A Non-Shannon-Type Conditional Inequality of Information Quantities

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 \( \alpha \) 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 \( \Omega \). 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 is two-alternative. These properties are the so-called basic information inequalities of Shannon’s information measures. An entropy function can be viewed as a \( 2^n \)-dimensional vector where the coordinates are indexed by the subsets of the ground set \( \{1, 2, \cdots, n\} \). As introduced in [4], \( \Gamma^*_\alpha \) stands for the cone in \( \mathbb{R}^{2^n} \) consisting of all vectors which have all these properties. Let \( \Gamma^*_\alpha \) be the set of all \( 2^n \)-dimensional vectors which correspond to the entropy functions of some sets of \( n \) discrete random variables. A fundamental information-theoretic problem is whether or not \( \Gamma^*_\alpha = \Gamma_n \). Here \( \Gamma^*_\alpha \) stands for the closure of the set \( \Gamma^*_\alpha \). In this correspondence, we show that \( \Gamma^*_\alpha \) is a convex cone, \( \Gamma^*_2 = \Gamma_2 \), \( \Gamma^*_3 \neq \Gamma_3 \), but \( \Gamma^*_4 = \Gamma_3 \). For four random variables, we have discovered a conditional inequality which is not implied by the basic information inequalities of the same set of random variables. This lends an evidence to the plausible conjecture that \( \Gamma^*_n \neq \Gamma_n \) for \( n > 3 \).

Index Terms—Entropy, \( I \)-Measure, information inequalities, mutual information.
Let $\mathcal{F}_n$ be the set of all functions defined on $2^N_n$ taking values in $[0, \infty)$. Define

$$
\Gamma_n \overset{\text{def}}{=} \{ F_n \in \mathcal{F}_n : F_n(\emptyset) = 0; \alpha \subseteq \beta \Rightarrow F_n(\alpha) \leq F_n(\beta); 
\forall \alpha, \beta \in 2^N_n, F_n(\alpha) + F_n(\beta) \geq F_n(\alpha \cup \beta) + F_n(\alpha \cap \beta) \}.
$$

(12)

Since for each $F_n \in \Gamma_n$, $F_n(\emptyset)$ takes the constant value 0, $F_n$ is completely specified by the set of values it takes on the nonempty subsets (a total of $2^n - 1$) of $N_n$. Thus each $F_n$ is represented by a point in $\mathbb{R}^{2^n-1}$ with the coordinates being the values of $F_n$ on the nonempty subsets of $N_n$, and $\Gamma_n$ is represented by a pyramid in the positive quadrant.

Apparently, for any $\Omega_n = \{ X_i : i = 1, \cdots, n \}$, the associated entropy function $H \in \Gamma_n$. This means that the set $\Gamma_n$ characterizes some of the properties of the entropy function. A natural question is whether this set "fully" characterizes the entropy function [3]. To make the question more precise, we introduce the following definitions.

Definition 1: A function $F_n \in \Gamma_n$ is constructible if and only if there exists a set of $n$ jointly distributed random variables with finite entropies such that the joint entropy function $H$ for these random variables satisfies $H = F_n$.

Define

$$
\Gamma_n^* = \{ F_n \in \Gamma_n : F_n \text{ is constructible} \}.
$$

Evidently, $\Gamma_n^* \subseteq \Gamma_n$. The question is whether $\Gamma_n^*$ is equal to $\Gamma_n$.

Let us first consider $n = 2$ and let $X$ and $Y$ be the two random variables. To see that $\Gamma_2^* = \Gamma_2$, we only have to consider $X = (T_1, T_2)$ and $Y = (T_2, T_3)$ for independent random variables $T_1, T_2, T_3$ and observe that $H(T_i)$ for $i = 1, 2, 3$ can take any nonnegative value. The details are omitted here.

For $n = 3$, consider the following information measures of three random variables $X$, $Y$, and $Z$: 

1) $U_1(X; Y; Z) = H(X|Y, Z)$,
2) $U_2(X; Y; Z) = H(Y|X, Z)$,
3) $U_3(X; Y; Z) = H(Z|X, Y)$;

Since the joint entropies involving $X$, $Y$, and $Z$ are an invertible linear transformation of $U_i(X; Y; Z)$: $i = 1, \cdots, 7$ [3], $\Gamma_3$ can be converted into a region in $\mathbb{R}^7$ with $U_i(X; Y; Z)$ as the coordinates. Denoting $U_i(X; Y; Z)$ by $u_i$, this region is given as follows:

$$
\Psi \overset{\text{def}}{=} \{ \pi \in \mathbb{R}^7 : u_i \geq 0, i = 1, 2, 3, 4, 5, 6; 
u_7 + u_1 \geq 0, j = 4, 5, 6 \}.
$$

