On a Relation Between Information Inequalities and Group Theory

Terence H. Chan, *Member, IEEE,* and Raymond W. Yeung, *Senior Member, IEEE*

***Abstract*—In this paper, we establish a one-to-one correspondence between information inequalities and group inequalities. The major implication of our result is that we can prove information inequalities by proving the corresponding group inequalities, and *vice versa*. By giving a group-theoretic proof for all Shannon-type inequalities, we suggest that new inequalities could be discovered by making use of the rich set of tools in group theory. On the other hand, via a non-Shannon-type information inequality recently discovered by Zhang and Yeung, we obtain a new inequality in group theory whose meaning is yet to be understood.**

***Index Terms*—Entropy, groups, group-theoretic inequalities, information inequalities.**

I. Introduction

THE quest for inequalities in information theory has been driven by the need to solve various communication problems. These inequalities play a crucial role in the proofs of almost all converse coding theorems in source and channel coding problems. In essence, they govern the impossibility in information theory.

The focus of this paper is inequalities which involve only Shannon's information measures, namely, entropies, mutual information, and conditional versions of these information measures. We refer to inequalities involving only Shannon's information measures as information inequalities. All the random variables involved are discrete. In [6], these inequalities are referred to as "the laws of information theory."

The main contribution of this paper is a group-theoretical interpretation of information inequalities. With this interpretation, we can translate the problem of proving an information inequality to a group-theoretical problem. It opens the door to discovering and proving new information inequalities by means of tools in group theory.

II. A Framework for Information Inequalities

Let $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ be n nonempty sets. Let Ω be the collection of all nonempty subsets of $\mathcal{N}$. For any $\alpha \in \Omega$, we define $\mathcal{X}_\alpha = \prod_{i\in\alpha} \mathcal{X}_i$ to be the Cartesian product of $\mathcal{X}_i$ for $i \in \alpha$. Let $X_1, X_2, \ldots, X_n$ be n jointly distributed discrete random variables defined on $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$, respectively. For any $\alpha \in \Omega$, X_α denotes the joint random variable $(X_i\colon i \in \alpha)$. For example, $X_{\{1,2\}}$ is the joint random variable of X_1 and X_2. For simplicity, the parentheses in the subscript are usually omitted, i.e., $X_{\{1,2\}}$ is written as $X_{1,2}$. Also, the joint entropy of X_α is denoted by $H(X_\alpha)$.

Let $\mathcal{H}_n$ be the set of all real functions defined on Ω. In other words, $\mathcal{H}_n$ is the set of all real functions defined on the collection of nonempty subsets of $\mathcal{N}$ and, hence, is a $(2^n - 1)$-dimensional Euclidean space [10]. For simplicity, for any function $\boldsymbol{g} \in \mathcal{H}_n$, the function value $\boldsymbol{g}(\alpha)$ is denoted by g_α for all $\alpha \in \Omega$.

Definition 2.1: Let $\boldsymbol{g} \in \mathcal{H}_n$. Then $\boldsymbol{g}$ is an entropy function if there exists a set of random variables $X_1, X_2, \ldots, X_n$ such that $g_\alpha = H(X_\alpha)$ for all $\alpha \in \Omega$.

Every linear information inequality $\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha) \geq 0$ corresponds to a linear inequality $\boldsymbol{b}^\top \boldsymbol{h} \geq 0$ in $\mathcal{H}_n$, where $\boldsymbol{b} = [b_\alpha,\ \alpha \in \Omega]^\top$ is a column vector whose components are indexed by $\alpha \in \Omega$ and $\boldsymbol{h} = [H(X_\alpha),\ \alpha \in \Omega]^\top$. Hence, for simplicity, an information inequality will usually be written in the form $\boldsymbol{b}^\top \boldsymbol{h} \geq 0$.

