Proving and Disproving Information Inequalities

Siu-Wai Ho*, Chee Wei Tan†, and Raymond W. Yeung‡
siuwai.ho@unisa.edu.au*, cheewtan@cityu.edu.hk† and whyeung@ie.cuhk.edu.hk‡

***Abstract*—Proving an information inequality is a crucial step in establishing the converse results in coding theorems. However, an information inequality involving many random variables is difficult to be proved manually. In [1], Yeung developed a framework that uses linear programming for verifying linear information inequalities. Under this framework, this paper considers a few other problems that can be solved by using Lagrange duality and convex approximation. We will demonstrate how linear programming can be used to find an analytic proof of an information inequality. The way to find a shortest proof is explored. When a given information inequality cannot be proved, the sufficient conditions for a counterexample to disprove the information inequality are found by linear programming.**

I. Introduction

In information theory, we may need to prove different kinds of information inequalities in different problems. To prove an information inequality is non-trivial when it involves more than three random variables. For example, we may want to check the correctness of the following two inequalities:

$$\begin{aligned} &I(A;B|CD)+I(B;D|AC)\\ &\quad\leq\quad I(A;B|D)+I(B;D|A)+H(A)+I(B;D|C) \end{aligned} \tag{1}$$

and

$$\begin{aligned} &I(A;B|CD)+I(B;D|AC)\\ &\quad\leq\quad I(A;B|D)+I(B;D|A)+H(AB|D). \end{aligned} \tag{2}$$

To prove information inequalities, the author in [1] developed a framework for linear information inequalities and provided a software package known as Information Theoretic Inequality Prover (ITIP) [2]. ITIP and Xitip (similar to ITIP but it uses a C-based linear programming solver instead [3]) are widely-used software packages that can automatically verify an information inequality on a computer. For example, if we use ITIP to verify (1) and (2), we will get "Not provable by ITIP" and "True", respectively. However, after we know that (2) is true, we still need an analytic proof. An analytic proof is the formal way to verify an information inequality and, more importantly, it also provides us further insights about the inequality of interest. One important insight is about the necessary and sufficient conditions for the equality to hold. Consider (2) again which can be proved by showing

$$\begin{aligned} 0\leq&H(B|A,C,D)+H(A|B,C,D)+I(B;C|A)\\ &+I(A;B|D)+I(A;C|D) \qquad (3)\\ =&I(A;B|D)+I(B;D|A)+H(AB|D)\\ &-I(A;B|CD)-I(B;D|AC), \qquad (4) \end{aligned}$$

where (4) can be easily verified by expressing all the qualities on both sides in terms of joint entropies. Then we can further deduce that the equality in (2) holds if and only if all the qualities on the right side of (3) are equal to 0. In Section II, we will demonstrate how to use a linear program to obtain a proof like the one in (3)–(4).

Recently, [4] solved an open problem by showing that the rate-regions of the exact-repair regeneration codes can be different from the rate-regions of the functional-repair regeneration codes. To prove an information inequality involving 16 random variables in this work, it is very hard to manually construct a proof. The author has tailor-made a linear program to find the required information inequality for the converse result and its proof. This interesting result demonstrates that we may need a machine to find a proof when we are dealing with a large-size problem involving many random variables.

Now, when an information inequality cannot be proved by ITIP, there are two possible cases. The first case is that the inequality is indeed true but to verify it is outside the capability of ITIP. The existence of such inequalities were first found in [5][6] and an infinite number of such inequalities were later reported in [7] (see also [8]–[10]). These inequalities are called non-Shannon type inequalities, which will be explicitly defined in Section IV. Another case is that the given inequality is in fact not true in general. In other words, there exist counterexamples which can disprove the inequality. It is important to distinguish between these two cases. In Section IV, we will use linear programming to obtain some hints for constructing counterexamples when the given inequality is not provable.

This paper is organized as follows. Section II shows how to find a proof of an information inequality through linear programming. In Section III, we introduce the problem formulation of finding the shortest proof of information inequalities, and propose a convex approximation algorithm that is motivated by the sparse recovery technique in the compressive sensing literature that yields a feasible proof. Finally, we will illustrate how Lagrange duality can help us to disprove an information inequality in Section IV.