We now show that $\Gamma_3^* \neq \Gamma_3$. The point $(0, 0, 0, 0, a, a, -a)$ for any $a > 0$ is in $\Psi$. It is easy to check that such a point corresponds to $H(X|Y, Z) = H(Y|Z, X) = H(Z|X, Y) = 0$ and

$$
I(X; Y) = I(Y; Z) = I(X; Z) = 0
$$

i.e., each random variable is a function of the other two, and the three random variables are pairwise-independent. Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be the supports of $X$, $Y$, and $Z$, respectively. For any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, since $X$ and $Y$ are independent, we have

$$
p(x, y) = p(x)p(y) > 0.
$$

(15)

Since $Z$ is a function of $X$ and $Y$, there is a unique $z \in \mathcal{Z}$ such that $p(x, y, z) = p(x, y) = p(x)p(y) > 0$.

(16)

Now since $Y$ is a function of $X$ and $Z$, and $X$ and $Z$ are independent, we can write

$$
p(x, y, z) = p(x, z) = p(x)p(z).
$$

(17)

Equating (16) and (17), we have

$$
p(y) = p(z).
$$

(18)

Now consider any $y' \in \mathcal{Y}$ such that $y' \neq y$. Since $Y$ and $Z$ are independent, we have

$$
p(y', z) = p(y')p(z) > 0.
$$

(19)

Since $X$ is a function of $Y$ and $Z$, there is a unique $x' \in \mathcal{X}$ such that

$$
p(x', y', z) = p(y', z) = p(y')p(z) > 0.
$$

(20)

Therefore, $Y$ must have a uniform distribution on its support. The same can be proved for $X$ and $Z$. Now

$$\begin{align*}
H(X) &= H(X|Y, Z) + I(X; Y|Z) + I(X; Z|Y) + I(X; Y; Z) \\
&= a
\end{align*}
$$

and, similarly, $H(Y) = H(Z) = a$. Then the only values that $a$ can take are $\log M$, where $M$ (a positive integer) is the cardinality of the supports of $X$, $Y$, and $Z$. This proves that $\Gamma_3^* \neq \Gamma_3$, and thus $\Gamma_n^* \neq \Gamma_n$ in general.

Since $\Gamma_n^* \neq \Gamma_n$ for $n \geq 3$, further characterization of $\Gamma_n^*$ is necessary. We see from the example above that a full characterization of $\Gamma^*$ could be extremely difficult. Instead, we are motivated to attack an alternative problem. We first introduce the following definition.

Definition 2: A function $F_n \in \Gamma_n$ is asymptotically constructible if and only if there exist sets of $n$ random variables $\Omega_n$ with finite entropies for $k = 1, \cdots, s$ such that the joint entropy functions $H_k$ associated with $\Omega_n$ satisfy $\lim_{n \to \infty} H_k = F_n$.

Obviously, a function $F_n$ is asymptotically constructible if and only if $F_n \in \Gamma_n^*$, the closure of the set $\Gamma_n^*$.

We now present the results in this correspondence. The proofs of these results are in the next section.
Theorem 1: $\Gamma_n$ is a convex cone.

Theorem 2:

$$\Gamma'_n = \Gamma_n.$$  \hspace{1cm} (26)

Han [1] has found the smallest cone that contains $\Gamma'_n$. From Theorem 1, we can see that what Han has found is $\Gamma'_n$. That is, Han's result and Theorem 1 imply Theorem 2. Also, Theorem 2 is a consequence of the theorem in Matúš [2]. Nevertheless, in order to make the correspondence more readable, we will give an independent proof of Theorem 2 in the next section.

It is still unknown whether this theorem can be generalized. That is, we still do not know whether it is true that for any $n$

$$\Gamma'_n = \Gamma_n.$$  \hspace{1cm} (27)

To give a definite answer to this question is a fundamental problem in information theory. Nevertheless, we prove the following conditional inequality of Shannon's information measures which may lend some insight to this open problem.

