(X_1; X_2) = \log_2(13/12)$.

$$\begin{aligned} H(X_3) &= H(X_4) = \log_2 13 \\ H(X_3X_4) &= \log_2 6 + \log_2 13. \\ H(X_3|X_1) &= H(X_3|X_2) = H(X_4|X_1) \\ &= H(X_4|X_2) = \log_2 4 \\ H(X_3X_4|X_1) &= H(X_3X_4|X_2) = \log_2 12. \end{aligned}$$

Therefore,

$$I(X_3; X_4) = \log_2 13 - \log_2 6$$

and

$$I(X_3; X_4|X_1) = I(X_3; X_4|X_2) = \log_2(4/3).$$

This gives that for these four random variables

$$\begin{aligned} S(1, 2|3, 4) &= I(X_1; X_2) + I(X_3; X_4|X_1) \\ &\quad + I(X_3; X_4|X_2) - I(X_3; X_4) \\ &= \log(13/12) + 2 \times \log_2(4/3) - \log_2(13/6) \\ &= \log_2(52/27) - \log_2(13/6) = -\log_2(9/8) \\ &< 0. \end{aligned}$$

This example shows that the function S_F may be negative for some entropy functions F and that our inner bound is not tight.

III. Proof of Theorems 3 and 5

Let X, Y, Z, U be four jointly distributed discrete random variables with distribution $p(x, y, z, u)$. We denote all marginals of this distribution by the same letter p. For instance, its marginal on z, u is denoted by $p(z, u)$. Define

$$q(x, y, z, u, x_1, y_1) \overset{\text{def}}{=} \frac{p(x, y, z, u)p(x_1, y_1, z, u)}{p(z, u)}. \quad (44)$$

Since both $p(x, y, z, u)$ and $p(x_1, y_1, z, u)$ are absolutely continuous with respect to $p(z, u)$, we can see that q is a distribution of six random variables. Let X_1, Y_1 be two random variables jointly distributed with X, Y, Z, U according to the joint distribution q.

Actually, we can express $\Delta(Z, U|X, Y)$ in terms of the information quantities of the six random variables defined above.

Lemma 2:

$$\begin{aligned} \Delta(Z, U|X, Y) &= I(X; Y_1) - I(X; Y_1|U) - I(X; Y_1|Z) \\ &\quad - I(Z; U|X, Y_1). \end{aligned} \quad (45)$$

Proof:

$$\begin{aligned} \Delta(Z, U|X, Y) &= I(Z; U) - I(Z; U|X) - I(Z; U|Y) \\ &= I(Z; U) - I(Z; U|X) - I(Z; U|Y_1) \\ &= I(Z; U; X; Y_1) - I(Z; U|XY_1) \\ &= I(X; Y_1) - I(X; Y_1|U) - I(X; Y_1|Z) \\ &\quad + I(X; Y_1|ZU) - I(Z; U|XY_1) \\ &= I(X; Y_1) - I(X; Y_1|U) - I(X; Y_1|Z) \\ &\quad - I(Z; U|XY_1). \end{aligned}$$

The last step is due to the fact that $I(X; Y_1|ZU) = 0$. The lemma is proved. □

Proof of Theorem 3: From Lemma 2, we have

$$I(Z; U) - I(Z; U|X) - I(Z; U|Y) \le I(X; Y_1).$$

Similarly, we can prove

$$I(Z; U) - 2I(Z; U|X) \le I(X; X_1).$$

By means of these two inequalities, Theorem 3 is proved in the following way:

$$\begin{aligned}
&2I(Z; U) - 3I(Z; U|X) - I(Z; U|Y)\\
&\quad\le I(X; Y_1) + I(X; X_1)\\
&\quad= I(X; X_1Y_1) + I(X; X_1; Y_1)\\
&\quad= I(X; X_1Y_1) + I(X_1; Y_1) - I(X_1; Y_1|X)\\
&\quad\le I(X; X_1Y_1) + I(X_1; Y_1)\\
&\quad\le I(X; ZU) + I(X_1; Y_1)\\
&\quad= I(X; ZU) + I(X; Y).
\end{aligned}$$

In the penultimate step, we used the data processing inequality and the fact that $I(X; Y) = I(X_1; Y_1)$. The theorem is proved. □

Using the six random variables X, Y, U, Z, X_1, Y_1, we can actually determine all the missing terms of the inequality in Theorem 3. This is done as follows: from Lemma 2

$$\begin{aligned}
\Delta(Z, U|X, Y) &= I(Z; U) - I(Z; U|X) - I(Z; U|Y)\\
&= I(X; Y_1) - I(X; Y_1|U) - I(X; Y_1|Z)\\
&\quad - I(Z; U|XY_1).
\end{aligned}$$

Let $R_1 = I(X; Y_1|U) + I(X; Y_1|Z) + I(Z; U|XY_1)$. This equality is restated as

$$\Delta(Z, U|X, Y) = I(X; Y_1) - R_1.$$

Similarly, we have

$$\begin{aligned}
\Delta(Z, U|X, X_1) &= I(Z; U) - I(Z; U|X) - I(Z; U|X_1)\\
&= I(Z; U) - 2I(Z; U|X)\\
&= I(X; X_1) - I(X; X_1|U) - I(X; X_1|Z)\\
&\quad - I(Z; U|XX_1).
\end{aligned}$$