Let Γ_n^* [10] be the set of all entropy functions. This set plays an important role in information theory (see Theorem 2.1). It is a subset of $\mathcal{H}_n$ and it has a very complex structure. For $n \geq 3$, it is not even closed [11]. It was proved in [11] that $\overline{\Gamma}_n^*$, the closure of Γ_n^*, is a closed convex cone. Thus, $\overline{\Gamma}_n^*$ is much more manageable than Γ_n^*, and for many applications, it is sufficient to consider $\overline{\Gamma}_n^*$.

Theorem 2.1 [10]: An information inequality

$$\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha) \geq 0$$

is valid if and only if for all $\boldsymbol{h} \in \Gamma_n^*$, $\boldsymbol{b}^\top \boldsymbol{h} \geq 0$.

Theorem 2.1 has the following important consequence. Since $\{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{b}^\top \boldsymbol{h} \geq 0\}$ is closed and convex, an information inequality $\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha) \geq 0$ is valid if and only if for all $\boldsymbol{h} \in \overline{\Gamma}_n^*$, $\boldsymbol{b}^\top \boldsymbol{h} \geq 0$. In other words, the validity of an information inequality $\boldsymbol{b}^\top \boldsymbol{h} \geq 0$ depends only on Γ_n^* (or simply $\overline{\Gamma}_n^*$) and $\{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{b}^\top \boldsymbol{h} \geq 0\}$. Thus, if Γ_n^* (or $\overline{\Gamma}_n^*$) has an explicit characterization, then the information inequality can be proved or disproved by comparing the two corresponding regions. Hence, the study of the underlying structure of Γ_n^* (and $\overline{\Gamma}_n^*$) is fundamental in information theory. Although it is proved that $\overline{\Gamma}_n^*$ is a closed convex cone, $\overline{\Gamma}_n^*$ is not yet fully characterized for $n > 3$ [12].

Manuscript received April 14, 1999; revised December 4, 2001. The work of T. H. Chan was supported in part by The Chinese University of Hong Kong Postgraduate Student Grants for Overseas Academic Activities.

T. H. Chan was with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. He is now with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: terence@comm.utoronto.ca).

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

Communicated by I. Csiszár, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(02)05174-X.

III. Group-Characterizable Entropy Functions

Recall that the region Γ_n^* consists of all the entropy functions in $\mathcal{H}_n$ for n random variables. As a first step toward establishing the relation between entropy and groups, we discuss in this section entropy functions in Γ_n^* which can be described by a finite group G and subgroups $G_1, G_2, \ldots, G_n$. Such entropy functions are said to be *group-characterizable*. For simplicity, all the groups in this paper are finite.

Let $\alpha \in \Omega$ and $G_i, i \in \alpha$ be subgroups of a finite group G. The intersection of the subgroups, also a subgroup of G, is denoted by G_α. Let aG_α be the left cosets of G_α in G. By Lagrange's theorem, the left cosets of G_α in G partition G into $\frac{|G|}{|G_\alpha|}$ subsets, each of which is of size equal to $|G_\alpha|$.

Suppose we have a collection of left cosets a_iG_i for $i \in \alpha$. If $\bigcap_{i\in\alpha} a_iG_i$ is nonempty, say $b \in \bigcap_{i\in\alpha} a_iG_i$, then

$$\bigcap_{i\in\alpha} a_iG_i = \bigcap_{i\in\alpha} bG_i \tag{1}$$

$$= b\bigcap_{i\in\alpha} G_i \tag{2}$$

where the last step is easy to verify. Thus, the intersection $\bigcap_{i\in\alpha} a_iG_i$ is either empty or is of size $|G_\alpha|$.

Given subgroups $G_1, G_2, \ldots, G_n$ of a group G, let $\boldsymbol{h} \in \mathcal{H}_n$ be defined by $h_\alpha = \log \frac{|G|}{|G_\alpha|}$ for all nonempty subsets α of $\mathcal{N}$. Then $\boldsymbol{h}$ is called group-characterizable by $(G, G_1, \ldots, G_n)$.

Theorem 3.1: If $\boldsymbol{h}$ is group-characterizable, then it is an entropy function, i.e., $\boldsymbol{h} \in \Gamma_n^*$.

