

Proving Information Inequalities and Identities with Symbolic Computation

Laigang Guo, Raymond W. Yeung, and Xiao-Shan Gao

Abstract

Proving linear inequalities and identities of Shannon's information measures, possibly with linear constraints on the information measures, is an important problem in information theory. For this purpose, ITIP and other variant algorithms have been developed and implemented, which are all based on solving a linear program (LP). In particular, an identity $f = 0$ is verified by solving two LPs, one for $f \geq 0$ and one for $f \leq 0$. In this paper, we develop a set of algorithms that can be implemented by symbolic computation. Based on these algorithms, procedures for verifying linear information inequalities and identities are devised. Compared with LP-based algorithms, our procedures can produce analytical proofs that are both human-verifiable and free of numerical errors. Our procedures are also more efficient computationally. For constrained inequalities, by taking advantage of the algebraic structure of the problem, the size of the LP that needs to be solved can be significantly reduced. For identities, instead of solving two LPs, the identity can be verified directly with very little computation.

Index Terms

Entropy, mutual information, information inequality, information identity, machine proving, ITIP.

I. INTRODUCTION

In information theory, we may need to prove various information inequalities and identities that involve Shannon's information measures. For example, such information inequalities and identities play a crucial role in establishing the converse of most coding theorems. However, proving an information inequality or identity involving more than a few random variables can be highly non-trivial.

To tackle this problem, a framework for linear information inequalities was introduced in [1]. Based on this framework, the problem of verifying Shannon-type inequalities can be formulated as a linear program (LP), and a software package based on MATLAB called Information Theoretic Inequality Prover (ITIP) was developed [3]. Subsequently, different variations of ITIP have been developed. Instead of MATLAB, Xitip [4] uses a C-based linear programming solver, and it has been further developed into its web-based version, oXitip [7]. minitip [5] is a C-based version of ITIP that adopts a simplified syntax and has a user-friendly syntax checker. psitip [6] is a Python library that can verify unconstrained/constrained/existential entropy inequalities. It is a computer algebra system where random variables, expressions, and regions are objects that can be manipulated. AITIP [8] is a cloud-based platform that not only provides analytical proofs for Shannon-type inequalities but also give hints on constructing a smallest counterexample in case the inequality to be verified is not a Shannon-type inequality.

Using the above LP-based approach, to prove an information identity $f = 0$, two LPs need to be solved, one for the inequality $f \geq 0$ and the other for the inequality $f \leq 0$. Roughly speaking, the amount of computation for proving an information identity is twice the amount for proving an information inequality. If the underlying random variables exhibit certain Markov or functional dependence structures, there exist more efficient approaches to proving information identities [10][12].

The LP-based approach is in general not computationally efficient because it does not take advantage of the special structure of the underlying LP. In this paper, we take a different approach. Instead of transforming the problem into a general LP to be solved numerically, we develop algorithms that can be implemented by symbolic computation, and based on these algorithms, procedures for proving information inequalities and identities are devised. Our specific contributions are:

- 1) Analytical proofs for information inequalities and identities that are free of numerical errors can be produced.
- 2) Compared with the LP-based approach, the computational efficiency of our procedure is in general much higher.

L. Guo is with the Laboratory of Mathematics and Complex Systems (Ministry of Education), School of Mathematical Sciences, Beijing Normal University, Beijing, China. e-mail: (lguo@bnu.edu.cn).

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. e-mail: (whyung@ie.cuhk.edu.hk).

X.-S. Gao is with the Key Laboratory of Mathematics Mechanization, Institute of Systems Science, AMSS, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing, China. e-mail: (xgao@mmrc.iss.ac.cn).

3) Information identities can be proved directly with very little computation instead of having to solve 2 LPs.

The rest of the paper is organized as follows. In Section II, we present the preliminaries for information inequalities. In Section III, we develop algorithms for simplifying a set of linear inequalities subject to linear inequality and equality constraints. In Section IV, we introduce a set of variables (inspired by the theory of I -Measure [13]) that facilitates the implementation of our algorithms. In Section V, the procedures for proving information inequalities and identities are presented. Two examples are given in Section VI to illustrate our procedures. Section VII concludes the paper.

II. INFORMATION INEQUALITY PRELIMINARIES

In this section, we present some basic results related to information inequalities and their verification. For a comprehensive discussion on the topic, we refer the reader to [9, Chs. 13-15].

It is well known that all Shannon's information measures, namely entropy, conditional entropy, mutual information, and conditional mutual information are always nonnegative. The nonnegativity of all Shannon's information measures forms a set of inequalities called the *basic inequalities*. The set of basic inequalities, however, is not minimal in the sense that some basic inequalities are implied by the others. For example,

$$H(X|Y) \geq 0 \text{ and } I(X;Y) \geq 0,$$

which are both basic equalities involving random variables X and Y , imply

$$H(X) = H(X|Y) + I(X;Y) \geq 0,$$

again a basic equality involving X and Y . In order to eliminate such redundancies, the minimal subset of the basic inequalities was found in [1].

Throughout this paper, all random variables are discrete. Unless otherwise specified, all information expressions involve some or all of the random variables X_1, X_2, \dots, X_n . The value of n will be specified when necessary. Denote the set $\{1, 2, \dots, n\}$ by \mathcal{N}_n and the sequence $[1, 2, \dots, n]$ by $[n]$.

Theorem II.1. [1] *Any Shannon's information measure can be expressed as a conic combination of the following two elemental forms of Shannon's information measures:*

- i) $H(X_i | X_{\mathcal{N}_n - \{i\}})$
- ii) $I(X_i; X_j | X_K)$, where $i \neq j$ and $K \subseteq \mathcal{N}_n - \{i, j\}$.

The nonnegativity of the two elemental forms of Shannon's information measures forms a proper subset of the set of basic inequalities. The inequalities in this smaller set are called the *elemental inequalities*. In [1], the minimality of the elemental inequalities is also proved. The total number of elemental inequalities is equal to

$$m = n + \sum_{r=0}^{n-2} \binom{n}{r} \binom{n-r}{2} = n + \binom{n}{2} 2^{n-2}.$$

In this paper, inequalities (identities) involving only Shannon's information measures are referred to as information inequalities (identities). The elemental inequalities are called *unconstrained* information inequalities because they hold for all joint distributions of the random variables. In information theory, we very often deal with information inequalities (identities) that hold under certain constraints on the joint distribution of the random variables. These are called *constrained* information inequalities (identities), and the associated constraints are usually expressible as linear constraints on the Shannon's information measures. We will confine our discussion on constrained inequalities of this type.

Example II.1. *The celebrated data processing theorem asserts that for any four random variables X, Y, Z and T , if $X \rightarrow Y \rightarrow Z \rightarrow T$ forms a Markov chain, then $I(X;T) \geq I(Y;Z)$. Here, $I(X;T) \geq I(Y;Z)$ is a constrained information inequality under the constraint $X \rightarrow Y \rightarrow Z \rightarrow T$, which is equivalent to*

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

or

$$I(X;Z|Y) + I(X,Y;T|Z) = 0$$

owing to the nonnegativity of conditional mutual information. Either way, the Markov chain can be expressed a set of linear constraint(s) on the Shannon's information measures.

Information inequalities (unconstrained or constrained) that are implied by the basic inequalities are called *Shannon-type* inequalities. Most of the information inequalities that are known belong this type. However, *non-Shannon-type* inequalities do exist, e.g., [11]. See [9, Ch. 15] for a discussion.

Shannon's information measures, with conditional mutual informations being the general form, can be expressed as a linear combination of joint entropies by means of following identity:

$$I(X_G; X_{G'} | X_{G''}) = H(X_G, X_{G''}) + H(X_{G'}, X_{G''}) - H(X_G, X_{G'}, X_{G''}) - H(X_{G'}).$$

where $G, G', G'' \subseteq \mathcal{N}_n$. For the random variables X_1, X_2, \dots, X_n , there are a total of $2^n - 1$ joint entropies. By regarding the joint entropies as variables, the basic (elemental) inequalities become linear inequality constraints in $\mathbb{R}^{2^n - 1}$. By the same token, the linear equality constraints on Shannon's information measures imposed by the problem under discussion become linear equality constraints in $\mathbb{R}^{2^n - 1}$. This way, the problem of verifying a (linear) Shannon-type inequality can be formulated as a linear program (LP), which is described next.