Let $R_2 = I(X; X_1|U) + I(X; X_1|Z) + I(Z; U|XX_1)$. This equality is restated as

$$\Delta(Z, U|X, X_1) = I(X; X_1) - R_2.$$

Therefore,

$$\begin{aligned}
&2I(Z; U) - 3I(Z; U|X) - I(Z; U|Y)\\
&\quad= I(X; Y_1) + I(X; X_1) - R_1 - R_2\\
&\quad= I(X; X_1Y_1) + I(X; X_1; Y_1) - R_1 - R_2\\
&\quad= I(X; ZU) - I(X; ZU|X_1Y_1) + I(X_1; Y_1)\\
&\qquad - I(X_1; Y_1|X) - R_1 - R_2\\
&\quad= I(X; ZU) + I(X; Y) - I(X; ZU|X_1Y_1)\\
&\qquad - I(X_1; Y_1|X) - R_1 - R_2.
\end{aligned}$$

This implies that the missing terms of the first inequality in Theorem 3 are

$$\begin{aligned}
&R(X, Y, Z, U, X_1, Y_1)\\
&\quad= \tfrac{1}{2}\left[I(X; ZU|X_1Y_1) + I(X_1; Y_1|X) + R_1 + R_2\right]\\
&\quad= \tfrac{1}{2}\left[I(X; X_1|U) + I(X; X_1|Z) + I(Z; U|XX_1)\right.\\
&\qquad + I(X_1; Y_1|X) + I(X; ZU|X_1Y_1)\\
&\qquad \left.+ I(X; Y_1|U) + I(X; Y_1|Z) + I(Z; U|XY_1)\right].
\end{aligned}$$

Apparently, the following function is in $\tilde{\Gamma}_4$:

$$\begin{aligned}
F(\phi) &= 0, & F(X) &= F(Y) = F(Z) = F(U) = 6a,\\
F(XY) &= 12a, & F(XZ) &= F(YZ) = F(YU) = F(XU) = 9a,\\
F(ZU) &= 10a, & F(XZU) &= F(YZU) = F(XZY)\\
& & &= F(XYU) = F(XZYU) = 12a.
\end{aligned}$$

For this function, one of our new inequalities is satisfied with equality. A natural question to ask is whether or not this function is asymptotically constructible. If this is true, then it is likely that

$$\overline{\Gamma}^*_4 = \tilde{\Gamma}_4.$$

Unfortunately, we were unable to prove this. Therefore, we doubt the correctness of this plausible conjecture.

IV. Proof of Theorem 6

In this section, we prove Theorem 6, the inner bound of $\overline{\Gamma}^*_4$. The result is proven via a series of basic constructions. Before we start to work on the proof of the result, we present first the basic constructions. In all constructions that follow, we use three independent and identically distributed ternary random variables W_1, W_2, and W_3 taking values in the set $\{0, 1, 2\}$. The common distribution of the random variables is the uniform distribution. Let W_0 represent a constant random variable taking a fixed value with probability 1. To make the notations simpler, in this section, we assume that the logarithmic function is based to 3. Therefore, the random variables W_i has entropy 1 for $i = 1, 2, 3$. The entropy of W_0 is zero. In this section, we are going to use the concept of atoms. There are 15 atoms in the case of four random variables. They are represented by the nonempty subsets of $\{1, 2, 3, 4\}$. For any function F, we will use the values of the function at atoms. The values of F at atoms are linear functions of the values of the function at subsets of $\{1, 2, 3, 4\}$. The values of the function at subsets will be denoted, for instance, for $\{1, 2, 3\}$, by $F(1, 2, 3)$. When F is an entropy function, this is the joint entropy. To distinguish from the values at subsets, the values of the function at atoms are represented, for instance, at atom $\{1, 2, 3\}$, by $F[1, 2, 3]$. When F is the entropy function of four discrete random variables X_1, X_2, X_3, X_4, we have

$$F[1, 2, 3] = I(X_1; X_2; X_3|X_4).$$

Construction 1: For any nonempty subset α of $\{1, 2, 3, 4\}$, let $X_i = W_1$ if $i \in \alpha$ and $X_i = W_0$, otherwise. The function defined by this construction is denoted by F^1_α. It is easy to check that, for this construction, for any $\beta \ne \alpha$, $F^1_\alpha[\beta] = 0$ and $F^1_\alpha[\alpha] = 1$.

Construction 2: $X_1 = W_1$, $X_2 = W_2$, $X_3 = W_3$, $X_4 = W_1 + W_2 + W_3 \mathrm{mod}\,(3)$. The function defined by this construction is denoted by F^2. For this construction, the function has value zero at all weight-one atoms, has value 1 at all weight-two and weight-four atoms, and has value -1 at all weight-three atoms.

Construction 3: $X_1 = W_1, X_2 = W_2, X_3 = W_1 + W_2 \pmod 3), X_4 = W_0$. The function defined by this construction is denoted by F_4^3 where 4 indicates that random variable $X_4 = W_0$. The construction is actually symmetric for the other three random variables. We also have the other three similar constructions obtained by permuting the four random variables. The functions so constructed are denoted by F_i^3 where i indicates that $X_i = W_0$. For this construction

$$F_4^3[1, 2] = F_4^3[1, 3] = F_4^3[2, 3] = 1$$

and

$$F_4^3[1, 2, 3] = -1.$$

At all other atoms the values are zero.

Construction 4: $X_1 = W_1, X_2 = W_2, X_3 = X_4 = W_1 + W_2 (\mathrm{mod}\, 3)$. The function so constructed is denoted by $F_{3,4}^4$. We can also construct a function $F_{i,j}^4$ and its meaning is self-explanatory. For this construction,

$$F_{3,4}^4[1, 2] = 1$$
$$F_{3,4}^4[2, 3, 4] = F_{3,4}^4[1, 3, 4] = 1$$

and

$$F_{3,4}^4[1, 2, 3, 4] = -1.$$

At all other atoms it has zero value.