Proof: Let Λ be a discrete random variable defined on the sample space G with uniform distribution. For any $i \in \mathcal{N}$, let random variable X_i be a function of Λ such that $X_i = aG_i$ if $\Lambda = a$. Let α be a nonempty subset of $\mathcal{N}$. Then

$$\Pr(X_i = a_iG_i\colon i \in \alpha) = \frac{|\bigcap_{i\in\alpha} a_iG_i|}{|G|} \tag{3}$$

$$= \begin{cases} \frac{|G_\alpha|}{|G|}, & \text{if } \bigcap_{i\in\alpha} a_iG_i \neq \emptyset \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Thus, it is trivial to see that the entropy of $(X_i, i \in \alpha)$ is $\log \frac{|G|}{|G_\alpha|} = h_\alpha$. Hence, $\boldsymbol{h}$ is an entropy function. □

Example 3.1: Fix any nonempty subset β of $\mathcal{N} = \{1, 2, \ldots, n\}$, and define a vector $\boldsymbol{h} \in \mathcal{H}_n$ as follows:

$$h_\alpha = \begin{cases} 1, & \text{if } \alpha \cap \beta \neq \emptyset \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Then $(G, G_1, G_2, \ldots, G_n)$ is a group characterization of $\boldsymbol{h}$, where G is the group of modulo 2 addition, $G_i = \{0\}$ if $i \in \beta$, and $G_i = G$ otherwise. By letting $\beta = \emptyset$, we have $h_\alpha = 0$ for all $\alpha \in \Omega$. Thus, we see that $(G, G_1, G_2, \ldots, G_n)$ is a group characterization of the origin of $\mathcal{H}_n$, with $G = G_1 = \cdots = G_n$.

Example 3.2: Let $\boldsymbol{x}$ be the following 2×6 matrix

$$\boldsymbol{x} = \begin{bmatrix} a, & a, & a, & b, & b, & b \\ c, & c, & d, & c, & c, & d \end{bmatrix}. \tag{6}$$

For any nonempty subset α of $\{1, 2\}$, $\boldsymbol{x}_\alpha$ is the submatrix of $\boldsymbol{x}$ by extracting the rows of $\boldsymbol{x}$ indexed by α. For $j = 1, \ldots, 6$, $x_{\alpha,j}$ is the jth column of the submatrix $\boldsymbol{x}_\alpha$.

Let G be the group of all permutations on $\{1, 2, 3, 4, 5, 6\}$, G_i be the subgroup of G such that

$$\sigma[\boldsymbol{x}_i] = [x_{i,\sigma(1)}, x_{i,\sigma(2)}, \ldots, x_{i,\sigma(6)}] = \boldsymbol{x}_i \tag{7}$$

for all $\sigma \in G_i$ and $i = 1, 2$. Since $G_{1,2} = G_1 \bigcap G_2$, it can be checked easily that $G_{1,2}$ is the subgroup of G such that

$$\sigma[\boldsymbol{x}_{1,2}] = [x_{\{1,2\},\sigma(1)}, x_{\{1,2\},\sigma(2)}, \ldots, x_{\{1,2\},\sigma(r)}] = \boldsymbol{x}_{1,2}. \tag{8}$$

It can be checked easily that G_1 is a subgroup of G of order $3!3!$, G_2 is a subgroup of G of order $4!2!$, and $G_{1,2}$ is a subgroup of G of order $2!1!2!1!$. Then we can construct a group characterizable function $\boldsymbol{h}$ defined by

$$h_1 = \log \frac{6!}{3!\,3!}, \qquad h_2 = \log \frac{6!}{4!\,2!}, \qquad h_{1,2} = \log \frac{6!}{2!\,1!\,2!\,1!}. \tag{9}$$

IV. A Group Characterization of $\overline{\Gamma}_n^*$

Since all function values of group-characterizable functions are rational, there are only countably infinitely many group-characterizable functions. However, in general, there are uncountably infinitely many entropy functions. Hence, the number of group-characterizable functions are substantially less than the number of entropy functions. Although the number of group-characterizable functions is comparatively small, it turns out that the set of all group-characterizable functions is almost good enough to characterize the region Γ_n^*, as we will see next.