Siu-Wai Ho is with the Institute for Telecommunications Research, University of South Australia. Chee Wei Tan is with the Department of Computer Science, City University of Hong Kong. R. W. Yeung is with the Institute of Network Coding and the Department of Information Engineering, The Chinese University of Hong Kong, N.T., Hong Kong, and with the Key Laboratory of Network Coding Key Technology and Application and Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China.

This work was partially funded by a grant from the University Grants Committee of the Hong Kong Special Administrative Region (Project No. AoE/E-02/08) and Key Laboratory of Network Coding, Shenzhen, China (ZSDY20120619151314964).

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

II. Linear Information Inequalities Framework

Consider n random variables $(X_1, X_2, \ldots, X_n)$ and the (joint) entropies of all the non-empty subset of these random variables form a column vector $\mathbf{h}$. For example, if $n = 3$, then

$$\begin{aligned}\mathbf{h} = (&H(X_1), H(X_2), H(X_3), H(X_1, X_2), H(X_2, X_3),\\ &H(X_1, X_3), H(X_1, X_2, X_3)).\end{aligned} \tag{5}$$

The coefficients related to an information inequality can be denoted by a column vector $\mathbf{b}$. To illustrate, continue the example of $\mathbf{h}$ in (5). Then, the information inequality $-H(X_1, X_3) + H(X_1, X_2, X_3) \geq 0$ is denoted by $\mathbf{b}^T\mathbf{h} \geq 0$ with $\mathbf{b} = [0\ 0\ 0\ 0\ 0\ -1\ 1]^T$. Due to the nonnegativity of Shannon's information measures, we know that $\mathbf{h}$ must satisfy certain inequalities. For example,

$$H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2) \geq 0, \tag{6}$$

$$H(X_1, X_2) - H(X_2) = H(X_1|X_2) \geq 0. \tag{7}$$

A special subset of all these inequalities due to the nonnegativity of Shannon's information measures is defined as the elemental inequalities [11, P. 340] which are denoted by

$$\mathbf{Dh} \geq \mathbf{0} \tag{8}$$

in this paper. Obviously, any vector $\mathbf{h}$ must satisfy (8). An important property about this set is that all the inequalities due to the nonnegativity of Shannon's information measures, like $H(X_1) \geq 0$, $H(X_1, X_2|X_3) \geq 0$, etc., can be obtained as a conic combination (also known as a nonnegative linear combination) of the elemental inequalities. Therefore, an information inequality can be proved by using the nonnegativity of Shannon's information measures if and only if the inequality can be implied by the elemental inequalities. An information inequality, which is implied by the nonnegativity of Shannon's information measures, is called *Shannon-type inequality*.

Very often we want to prove an information inequality $\mathbf{b}^T\mathbf{h} \geq 0$ subject to a given set of equality constraints $\mathbf{Eh} = 0$. When there is no equality constraint, this set is simply empty.

The linear combination of the joint entropies $\mathbf{b}^T\mathbf{h}$ is a valid information inequality if and only if it is always nonnegative [1]. Consider the following linear program:

$$\begin{array}{ll}\text{minimize} & \mathbf{b}^T\mathbf{h}\\ \text{subject to} & \mathbf{Dh} \geq \mathbf{0},\\ & \mathbf{Eh} = \mathbf{0},\\ \text{variables:} & \mathbf{h},\end{array} \tag{9}$$

and its Lagrange dual problem (also a linear program):

$$\begin{array}{ll}\text{maximize} & \mathbf{y}^T\mathbf{0}\\ \text{subject to} & \mathbf{D}^T\mathbf{y} \leq \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu},\\ & \mathbf{y} \geq \mathbf{0},\\ \text{variables:} & \mathbf{y},\ \boldsymbol{\mu}.\end{array} \tag{10}$$

Remark: The optimal value of (9) is 0 if $\mathbf{b}^T\mathbf{h} \geq 0$ is a Shannon-type inequality, and is $-\infty$ otherwise [11, Theorem 14.4].

From the fundamental (strong duality) theorem of linear programming, i.e., the duality gap is zero in linear programs, [12, Theorem 4.4], the optimal values of (9) and (10) are equal. Furthermore, we have the following optimality results which is used to show an analytical proof for any Shannon-type inequality.[1]

Theorem 1: The inequality $\mathbf{b}^T\mathbf{h} \geq 0$ is a Shannon-type inequality if and only if

$$\mathbf{D}^T\mathbf{y}^* = \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu}^*, \tag{11}$$

where $[\mathbf{y}^{*T}\ \boldsymbol{\mu}^{*T}]^T$ gives the optimal solution to (10).