Construction 5: $X_1 = W_1, X_2 = W_2, X_3 = W_1 + W_2 \pmod 3)$, $X_4 = W_1 - W_2 \pmod 3)$. The function constructed by this method is denoted by F^5. For this construction, at all weight-one and -two atoms the value is zero; at all weight–three atoms the value is 1; and $F^5[1, 2, 3, 4] = -2$.

Construction 6: $X_1 = W_1, X_2 = W_2, X_3 = W_3, X_4 = (W_1 + W_2 \pmod 3), W_1 + W_3 \pmod 3))$. The function constructed by this method is denoted by F_4^6. We can also construct functions F_i^6 for other values of i by the same method. For this construction, $F_4^6[1, 2, 3] = 1$; at all weight-two atoms containing 4, the values are 1; and $F_4^6[1, 2, 3, 4] = -1$. At all other atoms the values are zero.

Construction 7: $X_1 = W_1, X_2 = W_2, X_3 = W_1 + W_2 \pmod 3)$, $X_4 = (W_1, W_2)$. The function constructed by this method is denoted by F_4^7. We can also construct F_i^7 for other values of i by the same method. For this construction, at all atoms of weight at most two the value is zero; at all weight-three atoms except the atom $\{1, 2, 3\}$ the values are 1; $F_4^7[1, 2, 3, 4] = -1$ and $F_4^7[1, 2, 3] = 0$.

To help the reader to understand the proof, we give charts for the values of the functions constructed in Constructions 2–7. Since these functions take zero values at all weight-one atoms, we give their values at atoms of weight at least two. The atoms of weight at least two are arranged as follows:

$$\begin{array}{ccccccc} & & 23 & & & & \\ & & | & & & & \\ 12 & - & 123 & - & 13 & & \\ | & & | & & | & & \\ 124 & - & 1234 & - & 134 & - & 14 \\ | & & | & & | & & \\ 24 & - & 234 & - & 34 & & \end{array}$$

The values of the functions 2–7 are given as follows using this chart.

$$\begin{array}{ccccccc} & & (1) & & & & \\ & & | & & & & \\ (1) & - & (-1) & - & (1) & & \\ | & & | & & | & & \\ (-1) & - & (1) & - & (-1) & - & (1) \\ | & & | & & | & & \\ (1) & - & (-1) & - & (1) & & \end{array}$$

Function F^2.

$$\begin{array}{ccccccc} & & (1) & & & & \\ & & | & & & & \\ (1) & - & (-1) & - & (1) & & \\ | & & | & & | & & \\ (0) & - & (0) & - & (0) & - & (0) \\ | & & | & & | & & \\ (0) & - & (0) & - & (0) & & \end{array}$$

Function F_4^3.

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (1) & - & (0) & - & (0) & & \\ | & & | & & | & & \\ (0) & - & (-1) & - & (1) & - & (0) \\ | & & | & & | & & \\ (0) & - & (1) & - & (0) & & \end{array}$$

Function $F_{3,4}^4$.

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (0) & - & (1) & - & (0) & & \\ | & & | & & | & & \\ (1) & - & (-2) & - & (1) & - & (0) \\ | & & | & & | & & \\ (0) & - & (1) & - & (0) & & \end{array}$$

Function F^5.

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (0) & - & (1) & - & (0) & & \\ | & & | & & | & & \\ (0) & - & (-1) & - & (0) & - & (1) \\ | & & | & & | & & \\ (1) & - & (0) & - & (1) & & \end{array}$$

Function F_4^6.

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (0) & - & (0) & - & (0) & & \\ | & & | & & | & & \\ (1) & - & (-1) & - & (1) & - & (0) \\ | & & | & & | & & \\ (0) & - & (1) & - & (0) & & \end{array}$$

Function F_4^7.

In [39], we proved the following result:

Lemma 3: If $F \in \overline{\Gamma}^*_n$ and $\alpha > 0$, then $\alpha F \in \overline{\Gamma}^*_n$. If $F_1, F_2 \in \overline{\Gamma}^*_n$, then $F_1 + F_2 \in \overline{\Gamma}^*_n$.

That is, $\overline{\Gamma}^*_n$ is a convex cone.

Proof of Theorem 6: A function F in the region $\hat{\Gamma}_4$ should satisfy the following conditions in terms of atoms: Let the four indices of the random variables be $\{i, j, k, l\}$ which is a permutation of the set $\{1, 2, 3, 4\}$. From the definition of the region $\hat{\Gamma}_4$, the inequalities the function F should satisfy include:

1) F is nonnegative at all atoms of weight one;

2) $F[i, j|\phi] = F[i, j] + F[i, j, k] + F[i, j, l] + F[i, j, k, l] \geq 0$, where ϕ is the empty set;

$$\begin{array}{ccccccc}
 & & 23 & & & & \\
 & & | & & & & \\
(12) & - & (123) & - & 13 & & \\
| & & | & & | & & \\
(124) & - & (1234) & - & 134 & - & 14 \\
| & & | & & | & & \\
24 & - & 234 & - & 34 & &
\end{array}$$

The atoms involved in inequality (2).

3) $F[i, j|k] = F[i, j] + F[i, j, l] \geq 0$;

$$\begin{array}{ccccccc}
 & & 23 & & & & \\
 & & | & & & & \\
(12) & - & (123) & - & 13 & & \\
| & & | & & | & & \\
124 & - & 1234 & - & 134 & - & 14 \\
| & & | & & | & & \\
24 & - & 234 & - & 34 & &
\end{array}$$

The atoms involved in inequality (3).