Definition 4.1: Define the following region in $\mathcal{H}_n$:

$$\Upsilon_n = \{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{h} \text{ is group-characterizable}\}. \tag{10}$$

By Theorem 3.1, if $\boldsymbol{h}$ is group-characterizable, then $\boldsymbol{h} \in \Gamma_n^*$. This implies that $\Upsilon_n \subset \Gamma_n^*$. We will prove as a corollary of the next theorem that $\overline{\text{con}}(\Upsilon_n)$, the convex closure of Υ_n, is in fact equal to $\overline{\Gamma}_n^*$, the closure of Γ_n^*.

Theorem 4.1: For any $\boldsymbol{h} \in \Gamma_n^*$, there exists a sequence $\{\boldsymbol{f}^{(r)}\}$ in Υ_n such that $\lim_{r\to\infty} \frac{1}{r}\boldsymbol{f}^{(r)} = \boldsymbol{h}$.

We need the following lemma to prove this theorem. The proof of the following lemma can be found in [3, p. 282].

Lemma 4.1: Let X be a random variable with finite sample space $\mathcal{X}$ and its probability distribution $p(x)$ be rational (i.e., $p(x)$ is a rational number for all $x \in \mathcal{X}$). Without loss of generality, assume $p(x)$ is a rational number with denominator q for all $x \in \mathcal{X}$. Then for $r = q, 2q, 3q, \ldots$

$$H(X) - |\mathcal{X}|\frac{\log(r+1)}{r} \leq \frac{1}{r}\log\frac{r!}{\prod_{x\in\mathcal{X}}(rp(x))!} \leq H(X). \tag{11}$$

Hence, as a corollary

$$\lim_{r\to\infty}\frac{1}{r}\log\frac{r!}{\prod_x (rp(x))!} = H(X). \tag{12}$$

Proof of Theorem 4.1: For any $\boldsymbol{h} \in \Gamma_n^*$, there exists a collection of random variables $X_1, X_2, \ldots, X_n$ such that

$$h_\alpha = H(X_\alpha) \tag{13}$$

for all nonempty subsets α of $\mathcal{N}$. We first consider the special case that $|\mathcal{X}_i| < \infty$ for all $i \in \mathcal{N}$ and the joint distribution of $X_1, X_2, \ldots, X_n$ is rational. We want to show that there exists a sequence $\{\boldsymbol{f}^{(r)}\}$ in Υ_n such that $\lim_{r\to\infty} \frac{1}{r}\boldsymbol{f}^{(r)} = \boldsymbol{h}$.

For any $\alpha \in \Omega$, let Q_α be the marginal distribution of X_α. Assume without loss of generality that for any nonempty subset α of $\mathcal{N}$ and for all $a \in \mathcal{X}_\alpha$, $Q_\alpha(a)$ is a rational number with denominator q.

For each $r = q, 2q, 3q, \ldots$, fix an $n \times r$ matrix $\boldsymbol{x}$

$$\boldsymbol{x} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,r} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,r} \end{bmatrix} \tag{14}$$

such that for all $(a_1, \ldots, a_n) \in \mathcal{X}_\mathcal{N}$, the number of columns in $\boldsymbol{x}$ being $[a_1 \ldots a_n]^\top$ is $rQ_\mathcal{N}(a)$. The existence of such a matrix is guaranteed by all the values of the joint distribution of $X_\mathcal{N}$ being rational numbers with denominator q.

For any nonempty subset α of $\{1, \ldots, n\}$, $\boldsymbol{x}_\alpha$ is the submatrix of $\boldsymbol{x}$ obtained by extracting the rows of $\boldsymbol{x}$ indexed by α. For $j = 1, \ldots, r$, $x_{\alpha,j}$ is the jth column of the submatrix $\boldsymbol{x}_\alpha$. It is easy to prove that for all $(a_i\colon i \in \alpha) \in \mathcal{X}_\alpha$, the number of columns in $\boldsymbol{x}_\alpha$ being the transpose of $(a_i\colon i \in \alpha)$ is $rQ_\alpha(a_i\colon i \in \alpha)$, where Q_α is the marginal distribution of random variables X_i for $i \in \alpha$.