Proof: If (11) is true, then $\mathbf{b}^T\mathbf{h} = (\mathbf{y}^*)^T\mathbf{Dh} - (\boldsymbol{\mu}^*)^T\mathbf{Eh}$. Hence, $\mathbf{b}^T\mathbf{h} \geq 0$ follows from that $\mathbf{Dh} \geq \mathbf{0}$, $\mathbf{Eh} = \mathbf{0}$ and $\mathbf{y}^* \geq 0$.

If $\mathbf{b}^T\mathbf{h} \geq 0$ is a Shannon-type inequality, then the optimization problem in (9) has an optimal value equal to 0 [11, Theorem 14.4]. Using the Karush-Kuhn-Tucker (KKT) conditions, the respective optimal primal and dual solutions $\mathbf{h}^*$ and $[\mathbf{y}^{*T}\ \boldsymbol{\mu}^{*T}]^T$ satisfy

$$\begin{aligned}\mathbf{D}^T\mathbf{y}^* &= \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu}^*, \quad \text{(Stationarity of Lagrangian)}\\ \mathbf{y}^{*T}\mathbf{Dh}^* &= \mathbf{0}, \quad \text{(Complementarity slackness)}\\ \mathbf{Dh}^* \geq \mathbf{0}, \mathbf{Eh}^* &= \mathbf{0}, \mathbf{h}^* \geq \mathbf{0}, \quad \text{(Primal feasibility)}\\ \mathbf{D}^T\mathbf{y}^* \leq \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu}^*, &\mathbf{y}^* \geq \mathbf{0}. \quad \text{(Dual feasibility)}\end{aligned}$$

Hence, (11) follows because it is the stationarity of the Lagrangian in the KKT conditions. ∎

Then we can have the following direct consequence.

Corollary 2: If $\mathbf{b}^T\mathbf{h} \geq 0$ is a Shannon-type inequality, then let $[\mathbf{y}^{*T}\ \boldsymbol{\mu}^{*T}]^T$ be the optimal solution to (10) and an analytical proof can be written as follows. For all feasible $\mathbf{h}$,

$$\begin{aligned}\mathbf{b}^T\mathbf{h} &= (\mathbf{y}^*)^T\mathbf{Dh} - (\boldsymbol{\mu}^*)^T\mathbf{Eh} &(12)\\ &\geq 0, &(13)\end{aligned}$$

where (13) follows from that $\mathbf{y}^* \geq 0$, $\mathbf{Dh} \geq \mathbf{0}$ and $\mathbf{Eh} = \mathbf{0}$.

So we can modify ITIP to get the following sample output.

Example 1: Suppose we want to prove $H(U) \leq H(R)$ subject to $I(U; X) = 0$ and $H(U|RX) = 0$. A modified ITIP gives the following output:

$ITIP('H(U) <= H(R)',' I(U;X) = 0',' H(U|RX) = 0')$

True. The inequality follows from

$$\begin{aligned}-H(U) + H(R) &= (-H(U, X) + H(U, R, X)) +\\ &\quad (H(R) + H(X) - H(R, X)) +\\ &\quad \{-H(U) - H(X) + H(U, X)\} +\\ &\quad \{H(R, X) - H(U, R, X)\}\\ &\geq 0,\end{aligned}$$

where $(\cdot)$ is non-negative as it is either conditional entropy or conditional mutual information. All $\{\cdot\}$ are equal to 0 due to the given constraints. Equality holds iff all $(\cdot)$ are equal to 0.

[1] The idea of using the Lagrange duality to find an analytical proof has also been used in [4].

III. The Shortest Proof

We have seen that an information inequality can be proved by expressing it into a linear combination of elemental inequalities. However, there can be more than one way to express the same information inequality. For example,

$$\begin{aligned} &H(XYZ)-H(X|YZ)-H(Y|XZ)-H(Z|XY)\\ =&I(X;Y)+I(X;Z|Y)+I(Y;Z|X) &(14)\\ =&I(X;Z)+I(X;Y|Z)+I(Y;Z|X) &(15)\\ =&I(Y;Z)+I(X;Z|Y)+I(X;Y|Z). &(16)\end{aligned}$$

So we can prove $H(XYZ)-H(X|YZ)-H(Y|XZ)-H(Z|XY)\geq 0$ by proving any of the equalities in (14)–(16). In fact,

$$\begin{aligned} &H(XYZ)-H(X|YZ)-H(Y|XZ)-H(Z|XY)\\ =&0.8(I(X;Y)+I(X;Z|Y)+I(Y;Z|X))+\\ &0.1(I(X;Z)+I(X;Y|Z)+I(Y;Z|X))+\\ &0.1(I(Y;Z)+I(X;Z|Y)+I(X;Y|Z)) &(17)\end{aligned}$$

is also true. Obviously, the proof given in (17) is longer than necessity and also not succinctly elegant.