4) $F[i, j] + F[i, j, k] + F[i, j, l] + F[k, l] \geq 0$;

$$\begin{array}{ccccccc}
 & & 23 & & & & \\
 & & | & & & & \\
(12) & - & (123) & - & 13 & & \\
| & & | & & | & & \\
(124) & - & 1234 & - & 134 & - & 14 \\
| & & | & & | & & \\
24 & - & 234 & - & (34) & &
\end{array}$$

The atoms involved in inequality (4).

5) $F[i, j] \geq 0$.

Notice that the fourth condition comes from the constraint

$$S_F(i, j, k, l) \geq 0.$$

Other inequalities come from the nonnegativity of conditional mutual informations of two random variables and the nonnegativity of the conditional entropies. Lemma 1 is useful in finding the atoms involved in these inequalities. These five conditions will be referred to as Conditions 1–5 in the proof. The readers can extremely reduce the difficulty in subsequent reading by familiarizing themselves with these five conditions in the atom chart for all permutations of i, j, k, and l.

A function $F[.]$ is called nonnegative if its values at all atoms are nonnegative.

Lemma 4: Nonnegative functions are asymptotically constructible.

Proof: If J is a function that takes nonnegative values at all atoms, then

$$J = \sum_{\alpha} J[\alpha] F^1_\alpha$$

where $J[\alpha]$ is nonnegative for all α. It is asymptotically constructible from Construction 1 and Lemma 4. The lemma is proved. □

The basic idea for the proof is that for any function F in $\hat{\Gamma}_4$, we can find a sequence of basic functions from Constructions 1–7 $F_1, \cdots, F_m$ for some $m > 0$ and a sequence of nonnegative reals $a_1, \cdots, a_m$ such that

$$F = \sum_{j=1}^{m} a_j F_j.$$

Once we can prove this, then the theorem is proved by invoking Lemma 4. Suppose F is in $\hat{\Gamma}_4$ and F' is a basic function from Constructions 1–7 and $a > 0$. If $F - aF' \in \hat{\Gamma}_4$, then we say that subtracting aF' from F is a legal operation. We prove the theorem by finding a sequence of legal operations to reduce F to a nonnegative function, which is asymptotically constructible by Lemma 5. This implies by invoking Lemma 4 that the original function is asymptotically constructible.

Construction 1 is used only in Lemma 5. In the proof that follows, we use only the other six constructions. We notice that in Constructions 2–7, no atom of weight one is involved. As long as the values of the function at the weight-one atoms are nonnegative to start with, upon subtracting a nonnegative multiple of any of these constructions, Condition 1 remains satisfied. Therefore, we can always ignore the weight-one atoms when we consider subtracting a nonnegative multiple of any of these constructions.

We find these legal operations in the following steps.

Step 1: We notice that F^2 satisfies all inequalities in Conditions 2–4 with equalities. So subtracting aF^2 from F where

$$a = \min_{\{i, j\} \subset \{1, 2, 3, 4\}} F[i, j]$$

is a legal operation because 1) Conditions 2–4 will remain to be satisfied since subtracting zero does not change the direction of the inequalities and 2) since a is defined as the minimum of the values of the function at all weight-two atoms, after subtracting aF^2 the values of F at these atoms are at least 0. Therefore, Condition 5 holds. Without loss of generality, we assume that

$$F[1, 2] = \min_{\{i, j\} \subset \{1, 2, 3, 4\}} F[i, j].$$

Let $F' = F - aF^2$. We have

i) $F' \in \hat{\Gamma}_4$;

ii) $F'[1, 2] = 0$.

In the following chart, the atoms are marked either by a 0 indicating that the value of the function F' at this atom is zero, or by a $+$ indicating that the value of the function F' at this atom is nonnegative, or by an X indicating that the value of the function F' at this atom may be negative.

$$\begin{array}{ccccccc}
 & & (+) & & & & \\
 & & | & & & & \\
(0) & - & (+) & - & (+) & & \\
| & & | & & | & & \\
(+) & - & (X) & - & (X) & - & (+) \\
| & & | & & | & & \\
(+) & - & (X) & - & (+) & &
\end{array}$$

The Function F'.

Step 2: A function is called seminonnegative if its values at all atoms of weight up to three are nonnegative. In this step, we prove that F' can be reduced to a seminonnegative function via a series of legal operations. From the chart for F', we see F' is not seminonnegative if and only if at least one of two values of F', $F'[1, 3, 4]$ and $F'[2, 3, 4]$, is negative. Suppose

$$F'[1, 3, 4] < 0.$$

Let $a = -F'[1, 3, 4]$. We prove that subtracting aF_2^3 from F' is legal. Let $G = F' - aF_2^3$. We notice that

$$F_2^3[i, j|\phi] = 0$$
$$F_2^3[i, j|k] = 0,$$

for any $k \notin \{i, j\}$. These observations and $F' \in \hat{\Gamma}_4$ imply that

$$G[i, j|\phi] \geq 0$$
$$G[i, j|k] \geq 0.$$

From $G[1, 3, 4] = 0$, we see that

$$G[1, 3] = G[1, 3] + G[1, 3, 4] = F'[1, 3] + F'[1, 3, 4] \geq 0.$$

Similarly, we have

$$G[1, 4] \geq 0$$
$$G[3, 4] \geq 0.$$

For other pairs i, j, since the values of F' are not affected by the operation, we still have