Let G be the group of all permutations on $\{1, \ldots, r\}$. The group G depends on r, but in order to keep the notation simple, we do not state this dependency explicitly. For any $i \in \{1, \ldots, n\}$, let G_i be the subgroup of G such that

$$\sigma[\boldsymbol{x}_i] = [x_{i,\sigma(1)}, x_{i,\sigma(2)}, \ldots, x_{i,\sigma(r)}] = \boldsymbol{x}_i \tag{15}$$

for all $\sigma \in G_i$ and $i = 1, \ldots, n$. Since $G_\alpha = \bigcap_{i\in\alpha} G_i$, it can be checked easily that G_α is the subgroup of G such that

$$\sigma[\boldsymbol{x}_\alpha] = [x_{\alpha,\sigma(1)}, x_{\alpha,\sigma(2)}, \ldots, x_{\alpha,\sigma(r)}] = \boldsymbol{x}_\alpha. \tag{16}$$

Since for all $(a_i\colon i \in \alpha) \in \mathcal{X}_\alpha$, the number of columns in $\boldsymbol{x}_\alpha$ being $[a_i\colon i \in \alpha]^\top$ is $rQ_\alpha(a_i\colon i \in \alpha)$, it can be checked easily that $|G_\alpha| = \prod_{a\in\mathcal{X}_\alpha}(rQ_\alpha(a))!$. By Lemma 4.1

$$\lim_{r\to\infty} \frac{1}{r} \log \frac{|G|}{|G_\alpha|} = H(X_\alpha) = h_\alpha. \tag{17}$$

Let $\boldsymbol{f}^{(r)}$ be defined by

$$f_\alpha^{(r)} = \log \frac{|G|}{|G_\alpha|} \tag{18}$$

for all $\alpha \in \Omega$. Then $\boldsymbol{f}^{(r)} \in \Upsilon_n$ by construction, and hence,

$$\lim_{r\to\infty} \frac{1}{r}\boldsymbol{f}^{(r)} = \boldsymbol{h}. \tag{19}$$

In general, for any $\boldsymbol{h} \in \Gamma_n^*$, we can construct a sequence $\{\boldsymbol{h}^{(k)}\}$ in Γ_n^* such that $\lim_{k\to\infty} \boldsymbol{h}^{(k)} = \boldsymbol{h}$, where $\boldsymbol{h}^{(k)}$ is the entropy function of a collection of random variables $X_1, \ldots, X_n$ with finite sample space and a rational joint probability distribution. This completes the proof of the theorem. □

Corollary 4.1: $\overline{\mathrm{con}}(\Upsilon_n) = \overline{\Gamma}_n^*$

Proof: First of all, $\Upsilon_n \subset \Gamma_n^*$. By taking convex closure, we have $\overline{\mathrm{con}}(\Upsilon_n) \subset \overline{\mathrm{con}}(\Gamma_n^*)$. Since $\overline{\Gamma}_n^*$ is convex, $\overline{\mathrm{con}}(\Gamma_n^*) = \overline{\Gamma}_n^*$, and we have $\overline{\mathrm{con}}(\Upsilon_n) \subset \overline{\Gamma}_n^*$. On the other hand, we have shown in Example 3.1 that the origin of $\mathcal{H}_n$ has a group characterization and therefore is in Υ_n. It then follows from Theorem 4.1 that $\overline{\Gamma}_n^* \subset \overline{\mathrm{con}}(\Upsilon_n)$. Hence, we conclude that $\overline{\Gamma}_n^* = \overline{\mathrm{con}}(\Upsilon_n)$, completing the proof. □

V. Information Inequalities and Group Inequalities

As we have discussed in Section II, a linear information inequality

$$\boldsymbol{b}^\top \boldsymbol{h} \geq 0 \tag{20}$$

always holds if and only if

$$\overline{\Gamma}_n^* \subset \{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{b}^\top \boldsymbol{h} \geq 0\}. \tag{21}$$