Let \mathbf{h} be the column m -vector of the joint entropies of X_1, X_2, \dots, X_n . The set of elemental inequalities can be written as $G\mathbf{h} \geq 0$, where G is an $m \times (2^n - 1)$ matrix and $G\mathbf{h} \geq 0$ means all the components of $G\mathbf{h}$ are nonnegative. Likewise, the constraints on the joint entropies can be written as $Q\mathbf{h} = 0$. When there is no constraint on the joint entropies, Q is assumed to have zero row. The following theorem enables a Shannon-type inequality to be verified by solving an LP.

Theorem II.2. [1] $\mathbf{b}^\top \mathbf{h} \geq 0$ is a Shannon-type inequality under the constraint $Q\mathbf{h} = 0$ if and only if the minimum of the problem

$$\text{Minimize } \mathbf{b}^\top \mathbf{h}, \text{ subject to } G\mathbf{h} \geq 0 \text{ and } Q\mathbf{h} = 0$$

is zero.

III. LINEAR INEQUALITIES AND RELATED ALGORITHMS

In this section, we will develop some algorithms for simplifying a linear inequality set constrained by a linear equality set. These algorithms will be used as building blocks for the procedures to be developed in Section V for proving information inequalities and identities.

We will start by discussing some notions pertaining to linear inequality sets and linear equality sets. Then we will establish some related properties that are instrumental for developing the aforementioned algorithms.

Let $\mathbf{x} = [x_1, x_2, \dots, x_n]$, and let $\mathbb{R}_h[\mathbf{x}]$ be the set of all homogeneous linear polynomials in \mathbf{x} with real coefficients. In this paper, unless otherwise specified, we assume that all inequality sets have the form $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$, with $f_i \neq 0$ and $f_i \in \mathbb{R}_h[\mathbf{x}]$, and all the equality sets have the form $E_{\tilde{f}} = \{\tilde{f}_i = 0, i \in \mathcal{N}_{\tilde{m}}\}$ with $\tilde{f}_i \neq 0$ and $\tilde{f}_i \in \mathbb{R}_h[\mathbf{x}]$.

For a given set of polynomials $P_f = \{f_i, i \in \mathcal{N}_m\}$ and the corresponding set of inequalities $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$, and a given set of polynomials $P_{\tilde{f}} = \{\tilde{f}_i, i \in \mathcal{N}_{\tilde{m}}\}$ and the corresponding set of equalities $E_{\tilde{f}} = \{\tilde{f}_i = 0, i \in \mathcal{N}_{\tilde{m}}\}$, where f_i and \tilde{f}_i are polynomials in \mathbf{x} , we write $S_f = \mathcal{R}(P_f)$, $P_f = \mathcal{R}^{-1}(S_f)$, $E_{\tilde{f}} = \tilde{\mathcal{R}}(P_{\tilde{f}})$ and $P_{\tilde{f}} = \tilde{\mathcal{R}}^{-1}(E_{\tilde{f}})$.

Definition III.1. Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$ and $S_{f'} = \{f'_i \geq 0, i \in \mathcal{N}_{m'}\}$ be two inequality sets, and $E_{\tilde{f}}$ and $E_{\tilde{f}'}$ be two equality sets. We write $S_{f'} \subseteq S_f$ if $\mathcal{R}^{-1}(S_{f'}) \subseteq \mathcal{R}^{-1}(S_f)$, and $E_{\tilde{f}'} \subseteq E_{\tilde{f}}$ if $\tilde{\mathcal{R}}^{-1}(E_{\tilde{f}'}) \subseteq \tilde{\mathcal{R}}^{-1}(E_{\tilde{f}})$. Furthermore, we write $(f_i \geq 0) \in S_f$ to mean that the inequality $f_i \geq 0$ is included in S_f .

Definition III.2. Let $N_{>0} = \{1, 2, \dots\}$. For $a_i \in N_{>0}, i \in \mathcal{N}_n$, a sequence $[a_1, a_2, \dots, a_n]$ is said to be in descending order if $a_1 \geq a_2 \geq \dots \geq a_n$.

Definition III.3. Let $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$ be the sets of positive and nonnegative real numbers, respectively. A linear polynomial F in \mathbf{x} is called a positive (nonnegative) linear combination of polynomials f_j in \mathbf{x} , $j = 1, \dots, k$, if $F = \sum_{j=1}^k r_j f_j$ with $r_j \in \mathbb{R}_{>0}$ ($r_j \in \mathbb{R}_{\geq 0}$). A nonnegative linear combination is also called a conic combination.

Definition III.4. The inequalities $f_1 \geq 0, f_2 \geq 0, \dots, f_k \geq 0$ imply the inequality $f \geq 0$ if the following holds:

$$\mathbf{x} \text{ satisfying } f_1 \geq 0, f_2 \geq 0, \dots, f_k \geq 0 \text{ implies } \mathbf{x} \text{ satisfies } f \geq 0.$$

Definition III.5. Given a set of inequalities $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$, for some $i \in \mathcal{N}_m$, $f_i \geq 0$ is called a redundant inequality if $f_i \geq 0$ is implied by the inequalities $f_j \geq 0$, where $j \in \mathcal{N}_m$ and $j \neq i$.

Definition III.6. Two inequalities $f \geq 0$ and $g \geq 0$ are trivially equivalent if $f = cg$ for some $c \in \mathbb{R}_{>0}$. Given two sets of inequalities $S_f = \{f_i \geq 0, i \in \mathcal{N}_{m_1}\}$ and $S_g = \{g_i \geq 0, i \in \mathcal{N}_{m_2}\}$, we say that S_f and S_g are trivially equivalent if

- 1) S_f and S_g have exactly the same number of inequalities;

- 2) for every $i \in \mathcal{N}_{m_1}$, $f_i \geq 0$ is trivially equivalent to $g_j \geq 0$ for some $j \in \mathcal{N}_{m_2}$;
 3) for every $i \in \mathcal{N}_{m_2}$, $g_i \geq 0$ is trivially equivalent to $f_j \geq 0$ for some $j \in \mathcal{N}_{m_1}$.

Furthermore, if S_f and S_g are trivially equivalent, then we regard S_f and S_g as the same set of inequalities.

Lemma III.1 (Farkas' Lemma[14], [15]). *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then exactly one the following two assertions is true:*

1. *There exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq 0$.*
2. *There exists a $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{A}^T \mathbf{y} \geq 0$ and $\mathbf{b}^T \mathbf{y} < 0$.*

Lemma III.2. *Given $h_1, \dots, h_k, h \in \mathbb{R}_h[\mathbf{y}]$, $h_1 \geq 0, \dots, h_k \geq 0$ imply $h \geq 0$ if and only if h is a conic combination of h_1, \dots, h_k .*

Proof. It is straightforward that $h_1 \geq 0, \dots, h_k \geq 0$ imply $h \geq 0$ if h is a conic combination of h_1, \dots, h_k . We need only to prove the converse.

Assume that $h_1 \geq 0, \dots, h_k \geq 0$. Define a vector $\mathbf{h} = (h_1, \dots, h_k)^T$, and the variable vector $\mathbf{y} = (y_1, \dots, y_m)^T$. Since $h_1, \dots, h_k, h \in \mathbb{R}_h[\mathbf{y}]$, we can let $\mathbf{h} = \mathbf{A}^T \mathbf{y}$ and $h = \mathbf{b}^T \mathbf{y}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Since $h_1 \geq 0, \dots, h_k \geq 0$ imply $h \geq 0$, there exists no $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{A}^T \mathbf{y} \geq 0$ and $\mathbf{b}^T \mathbf{y} < 0$, which means Assertion 2 in Lemma III.1 is false. Then by the lemma, Assertion 1 must be true, that is, there exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq 0$. Then we have

$$\mathbf{Ax} = \mathbf{b} \Rightarrow (\mathbf{Ax})^T = \mathbf{b}^T \Rightarrow \mathbf{x}^T \mathbf{A}^T = \mathbf{b}^T \Rightarrow \mathbf{x}^T \mathbf{A}^T \mathbf{y} = \mathbf{b}^T \mathbf{y} \Rightarrow \mathbf{x}^T \mathbf{h} = h,$$

which implies that h is a conic combination of h_1, \dots, h_k . The lemma is proved. \square

Note that this lemma generalizes Theorem 2 in [1].