$$G[i, j] \geq 0.$$

To show that $G \in \hat{\Gamma}_4$, we need to check only Condition 4

$$G[i, j] + G[i, j, k] + G[i, j, l] + G[k, l] \geq 0.$$

There are six of them for

$$(i, j) = (3, 4),\ (1, 4),\ (1, 3),\ (1, 2),\ (2, 3),\ (2, 4).$$

The inequalities for $(i, j) = (1, 2),\ (1, 3),\ (1, 4)$ are trivial because all entries are nonnegative. The proofs for $(i, j) = (2, 3),\ (2, 4)$ are the same. We prove it only for $(i, j) = (2, 4)$. We need also to prove it for $(i, j) = (3, 4)$. For $(i, j) = (2, 4)$

$$\begin{aligned}
G[1, 3]&+G[1, 2, 4]+G[2, 3, 4]+G[2, 4] \\
&= F'[1, 3]+F'[1, 2, 4]+F'[2, 3, 4]+F'[2, 4] \\
&\quad - a(F_2^3[1, 3]+F_2^3[1, 2, 4]+F_2^3[2, 3, 4]+F_2^3[2, 4]) \\
&= F'[1, 3]+F'[1, 2, 4]+F'[2, 3, 4]+F'[2, 4]-aF_2^3[1, 3] \\
&= F'[1, 3]+F'[1, 2, 4]+F'[2, 3, 4]+F'[2, 4]+F'[1, 3, 4] \\
&\geq F'[1, 3]+F'[1, 3, 4]+F'[2, 3, 4]+F'[2, 4] \\
&\geq 0.
\end{aligned}$$

In the next to the last step, we used the fact that $F'[1, 2, 4] \geq 0$ and in the last step, we used the fact that $F' \in \hat{\Gamma}_4$. For $(i, j) = (3, 4)$

$$\begin{aligned}
G[1, 2] &+ G[2, 3, 4] + G[1, 3, 4] + G[3, 4] \\
&= F'[1, 2] + F'[1, 3, 4] + F'[2, 3, 4] + F'[3, 4] \\
&\quad - a(F_2^3[1, 2] + F_2^3[1, 3, 4] + F_2^3[2, 3, 4] + F_2^3[3, 4]) \\
&= F'[1, 2] + F'[1, 3, 4] + F'[2, 3, 4] + F'[3, 4] \\
&\quad - a(F_2^3[1, 3, 4] + F_2^3[3, 4]) \\
&= F'[1, 2] + F'[1, 3, 4] + F'[2, 3, 4] + F'[3, 4] \\
&\geq 0.
\end{aligned}$$

In the next to the last step, we used the fact that $F_2^3[1, 3, 4] + F_2^3[3, 4] = 0$ and in the last step, we used the fact that $F' \in \hat{\Gamma}_4$. This proves that $G \in \hat{\Gamma}_4$. If $G[2, 3, 4] \geq 0$, then G is already seminonnegative. Otherwise, repeating the same proof by replacing atom $\{1, 3, 4\}$ by $\{2, 3, 4\}$, we can obtain a seminonnegative function. Therefore, without loss of generality, we assume that G is already seminonnegative.

Step 3: Since G is seminonnegative, if the value of G at $\{1, 2, 3, 4\}$ is nonnegative, then the function is already nonnegative and therefore asymptotically constructible from Lemma 5. Otherwise, we continue to find legal operations to convert the function to a nonnegative function. In doing so, the inequalities we need to consider are those in which the atom of weight four is involved. That is, the following six inequalities

$$\sum_{\{i, j\} \subset \alpha} G[\alpha] \geq 0$$

for all six pairs i, j from $\{1, 2, 3, 4\}$. The following observations will be useful in the remaining part of the proof.

Observation 1: Let i, j, k, l be a permutation of 1, 2, 3, 4, and let J be a seminonnegative function. Then

$$J[i, j, k] + J[1, 2, 3, 4] \geq 0$$

implies that subtracting $aF_{i,j}^4$ from J is legal and results in a seminonnegative function where

$$a = \min\{J[k, l],\ J[i, k, l],\ J[j, k, l]\}.$$

Observation 2: Let i, j, k, l be a permutation of $1, 2, 3, 4$, and let J be a seminonnegative function. Then

$$\begin{aligned} J[i, j, k] + J[1, 2, 3, 4] &\geq 0\\ J[i, j, l] + J[1, 2, 3, 4] &\geq 0\\ J[i, k, l] + J[1, 2, 3, 4] &\geq 0 \end{aligned}$$

implies that subtracting aF_i^7 from J is legal and results in a nonnegative function where

$$a = \min\{J[i, k, l],\ J[i, j, l],\ J[i, j, k]\}.$$

The validity of these propositions is obvious.

Since functions F_i^3, F^5, and F_i^6 satisfy all these six inequalities with equalities, as long as at atoms of weight up to three the values of the function are not reduced below zero, subtracting a nonnegative multiple of one of these functions is always legal and results in a seminonnegative function. Suppose we keep performing these legal operations until no more legal operations resulting in a seminonnegative function using these three functions are possible. We distinguish the following cases according to the function G' that is resulted in.

Because no operation using function F_i^3 is legal, for any subset $\{i, j, k\}$ of $\{1, 2, 3, 4\}$, G' is zero at least one of the following atoms: $\{i, j\}$, $\{i, k\}$, $\{k, j\}$ (cf., the atom chart for F_i^3). There are only two possible cases:

Case 1: There exists a 3-subset, say $\{i, j, k\}$, such that G' is zero at all three atoms: $\{i, j\}$, $\{i, k\}$, $\{k, j\}$.