In other words, all linear information inequalities are fully characterized by $\overline{\Gamma}_n^*$. We also have proved at the end of the last section that $\overline{\mathrm{con}}(\Upsilon_n) = \overline{\Gamma}_n^*$. Since $\Upsilon_n \subset \Gamma_n^* \subset \overline{\Gamma}_n^*$, if (21) holds, then

$$\Upsilon_n \subset \{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{b}^\top \boldsymbol{h} \geq 0\}. \tag{22}$$

On the other hand, since $\{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{b}^\top \boldsymbol{h} \geq 0\}$ is close and convex, by taking convex closure in (22), we obtain

$$\overline{\Gamma}_n^* = \overline{\mathrm{con}}(\Upsilon_n) \subset \{\boldsymbol{h} \in \mathcal{H}_n\colon \boldsymbol{b}^\top \boldsymbol{h} \geq 0\}. \tag{23}$$

Therefore, (21) and (22) are equivalent.

For each $\boldsymbol{h} \in \Upsilon_n$

$$h_\alpha = \log \frac{|G|}{|G_\alpha|} \tag{24}$$

for all nonempty subset α of $\mathcal{N}$ for some finite group G and subgroups $G_1, G_2, \ldots, G_n$. Hence, the information inequality $\sum_{\alpha\in\mathcal{H}_n} b_\alpha H(X_\alpha) \geq 0$ holds for all random variables $X_1, X_2, \ldots, X_n$ if and only if the corresponding group inequality

$$\sum_{\alpha\in\Omega} b_\alpha \log \frac{|G|}{|G_\alpha|} \geq 0 \tag{25}$$

holds for all finite group G and subgroups $G_1, G_2, \ldots, G_n$. In other words, for every linear information inequality, there is a corresponding group inequality in the form as in (25), and *vice versa*. Therefore, inequalities in information theory can be proved by methods in group theory, and inequalities in group theory can be proved by methods in information theory.

Example 5.1: Let G_1 and G_2 be subgroups of a finite group G with group operation $\circ$. Define

$$G_1 \circ G_2 = \{a \circ b\colon a \in G_1 \text{ and } b \in G_2\}. \tag{26}$$

It is easy to prove that $|G_1 \circ G_2| = \frac{|G_1||G_2|}{|G_1 \cap G_2|}$. As a corollary

$$|G_{13} \circ G_{23}| = \frac{|G_{13}||G_{23}|}{|G_{13} \cap G_{23}|} \tag{27}$$

$$= \frac{|G_{13}||G_{23}|}{|G_{123}|}. \tag{28}$$

Since $G_{13} \circ G_{23}$ is a subset of G_3

$$|G_{13} \circ G_{23}| = \frac{|G_{13}||G_{23}|}{|G_{123}|} \leq |G_3|. \tag{29}$$

Rearranging the terms, we obtain

$$\log \frac{|G|}{|G_{13}|} + \log \frac{|G|}{|G_{23}|} \geq \log \frac{|G|}{|G_3|} + \log \frac{|G|}{|G_{123}|}. \tag{30}$$

This group inequality corresponds to the information inequality

$$H(X_1, X_3) + H(X_2, X_3) \geq H(X_3) + H(X_1, X_2, X_3). \tag{31}$$

Hence,

$$H(X_1, X_3) + H(X_2, X_3) \geq H(X_3) + H(X_1, X_2, X_3)$$

or, equivalently, $I(X_1; X_2|X_3) \geq 0$, is a valid information inequality for all random variables X_1, X_2, X_3.

The above example shows how an information inequality can be proved by methods in group theory. In fact, all so-called Shannon-type inequalities are consequences of the nonnegativity of conditional mutual information [10]. Therefore, all Shannon-type inequalities can be proved by methods in group theory.

On the other hand, information inequalities can also give rise to new inequalities in group theory by virtue of our result. This is discussed in the next example.