The *shortest proof* of an information inequality is considered as the proof involving the least number of elemental inequalities. Let us consider the following combinatorial optimization problem:

$$\begin{array}{ll}\text{minimize} & \|[\mathbf{y}^T\ \boldsymbol{\mu}^T]^T\|_0\\ \text{subject to} & \mathbf{D}^T\mathbf{y}=\mathbf{b}+\mathbf{E}^T\boldsymbol{\mu},\ \mathbf{y}\geq\mathbf{0},\\ \text{variables:} & \mathbf{y},\ \boldsymbol{\mu},\end{array} \tag{18}$$

where $\|\mathbf{x}\|_0$ is the cardinality or the number of nonzero components in the vector $\mathbf{x}$. Now, (18) is a combinatorial problem that is generally hard to solve. Suppose there exists a feasible dual variable $[\mathbf{y}^T\ \boldsymbol{\mu}^T]^T$. Consider the following nonempty and bounded polyhedron:

$$P=\{[\mathbf{y}^T\ \boldsymbol{\mu}^T]^T\in\mathcal{R}^{m+q}\mid \mathbf{y}\geq\mathbf{0},\ \mathbf{D}^T\mathbf{y}=\mathbf{b}+\mathbf{E}^T\boldsymbol{\mu}\},$$

where m and q are the number of rows in $\mathbf{D}$ and $\mathbf{E}$, respectively. From Lemma 2 in [13], (18) has a solution that is a vertex of P. So rather than solving (18) directly as an optimization problem, we can also enumerate all the vertices of P. Then the vertex that has the least cardinality is guaranteed to be a solution to (18). A practical pivot-based algorithm has been proposed in [14] to find the v vertices of a polyhedron in $\mathbf{R}^d$ defined by a non-degenerate system of m inequalities in $O(mdv)$ time and $O(md)$ space. But vertex enumeration can be computationally inefficient especially in large problems.

The marriage of the information inequalities framework and recent developments in convex approximation for sparse recovery offers new directions to explore. Now, we consider the following convex approximation problem to tackle (18):

$$\begin{array}{ll}\text{minimize} & \mathbf{1}^T\mathbf{y}+\|\boldsymbol{\mu}\|_1\\ \text{subject to} & \mathbf{D}^T\mathbf{y}=\mathbf{b}+\mathbf{E}^T\boldsymbol{\mu},\ \mathbf{y}\geq\mathbf{0},\\ \text{variables:} & \mathbf{y},\ \boldsymbol{\mu}.\end{array} \tag{19}$$

Since (19) is a (non-smoothed) convex optimization problem, it can in principle be solved by an interior-point method numerically [15] or the homotopy algorithm in [16]. However, (19) can be further transformed to the following equivalent problem that is a linear program:

$$\begin{array}{ll}\text{minimize} & \mathbf{1}^T\mathbf{y}+\mathbf{1}^T\mathbf{z}\\ \text{subject to} & \mathbf{D}^T\mathbf{y}=\mathbf{b}+\mathbf{E}^T\boldsymbol{\mu},\ \mathbf{y}\geq\mathbf{0},\\ & -\mathbf{z}\leq\boldsymbol{\mu}\leq\mathbf{z},\\ \text{variables:} & \mathbf{y},\ \boldsymbol{\mu},\ \mathbf{z}.\end{array} \tag{20}$$

Remarks: 1) The interior-point method or the Simplex method (e.g., see [12]) can be used to solve (20) exactly. From an algorithmic perspective, it is well-known that the Simplex method has a worst-case exponential time complexity, while the interior point method, though with a worst-case polynomial time complexity, needs to be enhanced with the ability to locate an exact vertex solution [17]. Without the enhancement, the solution of (20) may lead us to a proof like (17) that lacks aesthetics.