Definition III.7. *Let $S_f = \{f_i(\mathbf{x}) \geq 0, i \in \mathcal{N}_m\}$ be an inequality set. If $f_k(\mathbf{x}) = 0$ for all solution \mathbf{x} of S_f , then $f_k(\mathbf{x}) = 0$ is called an implied equality of S_f . The inequality set S_f is called a pure inequality set if S_f has no implied equalities.*

Lemma III.3. *Let $S_f = \{f_i(\mathbf{x}) \geq 0, i \in \mathcal{N}_m\}$ be an inequality set. Then f_k is an implied inequality of S_f if and only if*

$$f_k(\mathbf{x}) \equiv \sum_{i=1, i \neq k}^m p_i f_i(\mathbf{x}), \quad (1)$$

where $p_i \leq 0$ for all $i \in \mathcal{N}_m \setminus \{k\}$.

Proof. Assume (1) holds and let \mathbf{x} be any solution of S_f . Then $f_k(\mathbf{x}) = \sum_{i=1, i \neq k}^m p_i f_i(\mathbf{x}) \leq 0$ since $p_i \leq 0$ and $f_i(\mathbf{x}) \geq 0$, for $i \in \mathcal{N}_m \setminus \{k\}$. On the other hand, from $f_k(\mathbf{x}) \geq 0$, we obtain $f_k(\mathbf{x}) = 0$. Therefore, $f_k(\mathbf{x}) = 0$ for all solution \mathbf{x} of S_f , i.e., f_k is an implied equality of S_f .

Now, assume that f_k is an implied inequality of S_f , i.e., $f_k(\mathbf{x}) = 0$ for all solution \mathbf{x} of S_f . This implies that if \mathbf{x} is a solution of S_f , then $f_k(\mathbf{x}) \leq 0$. In other words, the inequality $f_k(\mathbf{x}) \leq 0$ is implied by the S_f . By Lemma III.2, there exist $q_i \geq 0, i \in \mathcal{N}_m$ such that

$$-f_k(\mathbf{x}) \equiv \sum_{i=1}^m q_i f_i(\mathbf{x}).$$

Then,

$$(-1 - q_k) f_k(\mathbf{x}) \equiv \sum_{i=1, i \neq k}^m q_i f_i(\mathbf{x}),$$

or

$$f_k(\mathbf{x}) \equiv \sum_{i=1, i \neq k}^m \left(-\frac{q_i}{1 + q_k} \right) f_i(\mathbf{x}).$$

Upon letting $p_i = -\frac{q_i}{1 + q_k}$, where $p_i \leq 0$ since $q_i \geq 0$, we obtain (1). This completes the proof. \square

Let $E_{\bar{f}}$ be the set of all implied equalities of S_f . Evidently, $\widetilde{\mathcal{R}}^{-1}(E_{\bar{f}}) \subseteq \mathcal{R}^{-1}(S_f)$. Next, we give an example to show that if an equality set is imposed, a pure inequality set can become a non-pure inequality set.

Example III.1. *Let $S_f = \{f_1 \geq 0, f_2 \geq 0\}$, where $f_1 = x_1 + x_2, f_2 = x_1 - x_2$. Evidently, S_f is a pure inequality set. However, if we impose the constraint $x_1 = 0$, then S_f becomes $\{x_2 \geq 0, -x_2 \geq 0\}$, which is a non-pure inequality set.*

Proposition III.1. *A subset of a pure inequality set is a pure inequality set.*

Proof. The proposition follows immediately from Lemma III.3 and Definition III.7. \square

Definition III.8. *Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$ and $S_{f'} = \{f'_i \geq 0, i \in \mathcal{N}_{m'}\}$ be two inequality sets. If the solution sets of $S_{f'}$ and S_f are the same, then we say that S_f and $S_{f'}$ are equivalent.*

Proposition III.2. *If S_f and $S_{f'}$ are equivalent, then every inequality in S_f is implied by $S_{f'}$, and every inequality in $S_{f'}$ is implied by S_f .*

In the rest of the section, we will develop a few algorithms for simplifying a linear inequality set constrained by a linear equality set.

A. Dimension Reduction of a set of inequalities by an equality set

Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$ be an inequality set and $E_{\tilde{f}} = \{\tilde{f}_i = 0, i \in \mathcal{N}_{\tilde{m}}\}$ be an equality set. Recall that $P_f = \mathcal{R}^{-1}(S_f) = \{f_i, i \in \mathcal{N}_m\}$ and $P_{\tilde{f}} = \tilde{\mathcal{R}}^{-1}(E_{\tilde{f}}) = \{\tilde{f}_i, i \in \mathcal{N}_{\tilde{m}}\}$. The following proposition is well known (see for example [17, Chapter 1]).

Proposition III.3. *Under the variable order $x_1 \prec x_2 \prec \dots \prec x_n$, the linear equation system $E_{\tilde{f}}$ can be reduced by Gauss-Jordan elimination to the unique form*

$$\tilde{E} = \{x_{k_i} - U_i = 0, i \in \mathcal{N}_{\tilde{n}}\}, \quad (2)$$

where $k_1 < k_2 < \dots < k_{\tilde{n}}$, x_{k_i} is the leading term of $x_{k_i} - U_i$, \tilde{n} is rank of the linear system $E_{\tilde{f}}$ and U_i is a linear function in $\{x_j, \text{ for } k_i < j < k_{i+1}, i \in \mathcal{N}_{\tilde{n}}\}$, with $k_{i+1} = n + 1$ by convention. Furthermore, $\sum_{i \in \mathcal{N}_{\tilde{n}}} |U_i| = n - \tilde{n}$.

Algorithm 1 Dimension Reduction

Input: $S_f, E_{\tilde{f}}$.

Output: The remainder set R_f .

- 1: Compute \tilde{E} with $E_{\tilde{f}}$ by Proposition III.3.
 - 2: Substitute x_{k_i} by U_i in P_f to obtain a set R .
 - 3: Let $R_f = R \setminus \{0\}$.
 - 4: **return** $\mathcal{R}(R_f)$.
-

We call the equality set \tilde{E} the *Jordan normal form* of $E_{\tilde{f}}$. Likewise, we call the polynomial set $\tilde{\mathcal{R}}^{-1}(\tilde{E})$ the Jordan normal form of $\tilde{\mathcal{R}}^{-1}(E_{\tilde{f}})$. We say reducing S_f by $E_{\tilde{f}}$ to mean using Algorithm 1 to find $\mathcal{R}(R_f)$. We also say reducing P_f by $E_{\tilde{f}}$ to mean using Algorithm 1 to find R_f , called *the remainder set* (or remainder if R_f is a singleton).

Example III.2. *Given a variable order $x_1 \prec x_2 \prec x_3$, let $S_f = \{f_1 \geq 0, f_2 \geq 0\}$ and $E_{\tilde{f}} = \{\tilde{f}_1 = 0, \tilde{f}_2 = 0, \tilde{f}_3 = 0\}$, where $f_1 = x_1 + x_2 - x_3$, $f_2 = x_2 + x_3$, $\tilde{f}_1 = x_1 + x_2 + x_3$, $\tilde{f}_2 = x_1 + x_2$, and $\tilde{f}_3 = x_3$. We write $P_f = \mathcal{R}^{-1}(S_f) = \{f_1, f_2\}$ and $P_{\tilde{f}} = \tilde{\mathcal{R}}^{-1}(E_{\tilde{f}}) = \{\tilde{f}_1, \tilde{f}_2, \tilde{f}_3\}$.*

Firstly, we obtain that the rank of $E_{\tilde{f}}$ is $\tilde{n} = 2$. Then the Jordan normal form of $E_{\tilde{f}}$ is given by $\tilde{E} = \{x_{k_1} - U_1 = 0, x_{k_2} - U_2 = 0\}$, where $k_1 = 1$, $k_2 = 3$, $U_1 = -x_2$, $U_2 = 0$.