Example 5.2: Recently, the following highly nontrivial information inequality, which cannot be deduced by directly invoking the basic Shannon inequalities, has been proved in [12]:

$$\begin{aligned} &H(X_1) + H(X_2) + 2H(X_1, X_2) + 4H(X_3) + 4H(X_4) \\ &\quad + 5H(X_1, X_3, X_4) + 5H(X_2, X_3, X_4) \\ &\leq 6H(X_3, X_4) + 4H(X_1, X_3) + 4H(X_1, X_4) \\ &\quad + 4H(X_2, X_3) + 4H(X_2, X_4). \end{aligned} \tag{32}$$

This information inequality corresponds to the group inequality

$$\begin{aligned} &\log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|} + 2\log \frac{|G|}{|G_{12}|} + 4\log \frac{|G|}{|G_3|} + 4\log \frac{|G|}{|G_4|} \\ &\quad + 5\log \frac{|G|}{|G_{134}|} + 5\log \frac{|G|}{|G_{234}|} \\ &\leq 6\log \frac{|G|}{|G_{34}|} + 4\log \frac{|G|}{|G_{13}|} + 4\log \frac{|G|}{|G_{14}|} \\ &\quad + 4\log \frac{|G|}{|G_{23}|} + 4\log \frac{|G|}{|G_{24}|}. \end{aligned} \tag{33}$$

Upon rearranging the terms, we obtain

$$\begin{aligned} &|G_{34}|^6|G_{13}|^4|G_{14}|^4|G_{23}|^4|G_{24}|^4 \\ &\quad \leq |G_1||G_2||G_3|^4|G_4|^4|G_{12}|^2|G_{134}|^5|G_{234}|^5. \end{aligned} \tag{34}$$

The meaning of this inequality and its implications in group theory are yet to be understood.

VI. Conclusion

Information inequalities play a crucial role in the proofs of almost all converse coding theorems in source and channel coding problems. In essence, they govern the impossibility in information theory. However, due to lack of tools, to find new information inequalities is an extremely difficult task. In this paper, we have identified a class of group-characterizable entropy functions. The correspondence between group-characterizable entropy functions and their group characterizations provides an algebraic approach to proving information inequalities and *vice versa*. Since group theory is a well-studied branch in mathematics, it may be possible that we can use some existing results in this field to attack the corresponding problem in information theory.

Acknowledgment

The authors are grateful to Prof. František Matúš. His detailed comments and suggestions have greatly improved the readability of the paper.

References

[1] T. H. Chan, "A combinatorial approach to information inequalities," *Commun. Inform. and Syst.*, vol. 1, no. 3, pp. 241–254.
[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
[4] A. W. Ingleton, "Characterization of matroids," in *Combinatorial Mathematics and It's Application*. New York: Academic, 1971.
[5] F. Matúš, "Probabilistic conditional independence structures and matroid theory: Background," *Int. J. General Syst.*, vol. 22, pp. 185–196, 1994.
[6] N. Pippenger, "What are the laws of information theory?," in *Proc. 1986 Specific Problems on Communication and Computation Conf.*, Palo Alto, CA, Sept. 3–5, 1986.
[7] M. Studený, "Conditional independence relations have no finite complete characterization," in *Trans. 11th Prague Conf. Information Theory, Statistical Decision Functions and Random Processes*, ser. B. Prague, Czech Republic: Academia, 1992, pp. 377–396.
[8] R. W. Yeung, "A new outlook on Shannon's information measures," *IEEE Trans. Inform. Theory*, vol. 37, pp. 466–474, May 1991.
[9] 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, May 1992.
[10] R. W. Yeung, "A framework for linear information inequality," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1924–1934, Nov. 1997.
[11] Z. Zhang and R. W. Yeung, "A non-Shannon-type conditional information inequality," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1982–1986, Nov. 1997.
[12] ——, "On the characterization of entropy function via information inequalities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1440–1452, July 1998.
[13] R. W. Yeung and Y.-O. Yan. ITIP. [Online]. Available: http://personal.ie.cuhk.edu.hk/~ITIP