Case 2: There exist two disjoint weight-two atoms, say $\{i, j\}$ and $\{k, l\}$, such that the values of the function G' at these two atoms are both zero.

In Case 1, without loss of generality, we assume that

$$G'[1, 2] = G'[1, 3] = G'[2, 3] = 0.$$

Since F_4^6 does not give a legal operation resulting in a seminonnegative function, the function takes value zero at one of the following four atoms: $\{1, 4\}$, $\{2, 4\}$, $\{3, 4\}$, $\{1, 2, 3\}$. This gives two subcases,

Case 1.1: The function is zero at $\{1, 4\}$ (or equivalently one of two other weight-two atoms listed above).

Case 1.2: The function is zero at $\{1, 2, 3\}$.

In Case 1.1, since F^5 does not give a legal operation resulting in a seminonnegative function, at least one of the four weight-three atoms, the function takes zero value. We consider only the cases where at one of the three atoms $\{1, 2, 4\}$, $\{1, 3, 4\}$, and $\{2, 3, 4\}$, the value of the function is zero. The case where the function is zero at the atom $\{1, 2, 3\}$ is equivalent to Case 1.2. Since the first two atoms are symmetric in this context, we consider only the case that the function G' is zero at $\{1, 2, 4\}$. We can see that Condition 2 implies

$$G'[1, 3, 4] + G'[1, 2, 3, 4] \geq 0$$

because both $G'[1, 4]$ and $G'[1, 2, 4]$ are zero

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (0) & - & (+) & - & (0) & & \\ | & & | & & | & & \\ (0) & - & (X) & - & (+) & - & (0) \\ | & & | & & | & & \\ (+) & - & (+) & - & (+) & & \end{array}$$

The Function G' in Case 1.1 for $G'[1, 2, 4] = 0$.

Let

$$a = \min\{G'[2, 4],\ G'[1, 2, 3],\ G'[1, 3, 4]\}.$$

The inequalities above imply that subtracting $aF_{1,3}^4$ from G' is legal and results in a seminonnegative function G''. If G'' is zero at either $\{1, 2, 3\}$ or $\{1, 3, 4\}$, then $G''[1, 2, 3, 4] \geq 0$. The function obtained is already nonnegative. Otherwise, $G''[2, 4] = 0$. This implies

$$G''[2, 3, 4] + G''[1, 2, 3, 4] \geq 0.$$

Let $b = -G''[1, 2, 3, 4]$, if $b > 0$ then subtracting bF_3^7 is a legal operation and this results in a function that is nonnegative at all atoms.

If $G'[2, 3, 4] = 0$, we have

$$G'[1, 2, 3] + G'[1, 2, 3, 4] \geq 0$$

because both $G'[2, 3]$ and $G'[2, 3, 4]$ are zero.

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (0) & - & (+) & - & (0) & & \\ | & & | & & | & & \\ (+) & - & (X) & - & (+) & - & (0) \\ | & & | & & | & & \\ (+) & - & (0) & - & (+) & & \end{array}$$

The Function G' in Case 1.1 for $G'[2, 3, 4] = 0$.

Let

$$a = \min\{G'[3, 4],\ G'[1, 2, 3],\ G'[1, 2, 4]\}.$$

The inequalities above imply that subtracting $aF_{1,2}^4$ from G' is legal and results in a seminonnegative function G''. If G'' is zero at $\{1, 2, 3\}$ then $G''[1, 2, 3, 4] \geq 0$. If G'' is zero at $\{1, 2, 4\}$ then this goes back to the previous case. In both cases, either the function obtained is already nonnegative, or it can be reduced to a nonnegative function by legal operations. Otherwise, $G''[3, 4] = 0$. For this function, we have

$$G''[1, 2, 3, 4] + G''[1, 3, 4] \geq 0.$$

Let

$$a = \min\{G''[2, 4],\ G''[1, 2, 3],\ G''[1, 3, 4]\}.$$

The inequalities above imply that subtracting $aF_{1,2}^4$ from G'' is legal and results in a seminonnegative function G'''. If G''' is zero at either $\{1, 2, 3\}$ or $\{1, 3, 4\}$, then $G''[1, 2, 3, 4] \geq 0$. Otherwise, $G'''[2, 4] = 0$. For this function, we have

$$\begin{aligned} G'''[1, 2, 3, 4] + G'''[1, 2, 4] &\geq 0 \\ G'''[1, 2, 3, 4] + G'''[1, 2, 3] &\geq 0 \\ G'''[1, 2, 3, 4] + G'''[1, 3, 4] &\geq 0. \end{aligned}$$

Let $b = -G''[1, 2, 3, 4]$, if $b > 0$ then subtracting bF_1^7 is a legal operation and this results in a function that is nonnegative at all atoms.

In Case 1.2, we have

$$\begin{aligned} G'[1, 3, 4] + G'[1, 2, 3, 4] &\geq 0, \\ G'[1, 2, 4] + G'[1, 2, 3, 4] &\geq 0, \\ G'[2, 3, 4] + G'[1, 2, 3, 4] &\geq 0. \end{aligned}$$

$$\begin{array}{ccccccc} & & (0) & & & & \\ & & | & & & & \\ (0) & - & (0) & - & (0) & & \\ | & & | & & | & & \\ (+) & - & (X) & - & (+) & - & (+) \\ | & & | & & | & & \\ (+) & - & (+) & - & (+) & & \end{array}$$

The Function G' in Case 1.2.

Let $a = -G'[\{1, 2, 3, 4\}]$. If $a > 0$, then subtracting aF_4^7 is a legal operation. This results in a nonnegative function. Otherwise, the function is already nonnegative.