Using the equality constraints in \tilde{E} , we substitute $x_1 = -x_2$ and $x_3 = 0$ into $P_f = \{f_1, f_2\}$ to obtain $R = \{0, x_2\}$. Hence $R_f = R \setminus \{0\} = \{x_2\}$. In other words, the inequality set S_f is reduced to $\mathcal{R}(R_f) = \{x_2 \geq 0\}$ by the equality set $E_{\tilde{f}}$. Note that in $\mathcal{R}(R_f)$, only $n - \tilde{n} = 1$ variable, namely x_2 , appears.

Remark III.1. *After the execution of Algorithm 1, the inequality set S_f constrained by the equality set $E_{\tilde{f}}$ is reduced to the inequality set $\mathcal{R}(R_f)$ constrained by the equality set \tilde{E} . Therefore, the solution set of ‘ S_f constrained by $E_{\tilde{f}}$ ’ in \mathbb{R}^n is the same as the solution set of ‘ $\mathcal{R}(R_f)$ constrained by \tilde{E} ’ in \mathbb{R}^n .*

B. The implied equalities contained in a system of inequalities

In this subsection, we will show how to find all the implied equalities contained in a system of linear inequalities.

Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$ be a given inequality set, where f_i is a linear function in \mathbf{x} . The following algorithm, called the Implied Equalities Algorithm, finds all the implied equalities of S_f .

Algorithm 2 Implied Equalities Algorithm

Input: S_f .

Output: The implied equalities in S_f .

- 1: Let $E_0 := \sum_{i=1}^m v_i f_i$, where $V = \{v_i, i \in \mathcal{N}_m\}$ is a set of variables.
- 2: Set $E_0 \equiv \sum_{j=1}^n w_j x_j \equiv 0$. Then $W = \{w_j = 0, j \in \mathcal{N}_n\}$ is a linear system in V .
- 3: Solve the linear equations $\{w_j = 0, j \in \mathcal{N}_n\}$ by Gauss-Jordan elimination to obtain the solution set of v_i of the form $\{v_i = V_i, i \in \mathcal{N}_m\}$, where d is the rank of the linear system W and V_i is a linear function in $m - d$ variables of V .
- 4: For every $k \in \mathcal{N}_m$, let $L_k, k = 1, \dots, m$ be the following linear programming problem:

$$\begin{aligned} & \max(V_k) \\ & \text{s.t. } V_i \geq 0, \quad i = 1, 2, \dots, m. \end{aligned} \quad (3)$$

- 5: The equality $f_k = 0$ is an implied equality of S_f if and only if the optimal value of $L_k \max(V_k) > 0$.
 - 6: **return** All implied equalities f_k 's in S_f .
-

With Algorithm 2, we can obtain the set of implied equalities of S_f , denoted by $E_{\tilde{f}}$. The following example illustrates how we can apply Algorithm 2 and then Algorithm 1 to reduce a given inequality set. A justification of Algorithm 2 is given after the example.

Example III.3. Fix the variable order $x_1 \prec x_2 \prec x_3$. Let $S_f = \{f_1 \geq 0, f_2 \geq 0, f_3 \geq 0, f_4 \geq 0, f_5 \geq 0\}$, where $f_1 = x_1$, $f_2 = x_2 - x_1$, $f_3 = -x_1$, $f_4 = -x_2$ and $f_5 = x_2 + x_3$. An application of Algorithm 2 to S_f yields the following:

- Firstly, we let $E_0 = \sum_{i=1}^5 v_i f_i = \sum_{j=1}^3 w_j x_j$. Then we have $V = \{v_1, v_2, v_3, v_4, v_5\}$ and $W = \{w_1 = 0, w_2 = 0, w_3 = 0\}$ with $w_1 = v_1 - v_2 - v_3$, $w_2 = v_2 - v_4 + v_5$ and $w_3 = v_5$.
- The rank of W is $d = 3$. We then solve the linear equations W by Gauss-Jordan elimination to obtain $\{v_i = V_i, i \in \mathcal{N}_5\}$, where $V_1 = v_3 + v_4$, $V_2 = v_4$, $V_3 = v_3$, $V_4 = v_4$ and $V_5 = 0$, from which we can see that V_i is a linear function of the two variables v_3 and v_4 .
- Finally, we have the following 5 linear programming problems:
 - L_1 : $\max(v_3 + v_4)$ s.t. $v_3 + v_4 \geq 0, v_3 \geq 0, v_4 \geq 0$.
 - L_2 : $\max(v_4)$ s.t. $v_3 + v_4 \geq 0, v_3 \geq 0, v_4 \geq 0$.
 - L_3 : $\max(v_3)$ s.t. $v_3 + v_4 \geq 0, v_3 \geq 0, v_4 \geq 0$.
 - L_4 : $\max(v_4)$ s.t. $v_3 + v_4 \geq 0, v_3 \geq 0, v_4 \geq 0$.
 - L_5 : $\max(0)$ s.t. $v_3 + v_4 \geq 0, v_3 \geq 0, v_4 \geq 0$.
- Observe that L_2 and L_4 are same, and the optimal value of L_5 is 0. Then, we solve L_1 to L_3 to obtain that the optimal values are all equal to $+\infty$. Thus, we obtain the implied equality set, denoted by $E_{\tilde{f}} = \{\tilde{f}_1 = 0, \tilde{f}_2 = 0, \tilde{f}_3 = 0, \tilde{f}_4 = 0\}$, where $\tilde{f}_1 = x_1$, $\tilde{f}_2 = x_2 - x_1$, $\tilde{f}_3 = -x_1$ and $\tilde{f}_4 = -x_2$.

Upon applying Algorithm 2, the inequality set S_f is reduced to the inequality set $S'_f = \{f_5 \geq 0\} = \{x_2 + x_3 \geq 0\}$ constrained by the equality set $E_{\tilde{f}}$. Finally, apply Algorithm 1 with S'_f and $E_{\tilde{f}}$ as inputs to obtain $R_f = \{x_3 \geq 0\}$. In other words, the inequality set S_f is reduced to $\{x_3 \geq 0\}$ constrained by the equality set $E_{\tilde{f}}$ after the applications of Algorithm 2 and then Algorithm 1.

Justification for Algorithm 2. In Algorithm 2, the optimal value of L_k being positive means that we can find a set of values of $v_i, i \in \mathcal{N}_m$ satisfying $v_k > 0$ and $v_j \geq 0$ for $j \neq k$, such that $\sum_{i=1}^m v_i f_i \equiv 0$, which can be rewritten as

$$f_k \equiv \sum_{i=1, i \neq k}^m \left(-\frac{v_i}{v_k} \right) f_i.$$

Since by Lemma III.3, $f_k = 0$ is an implied equality if and only if $f_k \equiv \sum_{i=1, i \neq k}^m p_i f_i$ with $p_i \leq 0$ for $i \in \mathcal{N}_m$, we see that the equality $f_k = 0$ is an implied equality of S_f if and only if the optimal value of L_k is positive.

C. Minimal characterization set

In this subsection, we first define a minimal characterization set of an inequality set and prove its uniqueness. Then we present an algorithm to obtain this set.

Definition III.9. Let $S_g = \{g_i \geq 0, i \in \mathcal{N}_m\}$ be an inequality set and $S_{g'} = \{g'_i \geq 0, i \in \mathcal{N}_{m'}\}$ be a subset of S_g . If

- 1) S_g and $S_{g'}$ are equivalent, and
- 2) there is no redundant inequalities in $S_{g'}$,

we say that $S_{g'}$ is a minimal characterization set of S_g .