2) A numerically exact solution is crucial in solving (20) than an asymptotically or approximate solution obtained by a numerical solver since the goal is to yield an exact proof as opposed to having a numerical value. In other words, we are looking for $[\mathbf{y}^*\boldsymbol{\mu}^*]$ such that we can claim "=" in (12). Using "$\approx$" in (12) is unacceptable.

Note that the convex approximation problem in (19) can be different from the original problem in (18) as shown below.

Example 2: Consider

$$\begin{aligned} &H(YZ)-H(Y|XZ)-H(Z|XY)+I(X;Y|Z) &(21)\\ =&I(X;Z)+2I(X;Y|Z)+I(Y;Z|X) &(22)\\ =&I(X;Y)+I(X;Y|Z)+I(Y;Z|X)+I(X;Z|Y). &(23)\end{aligned}$$

In order to prove the nonnegativity of (21), we just need to show the equality in either (22) or (23). The summation of coefficients in both (22) and (23) are the same and equal to 4. However, (22) involves less number of Shannon's information measures. Therefore, the optimization problems in (18) and (19) can have different results for some inequalities.

Now, we consider a slight modification of the linear program in (20) to obtain further results. Suppose we know that an information inequality is correct if the set of equality constraints $\mathbf{Eh}=\mathbf{0}$ is assumed. It is interesting to know the minimal set of equality constraints that is required. In other words, we want to know the minimal number of rows in $\mathbf{E}$ which are sufficient to prove the same inequality. To solve this problem, we just need to remove $\mathbf{y}^T$ in the objective function in (18). Since this problem is also NP-hard, we will consider the approximation of this problem by replacing $\mathbf{1}^T\mathbf{y}+\mathbf{1}^T\mathbf{z}$ by $\mathbf{1}^T\mathbf{z}$ in the linear program in (20). The following example revisits the implication problem in [11, Ex. 14.9] to demonstrate a much shorter proof. The modified linear program shows that not all the equality constraints in [11, (14.73)] are necessary.

Example 3: Under the assumption that

$$\begin{aligned} 0&=I(X;Y|Z)=I(X;T|Y)=I(X;Z|Y)\\ &=I(X;T|Z)=I(X;Z|T), &(24)\end{aligned}$$

we want to prove $I(X;Y|T) = 0$. The modified linear program shows

$$\begin{aligned} -I(X;Y|T) =& I(X;T|Z) + I(X;Z|Y) - \\ & I(X;Y|Z) - I(X;T|Y) - I(X;Z|T) \quad (25) \\ \geq& 0, \quad (26) \end{aligned}$$

where (26) follows from using only $I(X;Y|Z) = I(X;T|Y) = I(X;Z|T) = 0$ in (24). This proof can provide us further insights. By rearranging the terms in (25), $I(X;T|Z) = I(X;Z|Y) = 0$ can be implied by just assuming $I(X;Y|Z) = I(X;T|Y) = I(X;Z|T) = 0$. This further explains why some equality constraints in (24) are redundant.

Before the end of this section, we want to remark that a proof can be shorter if $\mathbf{D}$ is expanded to include all the inequalities due to the nonnegativity of Shannon's information measures. This expansion is equivalent to expanding $\mathbf{D}$ by adding some positive linear combinations of the rows in $\mathbf{D}$. If this new $\mathbf{D}$ is used in the linear program in (9), the same optimal solution will be obtained and hence, we can still obtain Theorem 1. However, we can now obtain a shorter proof. For example,

$$\begin{aligned} & H(XYZ) - I(Y;Z|X) - H(Z|X,Y) \quad (27) \\ =& H(X|Y,Z) + H(Y|X,Z) + I(X;Y) + I(X;Z|Y) \quad (28) \\ =& H(X) + H(Y|X,Z). \quad (29) \end{aligned}$$

When we prove the nonnegativity of (27), a longer proof in (28) is obtained if $\mathbf{D}$ contains only elemental inequalities. However, a shorter proof can be shown in (29) if the linear program *learns* more possibilities of conic combination through the matrix $\mathbf{D}$. However, we need to pay the price for a larger size in $\mathbf{D}$ that affects the computational time to solve the linear programs.