We now consider Case 2. Without loss of generality, we assume that $G'[1, 2] = G'[3, 4] = 0$. Since F^5 does not give a legal operation resulting in a seminonnegative function, the function G' has value zero at least one of the four weight-three atoms. Without loss of generality, we assume $G'[1, 2] = G'[3, 4] = G'[1, 2, 3] = 0$. Then we have

$$G'[1, 2, 4] + G'[1, 2, 3, 4] \geq 0.$$

$$\begin{array}{ccccccc} & & (+) & & & & \\ & & | & & & & \\ (0) & - & (0) & - & (+) & & \\ | & & | & & | & & \\ (+) & - & (X) & - & (+) & - & (+) \\ | & & | & & | & & \\ (+) & - & (+) & - & (0) & & \end{array}$$

The Function G' in Case 2.

Let $a = \min\{G'[2, 3],\ G'[1, 3, 4],\ G'[1, 2, 4]\}$. Then subtracting $aF_{1,4}^4$ is legal. The function G'' resulting from this legal operation takes zero value at either $\{1, 3, 4\}$, $\{1, 2, 4\}$, or $\{2, 3\}$. In the first case, $G''[1, 3, 4] = 0$, we have

$$G''[1, 3] + G''[1, 2, 3, 4] \geq 0.$$

Apparently, subtracting $bF_{2,4}^4$ is legal where

$$b = \min\{G''[1, 3],\ G''[1, 2, 4],\ G''[2, 3, 4]\}.$$

This results in a nonnegative function. In the second case, $G''[2, 3] = 0$, we have

$$G'[2, 3, 4] + G'[1, 2, 3, 4] \geq 0.$$

Let

$$a = \min\{G'[1, 3],\ G'[1, 2, 4],\ G'[2, 3, 4]\}.$$

Subtracting $aF_{2,4}^4$ is legal. Let G'' be $G' - aF_{2,4}^4$. Then either $G''[1, 2, 4] = 0$ or $G''[1, 3, 4] = 0$. In both cases, we must have $G''[1, 2, 3, 4] \geq 0$, that is, the function is nonnegative. Otherwise, $G''[1, 3] = 0$. This implies

$$\begin{aligned} G''[1, 3, 4] + G''[1, 2, 3, 4] &\geq 0 \\ G''[1, 2, 4] + G''[1, 2, 3, 4] &\geq 0 \end{aligned}$$

and

$$G''[2, 3, 4] + G''[1, 2, 3, 4] \geq 0.$$

Then, subtracting F_4^7 is legal and results in a nonnegative function. In the third case, $G'[1, 2, 4] = 0$, $G'[1, 2, 3, 4]$ must be nonnegative. Hence, G' is a nonnegative function.

Thus we have proved that we can always reduce a function in $\hat{\Gamma}_4$ by legal operations to a function that takes nonnegative values at all atoms. By Lemma 5, Theorem 6 follows.

V. Concluding Remarks

The key result of this paper is Theorem 3. This discovery shows that the set of so-called basic information inequalities cannot fully characterize Shannon's entropy function in the sense of Theorem 4. That is, the region Γ_n is strictly greater than the region $\overline{\Gamma}_n^*$. This is a surprising result because based on intuition, one tends to believe that the opposite is true. Actually, when we started to look into this problem, we tried first to prove that

$$\overline{\Gamma}_4^* = \Gamma_4$$

by finding all kinds of constructions for four random variables as in the proof of Theorem 6. Only after we failed to find a construction in one of many cases, we started to doubt the correctness of our conjecture. This led to the discovery of this new information inequality.

The full characterization of the region $\overline{\Gamma}_n^*$ seems to be a highly nontrivial problem. Even in the case of $n = 4$, we were unable to determine the region. We, instead, provided an inner bound of the region. This is Theorem 6 of the paper. The inner bound and the outer bound we found in this paper differ. It has been shown by an example that the inner bound is not tight. Unfortunately, the construction method we used in this example is not powerful enough to show that our outer bound is tight.

The simplest case of the problem is the case of $n = 4$ because this number is the smallest integer for which Γ_n and $\overline{\Gamma}_n^*$ differ. Although we mainly have concentrated on this simplest case, we have proved Theorem 5 which is a generalization of Theorem 3 to any number of random variables.

We also determined the missing terms in the inequalities in Theorem 3. They are expressed in terms of some auxiliary random variables. We did so in hope that this may be helpful in further searching for new information inequalities, as well as in further searching for improved inner bounds.

To get a better understanding of the behavior of the entropy function, it is important to fully characterize the function at least in the simplest case of $n = 4$. That is, the simplest task in this research direction is to determine the region $\overline{\Gamma}_4^*$. Based on our experience, we do not believe our outer bound to be tight. That is, we believe that there may exist more linear unconditional information inequalities involving four random variables.

The meaning of the new information inequalities provided by Theorems 3 and 5 are still not fully understood. Although we have used the region Γ_n^* to study the so-called distributed source coding problem, it is still of great interest to find more applications of the inequalities in other information-theoretical problems, especially in multiuser channel coding or source coding problems.

The problems studied in this paper have close connection to some other areas such as probabilistic reasoning, relational database, and so on. To study the implication of our results in those areas is also of interest.

Acknowledgment

The authors are grateful to the two referees. Their detailed comments and suggestions have greatly improved the readability of the paper.

References

[1] N. M. Abramson, *Information Theory and Coding.* New York: McGraw-Hill, 1963.