Definition III.10. Let $S_g = \{g_i \geq 0, i \in \mathcal{N}_m\}$ and $S_{g'} = \{g'_i \geq 0, i \in \mathcal{N}_{m'}\}$ be two inequality sets. We say $P_{g'} = \mathcal{R}^{-1}(S_{g'})$ is a minimal characterization set of $P_g = \mathcal{R}^{-1}(S_g)$ if $S_{g'}$ is a minimal characterization set of S_g .

Proposition III.4. Let $S_g = \{g_i \geq 0, i \in \mathcal{N}_m\}$ be an inequality set. If $S_{g'} = \{g'_i \geq 0, i \in \mathcal{N}_{m'}\}$ is a minimal characterization set of S_g , then $m' \leq m$ and $0 \notin \mathcal{R}^{-1}(S_{g'})$.

Proof. Since $S_{g'} \subseteq S_g$ by Definition III.9, we have $m' \leq m$. In addition, if $0 \in \mathcal{R}^{-1}(S_{g'})$, then $0 \geq 0$ is a redundant inequality in $S_{g'}$, which contradicts that $S_{g'}$ is a minimal characterization set of S_g . Thus, $0 \notin \mathcal{R}^{-1}(S_{g'})$. \square

The following corollary is immediate from Definition III.9 and Proposition III.1.

Corollary III.1. A minimal characterization set of a pure inequality set is also a pure inequality set.

Theorem III.1. Let $h_1, \dots, h_m \in \mathbb{R}_h[\mathbf{x}]$ and $S_h = \{h_i \geq 0, i \in \mathcal{N}_m\}$ be a pure inequality set. Then the minimal characterization set of S_h is unique.

Proof. Consider two minimal characterization sets of a pure set of linear inequalities S_h , denoted by $S_{h'} = \{h'_i \geq 0, i \in \mathcal{N}_{m_1}\}$ and $S_{\bar{h}} = \{\bar{h}_i \geq 0, i \in \mathcal{N}_{m_2}\}$. By Definition III.9, $S_{h'}$ and $S_{\bar{h}}$ are equivalent, and by Corollary III.1, they are both pure inequality sets. We will prove by contradiction that $S_{h'}$ and $S_{\bar{h}}$ are trivially equivalent.

Assume that for some inequality $(h'_j \geq 0) \in S_{h'}$, we cannot find $(\bar{h}_i \geq 0) \in S_{\bar{h}}$ that is trivially equivalent to $h'_j \geq 0$. By Proposition III.2 and Lemma III.2, we have

$$h'_j \equiv \sum_{i=1}^{m_2} p_i \bar{h}_i,$$

with $p_i \geq 0$. Without loss of generality, assume that $p_i > 0$ for $i = 1, \dots, \bar{m}_2$ and $p_i = 0$ for $i = \bar{m}_2 + 1, \dots, m_2$, where $2 \leq \bar{m}_2 \leq m_2$. Again by Lemma III.2, for all $i \in \mathcal{N}_{m_2}$,

$$\bar{h}_i \equiv \sum_{k=1}^{m_1} q_{i,k} h'_k, \quad (4)$$

where $q_{i,k} \geq 0$. Then

$$h'_j \equiv \sum_{i=1}^{\bar{m}_2} p_i \bar{h}_i \equiv \sum_{i=1}^{\bar{m}_2} p_i \sum_{k=1}^{m_1} q_{i,k} h'_k. \quad (5)$$

Rewrite (5) as

$$\left(1 - \sum_{i=1}^{\bar{m}_2} p_i q_{i,j}\right) h'_j(\mathbf{x}) \equiv \sum_{i=1}^{\bar{m}_2} p_i \sum_{k \in \mathcal{N}_{m_1} \setminus \{j\}} q_{i,k} h'_k(\mathbf{x}). \quad (6)$$

By collecting the coefficients of $h'_k(\mathbf{x})$ on the RHS, we have

$$\left(1 - \sum_{i=1}^{\bar{m}_2} p_i q_{i,j}\right) h'_j(\mathbf{x}) \equiv \sum_{k \in \mathcal{N}_{m_1} \setminus \{j\}} a_k h'_k(\mathbf{x}). \quad (7)$$

where

$$a_k = \sum_{i=1}^{\bar{m}_2} p_i q_{i,k}. \quad (8)$$

Now in (4), for a fixed $i \in \mathcal{N}_{m_2}$, if $q_{i,k} = 0$ holds for all $k = 1, \dots, m_1$ such that $k \neq j$, then we have

$$\bar{h}_i \equiv \sum_{k=1}^{m_1} q_{i,k} h'_k \equiv q_{i,j} h'_j. \quad (9)$$

If $q_{i,j} > 0$, then \bar{h}_i and h'_j are trivially equivalent, contradicting our assumption that there exists no $\bar{h}_i \in \mathcal{S}_{\bar{h}}$ which is trivially equivalent to h'_j . On the other hand, if $q_{i,j} = 0$, then $\bar{h}_i \equiv 0$, which by Proposition III.4 contradicts the assumption that $\mathcal{S}_{\bar{h}}$ is a minimal characterization set of S_h . Thus we conclude that for every $i \in \mathcal{N}_{m_1}$, there exists at least one $k \in \mathcal{N}_{m_1} \setminus \{j\}$ such that $q_{i,k} > 0$. From this and (8), it is not difficult to see that on the RHS of (7), there exists at least one $k \in \mathcal{N}_{m_1} \setminus \{j\}$ such that $a_k > 0$.

Consider a solution \mathbf{x}^* of $S_{h'}$ such that $h'_k(\mathbf{x}^*) > 0$ for all $k \in \mathcal{N}_{m_1}$. Such an \mathbf{x}^* exists because $S_{h'}$ is a pure inequality set. Substituting $\mathbf{x} = \mathbf{x}^*$ in (7) to yield

$$\left(1 - \sum_{i=1}^{\bar{m}_2} p_i q_{i,j}\right) h'_j(\mathbf{x}^*) = \sum_{k \in \mathcal{N}_{m_1} \setminus \{j\}} a_k h'_k(\mathbf{x}^*). \quad (10)$$

Since there exists at least one $k \in \mathcal{N}_{m_1} \setminus \{j\}$ such that $a_k > 0$, the RHS above is strictly positive, which implies that $1 - \sum_{i=1}^{\bar{m}_2} p_i q_{i,j} > 0$. It then follows that h'_j can be written as a conic combination of h'_k , $k \in \mathcal{N}_{m_1} \setminus \{j\}$. In other words, $h'_j \geq 0$ is implied by $h'_k \geq 0$, $k \in \mathcal{N}_{m_1} \setminus \{j\}$. This contradicts that $S_{h'}$ is a minimal characterization set of S_h .

Summarizing the above, we have proved that for every $(h'_j \geq 0) \in S_{h'}$, we can find an $(\bar{h}_i \geq 0) \in \mathcal{S}_{\bar{h}}$ which is trivially equivalent to $h'_j \geq 0$. Moreover, \bar{h}_i is unique, which can be seen as follows. If there exists another $(\bar{h}_{i'} \geq 0) \in \mathcal{S}_{\bar{h}}$ which is trivially equivalent to $h'_j \geq 0$, then $\bar{h}_i \geq 0$ and $\bar{h}_{i'} \geq 0$ are also trivially equivalent to each other, contradicting that $S_{h'}$ is a minimal characterization set of S_h . In the same way, we can prove that for every $(\bar{h}_i \geq 0) \in \mathcal{S}_{\bar{h}}$, we can find a unique $(h'_j \geq 0) \in S_{h'}$ which is trivially equivalent to $\bar{h}_i \geq 0$. Thus, $S_{h'}$ and $\mathcal{S}_{\bar{h}}$ are trivially equivalent and have exactly the same number of inequalities, which means that the minimal characterization set of a pure inequality set S_h is unique. This completes the proof of the theorem. \square

Theorem III.2. Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_{m_1}\}$ and $S_g = \{g_i, i \in \mathcal{N}_{m_2}\}$ be two pure inequality sets, and $S_{f'}$ and $S_{g'}$ be their minimal characterization sets respectively. If S_f and S_g are equivalent, then $S_{f'}$ and $S_{g'}$ are trivially equivalent.