IV. Disproving an information inequality

We have seen how to use a linear program to obtain a proof of a Shannon-type inequality. In this section, we explore how to disprove an information inequality. Suppose we are given an information inequality $\mathbf{b}^T\mathbf{h} \geq 0$ which cannot be proved by the linear program using $\mathbf{D}$. In other words, the linear program in (9) does not give 0 [11, Theorem 14.4]. It is still unclear whether a) $\mathbf{b}^T\mathbf{h} < 0$ for some $\mathbf{h}$ or b) $\mathbf{b}^T\mathbf{h} \geq 0$ is indeed true for all feasible $\mathbf{h}$, but proving $\mathbf{b}$) is beyond the capability of the linear program using $\mathbf{D}$ (in the case, $\mathbf{b}$) cannot be expressed as $\mathbf{D}^T\mathbf{y}$ for some $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^T\mathbf{h} \geq 0$ is called a *non-Shannon type inequality* [6]). Therefore, we need a counterexample to explicitly disprove $\mathbf{b}^T\mathbf{h} \geq 0$.

Suppose we can find $\mathbf{h}$ such that $\mathbf{Dh} \geq \mathbf{0}$ and $\mathbf{b}^T\mathbf{h} < 0$. This is still insufficient to be a counterexample because there may not exist any joint distribution $P_{X_1,X_2,\ldots,X_n}$ that realizes $\mathbf{h}$. An example is shown in [11, (15.85)]. In general, there is no known algorithm to construct $P_{X_1,X_2,\ldots,X_n}$ from any given $\mathbf{h}$, and hence, it seems that finding $\mathbf{h}$ may not give any immediate help. In the following, we will show that a linear program for finding $\mathbf{h}$ is still useful. Its dual problem will give us the sufficient conditions for the counterexample to disprove $\mathbf{b}^T\mathbf{h} \geq 0$.

Suppose the last element in $\mathbf{h}$ denotes the joint entropy $H(X_1, X_2, \ldots, X_n)$. Let $\mathbf{e}$ be a row vector such that $\mathbf{e}$ and $\mathbf{h}$ have the same length. Define $\mathbf{e} = [0\ 0\ \cdots\ 0\ 1]^T$. Consider the following linear program:

$$\begin{array}{rl} \text{minimize} & \mathbf{b}^T\mathbf{h} \\ \text{subject to} & \mathbf{Dh} \geq \mathbf{0}, \\ & \mathbf{Eh} = \mathbf{0}, \\ & \mathbf{e}^T\mathbf{h} = 1, \\ \text{variables:} & \mathbf{h}, \end{array} \quad (30)$$

and its dual problem (also a linear program):

$$\begin{array}{rl} \text{minimize} & -\gamma \\ \text{subject to} & \mathbf{D}^T\mathbf{y} \leq \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu} + \mathbf{e}\gamma, \\ & \mathbf{y} \geq \mathbf{0}, \\ \text{variables:} & \mathbf{y},\ \boldsymbol{\mu}, \gamma. \end{array} \quad (31)$$

Suppose the Simplex method is used to solve the linear program in (9) for an information inequality $\mathbf{b}^T\mathbf{h}$ which is not always true. Then the result of (9) is $-\infty$ from [11, P.344]. However in the linear program in (30), we have further fixed $H(X_1, X_2, \ldots, X_n) = 1$ by the extra constraint $\mathbf{e}^T\mathbf{h} = 1$. Together with the inequality constraints in $\mathbf{Dh} \geq 0$, all the elements in any feasible $\mathbf{h}$ (i.e., all the (joint) entropies) are upper bounded by 1. Therefore, $\mathbf{b}^T\mathbf{h}$ is a bounded negative value in the linear program in (30).

We have discussed the difficulty of using the optimal $\mathbf{h}^*$, which is obtained by solving the linear program in (30), to construct a counterexample. The following theorem states that the optimal $\mathbf{y}^*$ in the dual problem in (31) provides us a list of functional dependencies and (conditional) independencies between $\{X_1, X_2, \ldots, X_n\}$. This list gives hints on explicitly constructing a counterexample for disproving $\mathbf{b}^T\mathbf{h} \geq 0$.