Theorem 3: For four random variables $\Omega_4 = \{X, Y, Z, U\}$

$$I(X; Y) = I(X; Y | Z) = 0$$ \hspace{1cm} (28)

implies

$$I(U; Z) - I(U; Z | X) - I(U; Z | Y) \leq 0.$$ \hspace{1cm} (29)

The inequality in Theorem 3 cannot be derived by using the basic inequalities of the same set of random variables. This can be seen as follows: The following function is in $\Gamma_4$, but it does not satisfy the new conditional inequality:

$$F(\phi) = 0, \quad F(X) = F(Y) = F(Z) = F(U) = 2a > 0$$
$$F(X, Y) = 4a, \quad F(X, U) = F(X, Z) = F(Y, U)$$
$$= F(Y, Z) = F(Z, U) = 3a$$

It is easy to check that $F \in \Gamma_4$. If $F$ is constructible, then there exist random variables $X, Y, Z, U$ such that $H = F$. Then

$$I(Z; U) = F(U) + F(Z) - F(Z, U) = a > 0$$
$$I(X; Y) = -F(X, Y) + F(X) + F(Y) = 0$$
$$I(X; Y | Z) = -F(X, Y, Z) - F(Z) + F(X, Z) + F(Y, Z) = 0$$
$$I(Z; U | Y) = F(Y, U) + F(Y, Z) - F(Y) - F(Y, Z, U) = 0$$

and

$$I(Z; U | X) = F(X, U) + F(X, Z) - F(X) - F(X, Z, U) = 0.$$  \hspace{1cm} (30)

From Theorem 3, this is impossible. Therefore, there do not exist four jointly distributed random variables $X, Y, Z, U$ such that the joint entropy function of the four random variables is the same as $F$. That is, $F \not\in \Gamma_4$.

Theorem 3 provides an example of a conditional information inequality that is not implied by the basic information inequalities of the same set of random variables. It also lends an evidence for the following conjecture.

Conjecture 1:

$$\Gamma'_4 \neq \Gamma_4.$$  \hspace{1cm} (31)
Then it follows from the convexity of $\overline{\Psi}$ that

$$(0, 0, 0, a, a, a, -a) \in \overline{\Psi}$$

for all $a > 0$.

For an arbitrary $\bar{\tau} \in \Psi$, if $u_7 > 0$, we take $X = (V_1, V_4, V_5, V_7)$, $Y = (V_2, V_4, V_6, V_7)$, and $Z = (V_3, V_5, V_6, V_7)$ where $V_i; i = 1, \cdots, 7$ are seven independent random variables with entropies $H(V_i) = u_i$. Then, we can see that

$$U_i(X, Y, Z) = u_i, \quad \text{for } i = 1, \cdots, 7. \quad (35)$$

That is, $\bar{\tau}$ is constructible. If $u_7 < 0$, let

$$\bar{\tau} = (0, 0, 0, -u_7, -u_7, -u_7, u_7)$$

then $\bar{\tau} \equiv \tau$ is a nonnegative vector, and we can find random variables $X_1, Y_1, Z_1$ such that

$$U_i(X_1, Y_1, Z_1) = w_i. \quad (36)$$

For any $\epsilon > 0$, we can find $X_2, Y_2, Z_2$ which are independent of $(X_1, Y_1, Z_1)$ such that

$$|U_i(X_2, Y_2, Z_2) - t_i| < \epsilon, \quad \text{for } i = 1, \cdots, 7 \quad (37)$$

because $\bar{\tau}$ is asymptotically constructible. Let $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$, and $Z = (Z_1, Z_2)$. We have for $i = 1, \cdots, 7$

$$\tilde{U}_i(X, Y, Z) = U_i(X_1, Y_1, Z_1) + U_i(X_2, Y_2, Z_2).$$

This implies

$$|\tilde{U}_i(X, Y, Z) - u_i| < \epsilon, \quad \text{for } i = 1, \cdots, 7. \quad (38)$$

This proves that $\tau$ is asymptotically constructible. That is,

$$\overline{\Psi} = \Psi$$

which implies

$$\Gamma_3 = \Gamma_3. \quad (39)$$

This completes the proof of Theorem 2.

**Proof of Theorem 3**

$$I(Z; U) = I(Z; U|X) - I(Z; U|Y)$$

\begin{align*}
&= \sum_{x, y, z, u} p(x, y, z, u) \log \frac{p(z, u)p(x, z)p(y, z)p(y, u)}{p(z)p(u)p(x)p(y)p(x, z, u)p(y, z, u)} \\
&\quad - \sum_{x, y, z, u; p(x, y, z, u) > 0} \log \frac{p(z, u)p(x, z)p(y, z)p(y, u)}{p(z)p(u)p(x)p(y)p(x, z, u)p(y, z, u)} \\
&= \sum_{x, y, z, u; p(x, y, z, u) > 0} \tilde{p}(x, y, z, u) \log \frac{p(x, z)p(x, y)p(y, z)p(y, u)}{p(z)p(u)p(x)p(y)p(x, z, u)p(y, z, u)} \quad (40)
\end{align*}