Proof. If the two pure inequality sets S_f and S_g are equivalent, then $S_{f'}$ and $S_{g'}$ are pure and equivalent. Thus the theorem follows immediately from the proof of Theorem III.1. \square

Next, we give an example to show that the minimal characterization set of a non-pure inequality set may not be unique.

Example III.4. Let $S_f = \{f_1 \geq 0, f_2 \geq 0, f_3 \geq 0, f_4 \geq 0, \}$ be an inequality set, where $f_1 = x_1 - x_2$, $f_2 = x_2$, $f_3 = -x_2$, $f_4 = x_1$. Evidently, S_f is a non-pure inequality set, and it can readily be seen that both $S_{f'} = \{f_1 \geq 0, f_2 \geq 0, f_3 \geq 0\}$ and $S_{f''} = \{f_2 \geq 0, f_3 \geq 0, f_4 \geq 0\}$ are minimal characterization sets of S_f . However, $S_{f'}$ and $S_{f''}$ are not trivially equivalent. Thus, the minimal characterization set of S_f isn't unique.

Let $S_h = \{h_i \geq 0, i \in \mathcal{N}_m\}$ be an inequality set, where $h_i \in \mathbb{R}_h[\mathbf{x}]$. Based on Lemma III.2, the following algorithm, called Minimal Characterization Set Algorithm, can be used to obtain a minimal characterization set of S_h .

Algorithm 3 Minimal Characterization Set Algorithm

Input: S_h .

Output: A minimal characterization set of S_h .

 Set $P_h := \mathcal{R}^{-1}(S_h)$, $\mathcal{M} := \mathcal{N}_m$.

- 1: **for** k from 1 to m **do**
 - 2: Let $H_k := h_k - \sum_{i \in \mathcal{M} \setminus \{k\}} q_{i,k} h_i$, where $T_k = \{q_{i,k}, i \in \mathcal{M} \setminus \{k\}\}$ is a set of variables.
 - 3: Set $H_k \equiv \sum_{i=1}^n Q_{i,k} x_i \equiv 0$. Then $\tilde{T}_k = \{Q_{i,k} = 0, i \in \mathcal{N}_n\}$ is a linear system in T_k .
 - 4: Solve the linear equations of \tilde{T}_k .
 - 5: **if** the linear equations of \tilde{T}_k can be solved **then**
 - 6: Obtain the solution set of $q_{i,k}$ of the form $\{q_{i,k} = \mathcal{Q}_{i,k}, i \in \mathcal{M} \setminus \{k\}\}$, where d_1 is the rank of the linear system \tilde{T}_k and $\mathcal{Q}_{i,k}$ is a linear function in $N[\mathcal{M} \setminus \{k\}] - d_1$ variables of T_k .
 - 7: Let L_k be the following linear programming problem:

$$\begin{aligned} & \min(0) \\ \text{s.t. } & \mathcal{Q}_{i,k} \geq 0, i \in \mathcal{M} \setminus \{k\}. \end{aligned}$$
 - 8: **if** L_k can be solved **then**
 - 9: $P_h := P_h \setminus \{h_k\}$, $\mathcal{M} := \mathcal{M} \setminus \{k\}$.
 - 10: **end if**
 - 11: **end if**
 - 12: **end for**
 - 13: **return** $\mathcal{R}(P_h)$.
-

Justification for Algorithm 3. Steps 2 to 11 remove the polynomial h_k from P_h if it can be expressed as a conic combination of $h_i, i \in \mathcal{M} \setminus \{k\}$. Iterating over all k from 1 to m , the output inequality set $\mathcal{R}(P_h)$ is equivalent to S_h and it is a pure inequality set. Hence, it is a minimal characterization set of S_h .

D. The reduced minimal characterization set

In this subsection, we first define the reduced minimal characterization set of a linear inequality set and prove its uniqueness. Then we present an algorithm to obtain this set.

Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$ be a linear inequality set, and $E_{\bar{f}}$ be the set of implied equalities of S_f obtained by applying Algorithm 2. Then we obtain \tilde{E} , the Jordan normal form of $E_{\bar{f}}$, as in Proposition III.3. Let R_f be the remainder set obtained by reducing $\mathcal{R}^{-1}(S_f) \setminus \mathcal{R}^{-1}(E_{\bar{f}})$ by $\tilde{\mathcal{R}}^{-1}(\tilde{E})$ using Algorithm 1.

Theorem III.3. *The set $\mathcal{R}(R_f)$ is a pure inequality set.*

Proof. Let $\tilde{E} = \{E_i = 0, i \in \mathcal{N}_{\tilde{n}}\}$, and assume there is an implied equality $(\bar{f} = 0) \in \mathcal{R}(R_f)$. In the process of obtaining \bar{f} , we substitute $x_{k_i} = U_i, i \in \mathcal{N}_{\tilde{n}}$ into some polynomial $f \in \mathcal{R}^{-1}(S_f)$ (cf. 2). Therefore, we can write

$$\bar{f} \equiv f - \sum_{i=1}^{\tilde{n}} c_i E_i, \quad (11)$$

where c_i is the coefficient of x_{k_i} in f . Let \mathbf{x}^* be a solution of S_f . From Remark III.1, we see that \mathbf{x}^* is also a solution of $\mathcal{R}(R_f)$ constrained by \tilde{E} , so that $E_i(\mathbf{x}^*) = 0$ for all $i \in \mathcal{N}_{\tilde{n}}$. From (11), we have

$$f(\mathbf{x}^*) = \bar{f}(\mathbf{x}^*) - \sum_{i=1}^{\tilde{n}} c_i E_i(\mathbf{x}^*).$$

Since $\bar{f} = 0$ is an implied inequality of S_f , we have $\bar{f}(\mathbf{x}^*) = 0$. It follows from the above that $f(\mathbf{x}^*) = 0$. Since this holds for all solution \mathbf{x}^* of S_f , we see that $f = 0$ is an implied equality of S_f , i.e., $(f = 0) \in E_{\bar{f}}$, which is a contradiction to $f \in \mathcal{R}^{-1}(S_f) \setminus \mathcal{R}^{-1}(E_{\bar{f}})$. The theorem is proved. \square

Since $\mathcal{R}(R_f)$ is a pure inequality set, the minimal characterization set of $\mathcal{R}(R_f)$ is unique. We let $S_{r'}$ be the minimal characterization set of $\mathcal{R}(R_f)$.

Definition III.11. The set $S_M = \tilde{E} \cup S_{r'}$ is called the reduced minimal characterization set of S_f .

Theorem III.4. The reduced minimal characterization set of S_f is unique.

Proof. Fix the variable order $x_1 \prec x_2 \prec \dots \prec x_n$. By Proposition III.3, the reduced standard basis $\tilde{\mathcal{R}}^{-1}(\tilde{E})$ is unique, which yields that the remainder set R_f is unique. Since $\mathcal{R}(R_f)$ is a pure inequality set by Theorem III.1, the minimal characterization set of $\mathcal{R}(R_f)$ is unique. Hence, S_M is unique. \square

In the following, we present an algorithm to find the reduced minimal characterization set of a linear inequality set.

Algorithm 4 Reduced Minimal Characterization Set Algorithm

Input: S_f .

Output: The reduced minimal characterization set of S_f .

- 1: Apply Algorithm 2 to find the implied equality set of S_f , denoted by $E_{\tilde{f}}$.
 - 2: Apply Algorithm 1 to reduce $\mathcal{R}^{-1}(S_f) \setminus \tilde{\mathcal{R}}^{-1}(E_{\tilde{f}})$ by $E_{\tilde{f}}$ to obtain R_f .
 - 3: Apply Algorithm 3 to obtain the minimal characterization set of $\mathcal{R}(R_f)$, denoted by $S_{r'}$.
 - 4: **return** $S_M = \tilde{E} \cup S_{r'}$.
-

By Proposition III.3 and Theorems III.2 and III.4, we immediately obtain the following theorem.

Theorem III.5. For two equivalent inequality sets, their reduced minimal characterization sets are same.

Note that for a pure inequality set, the minimal characterization set is exactly the reduced minimal characterization set.