[2] L. L. Campbell, "Entropy as a measure," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 112–114, Jan. 1965.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems.* New York: Academic, and Budapest, Hungary: Akademiai Kiado, 1981.

[4] A. P. Dawid, "Conditional independence in statistical theory (with discussion)," *J. Roy. Statist. Soc., Ser. B*, vol. 41, pp. 1–31.

[5] S. Fujishige, "Polymatroidal dependence structure of a set of random variables," *Inform. Contr.*. vol. 39, pp. 55–72, 1978.

[6] T. Kawabata and R. W. Yeung, "The structure of the I-measure of a Markov chain," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1146–1149, 1992.

[7] T. S. Han, "Linear dependence structure of the entropy space," *Inform. Contr.*, vol. 29, pp. 337–368.

[8] ———, "Nonnegative entropy measures of multivariate symmetric correlations," *Inform. Contr.*, vol. 36, pp. 133–156, 1978.

[9] ———, "A uniqueness of Shannon's information distance and related nonnegativity problems," *J. Comb., Inform. Syst. Sci.*, vol. 6, pp. 320–321, 1981.

[10] G.-d. Hu, "On the amount of information," *Teor. Veroyatnost. i Primenen.*, vol. 4, pp. 447–455, 1962, in Russian.

[11] F. Matúš, private communication.

[12] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes.* Amsterdam, The Netherlands: North-Holland, Elsevier Science B.V., 1977.

[13] M. Matúš, "Abstract functional dependency structures," *Theor. Comput. Sci.*, vol. 81, pp. 117–126, 1991.

[14] ———, "On equivalence of Markov properties over undirected graphs," *J. Appl. Probab.*, vol. 29, pp. 745–749, 1992.

[15] ———, "Ascending and descending conditional independence relations," in *Trans. 11th Prague Conf. Information Theory, Statistical Decision Functions and Random Processes*, vol. B. Prague, Czechoslovakia: Academia, pp. 181–200, 1992.

[16] ———, "Probabilistic conditional independence structures and matroid theory: Background," *Int. J. Gen. Syst.*, vol. 22, pp. 185–196.

[17] ———, "Extreme convex set functions with many nonnegative differences," *Discr. Math.*, vol. 135, pp. 177–191, 1994.

[18] F. Matúš, "Conditional independences among four random variables II," *Combin., Prob. Comput.*, vol. 4, pp. 407–417, 1995.

[19] ———, "Conditional independence structures examined via minors," *Ann. Math., Artificial Intell.*, vol. 21, pp. 99–128, 1997.

[20] F. Matúš and M. Studený, "Conditional independences among four random variables I," *Combin., Prob. Comput.*, vol. 4, pp. 269–278, 1995.

[21] W. J. McGill, "Multivariate information transmission," in *Trans. Prof. Group Inform. Theory, 1954 Symp. Information Theory*, vol. PGIT-4, 1955, pp. 93–111.

[22] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, 1984.

[23] J. Pearl, *Probabilistic Reasoning in Intelligent Systems.* San Mateo, CA: Morgan Kaufman, 1988.

[24] *An Introduction to Information Theory.* New York: McGraw-Hill, 1961.

[25] M. Studený, "Attempts at axiomatic description of conditional independence," in *Proc. Workshop on Uncertainty Processing in Expert Systems*, supplement to *Kybernetika*, vol. 25, nos. 1–3, pp. 65–72, 1989.

[26] ———, "Multiinformation and the problem of characterization of conditional independence relations," *Probl. Contr. Inform. Theory*, vol. 18, pp. 3–16, 1989.

[27] ———, "Conditional independence relations have no finite complete characterization," in *Trans. 11th Prague Conf. Information Theory, Statistical Decision Functions and Random Processes*, vol. B. Prague, Czechoslovaka: Academia, pp. 377–396, 1992.

[28] ———, "Structural semigraphoids," *Int. J. Gen. Syst.*, vol. 22, no. 2, pp. 207–217, 1994.

[29] ———, "Descriptions of structures of stochastic independence by means of faces and imsets (in three parts)," *Int. J. Gen. Syst.*, vol. 23, pp. 123–137, pp. 201–219, pp. 323–341, 1994/1995.

[30] T. Tsujishita, "On triple mutual information," *Adv. Appl. Math.*, vol. 16, pp. 269–274, 1995.

[31] S. Watanabe, "A study of ergodicity and redundancy on intersymbol correlation of finite range," in *Trans. 1954 Symp. Inform. Theory* (Cambridge, MA, Sept. 15–17, 1954), p. 85.

[32] ———, "Information theoretical analysis of multivariate correlation," *IBM J.*, pp. 66–81, 1960.

[33] D. J. A. Welsh, *Matroid Theory.* New York: Academic 1976.

[34] R. W. Yeung, "A new outlook on Shannon's information measures," *IEEE Trans. Inform. Theory*, vol. 37, pp. 466–474, 1991.

[35] ———, "A framework for linear information inequalities," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1924–1934, Nov. 1997.

[36] R. W. Yeung, T. T. Lee, and Z. Ye, "An information-theoretic characterization of Markov random fields and its applications," *IEEE Trans. Inform. Theory*, submitted for publication.

[37] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communication," *IEEE Trans. Inform. Theory*, submitted for publication.

[38] R. W. Yeung and Y.-O. Yan, "Information theoretic inequality prover." [Online] Available: http://www.ie.cuhk.edu.hk/ ITIP or http://it.ucsd.edu/\verb+˜+whyeung(mirrorsite)..

[39] Z. Zhang and R. W. Yeung, "A non-Shannon type conditional inequality of information quantities," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1982–1985, Nov. 1997.