We claim that

$$\sum_{x, y, z, u; p(x, y, z, u) > 0} \tilde{p}(x, y, z, u) \log \frac{p(x, z)p(x, y)p(y, z)p(y, u)}{p(z)p(u)p(x)p(y)p(x, z, u)p(y, z, u)}$$

where

$$\tilde{p}(x, y, z, u) \equiv \begin{cases} 
\frac{p(x, z)p(y, z)}{p(z)}, & \text{if } p(z) \neq 0 \\
0, & \text{otherwise}.
\end{cases} \quad (41)$$

The last equality is justified by observing that if one of $p(x), p(y), p(z), p(u)$ is zero, then $\tilde{p}(x, y, z, u) = 0$, and the quadruple $(x, y, z, u)$ is not involved in the summation and, otherwise,

$$\tilde{p}(x, y, z, u) = \frac{p(x, z)p(x, y)p(y, z)p(y, u)}{p(z)p(u)p(x)p(y)p(x, z, u)p(y, z, u)}.$$

To prove the result, we need to show that $\tilde{p}(x, y, z, u)$ is a probability distribution. If this is proven, then the conclusion of the theorem is merely a consequence of the inequality

$$D(\mu || \nu) \geq 0$$

where $D(\mu || \nu)$ stands for the divergence of two distributions $\mu$ and $\nu$ defined on the same set. Notice that for $x, y, z$ such that $p(z) > 0$

$$p(x, y, z) = \frac{p(x, z)p(y, z)}{p(z)}$$

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

$$p(x, y) = p(x)p(y)$$
by $I(X; Y) = 0$. Therefore,
\[
\sum_{x, y, z, u} \hat{p}(x, y, z, u) = \sum_{x, y, z, u} \hat{p}(x, y, z, u) = \sum_{x, y, z, u} \frac{p(x, z)p(x, u)p(y, z)p(y, u)}{p(z)p(u)} = \sum_{x, y, z, u} \frac{p(x, y)p(u)p(y, u)}{p(u)} = 1.
\]
The theorem is proved.

ACKNOWLEDGMENT

The authors wish to thank F. Matúš for his careful reading of the manuscript. R. W. Yeung wishes to thank B. Hajek for his useful discussion.

REFERENCES


Capacity of Fading Channels with Channel Side Information

Andrea J. Goldsmith, Member, IEEE, and Pravin P. Varaiya, Fellow, IEEE

Abstract—We obtain the Shannon capacity of a fading channel with channel side information at the transmitter and receiver, and at the receiver alone. The optimal power adaptation in the former case is “water-pouring” in time, analogous to water-pouring in frequency for time-invariant frequency-selective fading channels. Inverting the channel results in a large capacity penalty in severe fading.

Index Terms—Capacity, channel side information, fading channels, power adaptation.

I. INTRODUCTION

The growing demand for wireless communication makes it important to determine the capacity limits of fading channels. In this correspondence, we obtain the capacity of a single-user fading channel when the channel fade level is tracked by both the transmitter and receiver, and by the receiver alone. In particular, we show that the fading-channel capacity with channel side information at both the transmitter and receiver is achieved when the transmitter adapts its power, data rate, and coding scheme to the channel variation. The optimal power allocation is a “water-pouring” in time, analogous to the water-pouring used to achieve capacity on frequency-selective fading channels [1], [2].

We show that for independent and identically distributed (i.i.d.) fading, using receiver side information only has a lower complexity and the same approximate capacity as optimally adapting to the channel, for the three fading distributions we examine. However, for correlated fading, not adapting at the transmitter causes both a decrease in capacity and an increase in encoding and decoding complexity. We also consider two suboptimal adaptive techniques: channel inversion and truncated channel inversion, which adapt the transmit power but keep the transmission rate constant. These techniques have very simple encoder and decoder designs, but they exhibit a capacity penalty which can be large in severe fading. Our capacity analysis for all of these techniques neglects the effects of estimation error and delay, which will generally degrade capacity.

The tradeoff between these adaptive and nonadaptive techniques is therefore one of both capacity and complexity. Assuming that the channel is estimated at the receiver, the adaptive techniques require a feedback path between the transmitter and receiver and some complexity in the transmitter. The optimal adaptive technique uses variable-rate and power transmission, and the complexity of its decoding technique is comparable to the complexity of decoding a sequence of additive white Gaussian noise (AWGN) channels in parallel. For the nonadaptive technique, the code design must make use of the channel correlation statistics, and the decoder complexity is proportional to the channel decorrelation time. The optimal adaptive technique always has the highest capacity, but the increase relative

Manuscript received December 25, 1995; revised February 26, 1997. This work was supported in part by an IBM Graduate Fellowship and in part by the California PATH program, Institute of Transportation Studies, University of California, Berkeley, CA.
A. J. Goldsmith is with the California Institute of Technology, Pasadena, CA 91125 USA.
P. P. Varaiya is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720 USA.