Remark III.2. Since the basic inequalities contain no implied inequality and hence form a pure inequality set, the elemental inequalities form the minimal characterization set of the basic inequalities. In fact, for a fixed number of random variables, Algorithm 4 can be used to compute the reduced minimal characterization set of the basic inequalities under the constraint of an equality set and possibly an inequality set (used for example, for including some non-Shannon-type inequalities).

IV. THE s -VARIABLES

The I -Measure [13] gives a set-theoretic interpretation of Shannon's information measure. In this section, we first give a brief introduction to the I -Measure. The readers are referred to [9, Chapter 3] for a detailed discussion. Then we introduce the s -variables which facilitate the implementation of the algorithms to be developed in Section III.

Consider random variables X_i , $i = 1, \dots, n$ which are jointly distributed, and let \tilde{X}_i be a set variable corresponding to the random variable X_i . Define the universal set Ω to be $\cup_{i=1}^n \tilde{X}_i$ and let \mathcal{F}_n be the σ -field generated by $W = \{\tilde{X}_i, i = 1, \dots, n\}$. The atoms of \mathcal{F}_n have the form $\cap_{i=1}^n Y_i$, where Y_i is either \tilde{X}_i or \tilde{X}_i^c . Let $\mathcal{A}_n \subset \mathcal{F}_n$ be the set of all atoms of \mathcal{F}_n except for $\cap_{i=1}^n \tilde{X}_i^c$, which is \emptyset , the empty set. Note that $|\mathcal{A}_n| = 2^n - 1$. To simplify notations, we shall use X_G to denote $(X_i, i \in G)$, and \tilde{X}_G to denote $\cup_{i \in G} \tilde{X}_i$.

The I -measure μ^* , which is a signed measure on \mathcal{F}_n , is constructed by defining $\mu^*(\tilde{X}_G) = H(X_G)$ for all nonempty subsets G of \mathcal{N}_n . It is consistent with all Shannon's information measure in the sense that the following holds for all (not necessarily disjoint) subsets G, G', G'' of \mathcal{N}_n where G and G' are nonempty:

$$\mu^*(\tilde{X}_G \cap \tilde{X}_{G'} - \tilde{X}_{G''}) = I(X_G; X_{G'} | X_{G''}).$$

To facilitate the discussion in this paper, we introduce the concept of s -variables, which have the form s_{i_1, i_2, \dots, i_n} , $i_1, i_2, \dots, i_n \in \mathcal{N}_n$. For an integer set $S \subset \mathcal{N}_n$, we denote its minimum by $\min(S)$.

Definition IV.1. Let μ^* be an unspecified I -measure of \mathcal{F}_n . Let $A = \cap_{i=1}^n Y_i$ be an atom in \mathcal{A}_n and S be the subset of \mathcal{N}_n such that $Y_i = \tilde{X}_i$ for $i \in S$ and $Y_i = \tilde{X}_i^c$ for $i \in S^c = \mathcal{N}_n \setminus S$. Replace the integers $i \in S^c$ in the sequence $[n]$ by $*$ to yield the sequence B_* . Then replace all the $*$'s in B_* by $\min(S)$ to yield another sequence B_s . Let $s_{i_1, i_2, \dots, i_n} = \mu^*(A)$, where $[i_1, i_2, \dots, i_n] = B_s$. The variable s_{i_1, i_2, \dots, i_n} is called the s -variable associated with the atom A .

Note that in the above definition, there is a one-to-one correspondence between the s -variable s_{i_1, i_2, \dots, i_n} and the atom A . On the one hand, the s -variable can be obtained from an atom A as described above. On the other hand, we can determine the associated atom A from the s -variable s_{i_1, i_2, \dots, i_n} through S , with $A = \cap_{i=1}^n Y_i$, where $Y_i = \tilde{X}_i$ for $i \in S$ and $Y_i = \tilde{X}_i^c$ for $i \in S^c$. This is illustrated in the example below.

Example IV.1. Given the atom $A = \tilde{X}_1 \cap \tilde{X}_2^c \cap \tilde{X}_3 \cap \tilde{X}_4^c$, we have $S = \{1, 3\}$ and $S^c = \{2, 4\}$, and $B_* = [1, *, 3, *]$. Replace all the $*$'s in B_* by the smallest element in S to yield $B_s = [1, 1, 3, 1]$. Then $s_{1,1,3,1} = \mu^*(A)$ is the s -variable

corresponding to the atom A . On the other hand, from $s_{1,1,3,1}$, we can obtain $S = \{1, 1, 3, 1\} = \{1, 3\}$, from which A can be determined.

We now introduce some further notations. Let $t = s_{i_1, i_2, \dots, i_n}$ be an s -variable. The set $L(t) = \{i_1, i_2, \dots, i_n\}$ is called the *subscript set* of t . The sequence $\mathcal{L}(t) = [i_1, i_2, \dots, i_n]$ is called the *subscript sequence* of t . The number of elements in the subscript set is denoted by $N[L(t)]$, and the length of the subscript sequence, denoted by $N[\mathcal{L}(t)]$, is equal to n .

For splitting an s -variable s_{i_1, i_2, \dots, i_n} , we mean adding an element to $L(s_{i_1, i_2, \dots, i_n})$ and yielding two new s -variables $s_{i_1, i_2, \dots, i_n, i_1}$ and $s_{i_1, i_2, \dots, i_n, n+1}$. Note that if s_{i_1, i_2, \dots, i_n} corresponds to an atom $A \in \mathcal{A}_n$, then $s_{i_1, i_2, \dots, i_n, i_1}$ and $s_{i_1, i_2, \dots, i_n, n+1}$ correspond to the atoms $A \cap \tilde{X}_{n+1}^c$ and $A \cap \tilde{X}_{n+1}$ in \mathcal{A}_{n+1} , respectively.

Definition IV.2. For $S \subset N_{>0}$ and $a, b \in N_{>0}$, we introduce the following shorthand notations:

- ' $a \vee b \in S$ ' means $a \in S$ or $b \in S$,
- ' $a \wedge b \in S$ ' means $a \in S$ and $b \in S$,
- ' $a \setminus b \in S$ ' means $a \in S$ and $b \notin S$,
- ' $a \vee b \notin S$ ' means $a \notin S$ and $b \notin S$.¹

Based on Definition IV.2, for $c, d \in N_{>0}$, we further have the following:

- ' $a \wedge (b \vee c) \in S$ ' means $a \in S$ and $b \vee c \in S$,
- ' $(a \vee b) \setminus (c \vee d) \in S$ ' means $a \vee b \in S$ and $c \vee d \notin S$,
- ' $(a \wedge (b \vee c)) \setminus d \in S$ ' means $a \wedge (b \vee c) \in S$ and $d \notin S$.

For $n \in N_{>0}$, let S_n be the set of s -variables of all the atoms in \mathcal{A}_n . Note that S_{n+1} can be obtained from S_n . We first illustrate the case $n = 1$. First of all, $S_1 = \{s_1\}$, where $s_1 = \mu^*(\tilde{X}_1)$. Then, we split s_1 to obtain $s_{1,1}$ and $s_{1,2}$ in S_2 , where $s_{1,1} = \mu^*(\tilde{X}_1 - \tilde{X}_2)$ and $s_{1,2} = \mu^*(\tilde{X}_1 \cap \tilde{X}_2)$. By also including the additional variable $s_{2,2} = \mu^*(\tilde{X}_2 - \tilde{X}_1)$, we obtain $S_2 = \{s_{1,1}, s_{1,2}, s_{2,2}\}$.

In general, we can obtain S_{n+1} from S_n as follows. For every s -variable s_{i_1, i_2, \dots, i_n} in S_n , we split s_{i_1, i_2, \dots, i_n} to obtain $s_{i_1, i_2, \dots, i_n, i_1}$ and $s_{i_1, i_2, \dots, i_n, n+1}$ in S_{n+1} . Then we obtain S_{n+1} by including the additional variable $s_{n, n, \dots, n}$ with $N[\mathcal{L}(s_{n, n, \dots, n})] = n + 1$.