Theorem 3: Let $[\mathbf{y}^{*T}\ \boldsymbol{\mu}^{*T}]$ be the optimal dual solutions for the problem in (31). If there exists a joint distribution $P_{X_1,X_2,\ldots,X_n}$ such that its entropy vector $\tilde{\mathbf{h}}$ satisfies

$$\mathbf{y}^{*T}\mathbf{D}\tilde{\mathbf{h}} = \boldsymbol{\mu}^{*T}\mathbf{E}\tilde{\mathbf{h}} = \mathbf{0}, \text{ and } \mathbf{e}^T\tilde{\mathbf{h}} = 1, \quad (32)$$

then $\mathbf{b}^T\tilde{\mathbf{h}} < 0$ and $P_{X_1,\ldots,X_n}$ is a counterexample to disprove $\mathbf{b}^T\mathbf{h} \geq 0$.

Proof: Using the KKT conditions, the optimal dual solutions $[\mathbf{y}^{*T}\ \boldsymbol{\mu}^{*T}\ \gamma^*]^T$ for the problem in (31) satisfy

$$\mathbf{D}^T\mathbf{y}^* = \mathbf{b} + \mathbf{E}^T\boldsymbol{\mu}^* + \mathbf{e}^T\gamma^*. \text{ (Stationarity of Lagrangian)}$$

Together with (32), we have

$$\begin{aligned} \mathbf{b}^T\tilde{\mathbf{h}} =& \mathbf{b}^T\tilde{\mathbf{h}} + \boldsymbol{\mu}^{*T}\mathbf{E}\tilde{\mathbf{h}} + \gamma^*\mathbf{e}^T\tilde{\mathbf{h}} - \gamma^* \quad (33) \\ =& \mathbf{y}^{*T}\mathbf{D}\tilde{\mathbf{h}} - \gamma^* \quad (34) \\ =& -\gamma^*. \quad (35) \end{aligned}$$

From the fundamental (strong duality) theorem of linear programming, i.e., the duality gap is zero in linear programs, [12, Theorem 4.4], the optimal value of (30) and (31) are equal.

As the optimal value of (30) is negative, $-\gamma^*$ and $\mathbf{b}^T\tilde{\mathbf{h}}$ are both negative so that the theorem is proved. ■

In order to satisfy $\mathbf{y}^{*T}\mathbf{D}\tilde{\mathbf{h}} = \mathbf{0}$ in (32), we need to find out the positive elements in $\mathbf{y}^*$. Their corresponding rows in $\mathbf{D}$ tell us the extra equality constraints which are used together with $\mathbf{E}\tilde{\mathbf{h}} = \mathbf{0}$ and $\mathbf{e}^T\tilde{\mathbf{h}} = 1$ to construct a counterexample.

Example 4: A sample output from a modified ITIP:
$ITIP('I(X;Y) <= 0.9H(Y)')$
Not provable by ITIP.
It can be disproved by a probability distribution satisfying all the following Shannon's information measures equal to zero:

$$H(X|Y), H(Y|X). \tag{36}$$

From the above output from ITIP, we can deduce that the counterexample should be chosen as $X = Y$ to disprove $I(X;Y) <= 0.9H(Y)$.

Example 5: The modified ITIP can be used to disprove (1):
$ITIP('I(A;B|CD) + I(B;D|AC) <= I(A;B|D) + I(B;D|A) + H(A) + I(B;D|C)')$
Not provable by ITIP.
It can be disproved by a probability distribution satisfying all the following Shannon's information measures equal to zero:

$$\begin{aligned}&H(A|B,C,D), H(C|A,B,D), H(D|A,B,C), I(A;B|C),\\ &I(A;B|D), I(A;C|D), I(A;D), I(B;C|A), I(B;D|A),\\ &I(B;D|C), I(C;D|A).\end{aligned} \tag{37}$$

From the above output from ITIP, we can deduce the following counterexample. Let X, Y and Z be three independent binary random variables with entropy equal to 1. Let $(A,B,C,D) = (X \oplus Y, X, Y \oplus Z, Z)$ where $\oplus$ denotes "exclusive or". Then it is easy to check that $I(A;B|D) + I(B;D|A) + H(A) + I(B;D|C) - I(A;B|CD) - I(B;D|CA) = -1 < 0$.

In the above examples, ITIP gives some equality constraints that help us to construct the counterexample for an invalid information inequality. There are some tricks to construct the example from the output of ITIP. We can consider a set of auxiliary random variables which are mutually independent. By considering the condition entropy in (37), we can obtain some hints about the number of auxiliary random variables. Those mutual information and conditional mutual information in (37) can tell us where "exclusive or" should be used. Of course, the joint entropy of the random variables cannot be zero. Otherwise, the equality constraints are satisfied but the information inequality cannot be disproved.

Then it is natural to ask: Is it always possible to construct an example from the given equality constraints? Unfortunately, the answer is 'No' due to the existence of the non-Shannon type inequalities. See the following example.

Example 6: A sample output from ITIP:
$ITIP('2I(C;D) <= I(A;B) + I(A;C,D) + 3I(C;D|A) + I(C;D|B)')$
Not provable by ITIP.
It can be disproved by a probability distribution satisfying all the following Shannon's information measures equal to zero:

$$\begin{aligned}&H(C|ABD), I(A;B), I(A;B|C), I(A;C|D), I(A;D|C),\\ &I(B;C|D), I(C;D|A), I(C;D|B), I(C;D|AB).\end{aligned} \tag{38}$$

In the above example, one should not be able to find a joint distribution P_{ABCD} satisfying (38) with $H(ABCD) \neq 0$. Indeed, it is impossible to find such P_{ABCD} because $2I(C;D) \leq I(A;B) + I(A;C,D) + 3I(C;D|A) + I(C;D|B)$ is a non-Shannon type inequality which is always true for all P_{ABCD} [6]. Therefore, the counterexample does not exist.

Remarks: 1) If we assume that all the quantities in (36) are equal to 0, then we can prove $I(X;Y) \geq 0.9H(Y)$, i.e., the opposite of what we wanted to disprove. This property also holds for the inequalities in Examples 5 and 6. This can be seen from Theorem 3.
2) The result from Example 6 can lead us to the following constrained non-Shannon type inequality.

Proposition 4: If all Shannon's information measures in (38) are equal to zero, then $H(A,B,C,D) \leq 0$.

REFERENCES

[1] R. W. Yeung, *A framework for linear information inequalities*, IEEE Trans. on Information Theory, vol. 43, pp. 1924-1934, Nov 1997.
[2] R. W. Yeung and Y. O. Yan, *Information Theoretic Inequality Prover (ITIP)*, Matlab program software package available at: http://home.ie.cuhk.edu.hk/~ITIP
[3] R. Pulikkoonattu and S. Diggavi, *a ITIP-based C program software package*, available at: http://xitip.epfl.ch
[4] C. Tian, "Characterizing the rate-region of the (4,3,3) exact-repair regenerating codes". To appear in *JSAC on Communication Methodologies for the Next-Generation Storage Systems*, 2014.
[5] Z. Zhang and R. W. Yeung, "A non-Shannon-type conditional inequality of information quantities," *IEEE Trans. Info. Theory*, Vol. 43, pp. 1982–1986, 1997.
[6] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. Info. Theory*, Vol. 44, pp. 1440–1452, 1998.
[7] F. Matúš, "Infinitely Many Information Inequalities," in *Proc. IEEE International Symposium on Information Theory (ISIT2007)*, Nice, France, pp. 2101–2105, June, 2007
[8] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-Shannon information inequalities," *Proc. IEEE International Symposium on Information Theory, (ISIT2006)*, Seattle, Washington, pp. 233–236, July, 2006.
[9] R. Lněnička, "On the tightness of the Zhang-Yeung inequality for Gaussian vectors," *Communications in Information and Systems*, vol. 3, no. 1, pp. 41-46, June 2003.
[10] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Communications in Information and Systems*, no. 2, pp. 147–166, Dec. 2002.
[11] R. W. Yeung, *Information Theory and Network Coding*, Springer, 2008.
[12] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, 1997.
[13] M. Wang, C. W. Tan, W. Xu and A. Tang, "Cost of Not Splitting in Routing: Characterization and Estimation," *IEEE/ACM Trans. on Networking*, vol. 19, No. 6, pp. 1849-1859, Dec. 2011.
[14] D. Avis and K. Fukuda, "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra," *Discrete Computational Geometry*, vol. 8, No. 3, pp. 295-313, 1992.
[15] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[16] D. L. Donoho and Y. Tsaig, "Fast Solution of l1-Norm Minimization Problems When the Solution May Be Sparse," *IEEE Trans. on Information Theory*, vol. 54, No. 11, pp. 4789-4812, Nov. 2008.
[17] S. Mehrotra, "On finding a vertex solution using interior point methods," *Linear Algebra and its Applications*, vol. 152, pp. 233-253, Jul. 1991.