As illustrations of the use of the notations we have introduced, we state the following which can readily be verified:

- 1) $H(X_a, X_b) = \sum t$ for t such that $a \vee b \in L(t)$,
- 2) $I(X_a; X_b) = \sum t$ for t such that $a \wedge b \in L(t)$,
- 3) $H(X_a | X_b) = \sum t$ for t such that $a \setminus b \in L(t)$,
- 4) $I(X_a; (X_b, X_c)) = \sum t$ for t such that $a \wedge (b \vee c) \in L(t)$,
- 5) $H(X_a, X_b | X_c, X_d) = \sum t$ for t such that $(a \vee b) \setminus (c \vee d) \in L(t)$,
- 6) $I((X_a; (X_b, X_c)) | X_d) = \sum t$ for t such that $(a \wedge (b \vee c)) \setminus d \in L(t)$.

For example, for three random variables X_1, X_2, X_3 , we have the following:

$$H(X_1, X_2) = \mu^*(\tilde{X}_1 \cup \tilde{X}_2) = \sum_{1 \vee 2 \in L(s_{i,j,k})} s_{i,j,k} = s_{1,1,1} + s_{1,1,3} + s_{1,2,1} + s_{1,2,3} + s_{2,2,2} + s_{2,2,3},$$

$$I(X_1; X_2) = \mu^*(\tilde{X}_1 \cap \tilde{X}_2) = \sum_{1 \wedge 2 \in L(s_{i,j,k})} s_{i,j,k} = s_{1,2,1} + s_{1,2,3},$$

$$H(X_1, X_2 | X_3) = \mu^*(\tilde{X}_1 \cup \tilde{X}_2 - \tilde{X}_3) = \sum_{(1 \vee 2) \setminus 3 \in L(s_{i,j,k})} s_{i,j,k} = s_{1,1,1} + s_{1,2,1} + s_{2,2,2},$$

$$I(X_1; X_2 | X_3) = \mu^*(\tilde{X}_1 \cap \tilde{X}_2 - \tilde{X}_3) = \sum_{(1 \wedge 2) \setminus 3 \in L(s_{i,j,k})} s_{i,j,k} = s_{1,2,1}.$$

Using this set of notations, we can express a Shannon's information measure as a linear polynomial in the s -variables which are indexed by subscript sequences, so that they can be conveniently represented in a computer implementation.

Definition IV.3 (s -variable order). Let $t_1 = s_{i_1, i_2, \dots, i_n}$ and $t_2 = s_{j_1, j_2, \dots, j_n}$ be two s -variables. We write $t_1 \succ t_2$ if one of the following conditions is satisfied:

- 1) $N[L(t_1)] > N[L(t_2)]$,
- 2) $N[L(t_1)] = N[L(t_2)]$, $i_l = j_l$ for $l = 1, \dots, k-1$ and $i_k < j_k$.

Definition IV.4. For $n \in N_{>0}$, let S_n be the set of s -variables. The associated s -variable sequence S_n is obtained by ordering the elements in S_n according to the s -variable order.

¹Equivalently, $a \vee b \notin S$ means $\sim (a \in S \text{ or } b \in S)$.

For example, the s -variable sequence \mathcal{S}_3 is $[s_{1,2,3}, s_{1,1,3}, s_{1,2,1}, s_{2,2,3}, s_{1,1,1}, s_{2,2,2}, s_{3,3,3}]$. The s -variable order is employed in the computational procedures to be discussed in the next section for the convenience of implementation.

V. PROCEDURES FOR PROVING INFORMATION INEQUALITIES AND IDENTITIES

In this section, we present two procedures for proving information inequalities and identities under the constraint of an inequality set and/or an equality set. They are designed in the spirit of Theorem II.2.

A. Procedure I: Proving Information Inequalities

Input:

Objective information inequality: $\bar{F} \geq 0$.

Additional constraints: $\bar{C}_i = 0, i = 1, \dots, r_1; \bar{C}_j \geq 0, j = r_1 + 1, \dots, r_2$.

Element information inequalities: $\bar{C}_k \geq 0, k = r_2 + 1, \dots, r_3$.

// Here, \bar{F} , \bar{C}_i , \bar{C}_j , and \bar{C}_k are linear combination of information measures.

Output: A proof of $\bar{F} \geq 0$ if feasible.

Step 1. Construct the s -variable set \mathcal{S}_n and the associated s -variable sequence \mathcal{S}_n .

Step 2. Transform \bar{F} , \bar{C}_i , \bar{C}_j and \bar{C}_k to linear polynomials F , C_i , C_j and C_k in \mathcal{S}_n respectively.

// We need to solve

// **Problem P₁**: Determine whether $F \geq 0$ is implied by

$$\begin{aligned} C_i &= 0, i = 1, \dots, r_1, \\ C_j &\geq 0, j = r_1 + 1, \dots, r_2, \\ C_k &\geq 0, k = r_2 + 1, \dots, r_3. \end{aligned}$$

Step 3. Apply Algorithm 1 to reduce $\{C_l, l \in \mathcal{N}_{r_3} \setminus \mathcal{N}_{r_1}\}$ by $\{C_l = 0, l \in \mathcal{N}_{r_1}\}$ to obtain the Jordan normal form of $\{C_l, l \in \mathcal{N}_{r_1}\}$, denoted by B , and the remainder set, denoted by $\mathbf{C}_1 = \{g_i, i \in \mathcal{N}_r\}$.

Step 4. Apply Algorithm 4 to obtain the reduced minimal characterization set of $\mathcal{R}(\mathbf{C}_1)$, denoted by

$$S_M = \tilde{E} \cup S_{r'}. \text{ Write } S_{r'} = \{C_j \geq 0, j \in \mathcal{N}_{t_2}\}.$$

Step 5. Let $G = \tilde{\mathcal{R}}^{-1}(\tilde{E}) \cup B$ and compute the Jordan normal form of G , denoted by $\mathcal{B} = \{C_i, i \in \mathcal{N}_{t_1}\}$.

// In the above, the inequality set $\mathcal{R}(\mathbf{C}_1)$ is generated by reducing $\{C_l \geq 0, l \in \mathcal{N}_{r_3} \setminus \mathcal{N}_{r_1}\}$ by $\{C_l = 0, l \in \mathcal{N}_{r_1}\}$, and

// the inequality set $S_{r'}$ is generated by further reducing $\mathcal{R}(\mathbf{C}_1)$ by own implied equalities, which is equivalent to \tilde{E} .

// Therefore, in $S_{r'}$, only the free variables in the Jordan normal form \mathcal{B} are involved.

Step 6. Reduce F by $\tilde{R}(\mathcal{B})$ to obtain the remainder F_1 .

// In both F_1 and $S_{r'}$, only the free variables in the Jordan normal form \mathcal{B} are involved.

// The original Problem P₁ is now transformed into

// **Problem P₂**: Determine whether $F_1 \geq 0$ is implied by the inequalities in $S_{r'}$, i.e.,

$$\mathbb{C}_i \geq 0, j = 1, \dots, t_2.$$

// Since the equality set $\tilde{R}(\mathcal{B})$ contains only constraints on the pivot variables in \mathcal{B} , it is ignored in formulation of

// Problem P₂. The remaining steps follow Algorithm 3.

Step 7. Let $x_j, j \in \mathcal{N}_{n_1}$ be the variables in Problem P₂. Let $F_2 = F_1 - \sum_{i=1}^{t_2} p_i \mathbb{C}_i$, where

$$P = \{p_i, i \in \mathcal{N}_{t_2}\} \text{ is a set of variables. Set } F_2 \equiv \sum_{j=1}^{n_1} q_j x_j \equiv 0. \text{ Then } Q = \{q_j = 0, j \in \mathcal{N}_{n_1}\} \text{ is a linear system in } P.$$

Step 8. If the linear system Q has no solution, declare that the objective information inequality $\bar{F} \geq 0$ is ‘Not Provable’ and terminate the procedure.

Step 9. Otherwise, solve the linear equations $\{q_j = 0, j \in \mathcal{N}_{n_1}\}$ by Gauss-Jordan elimination to obtain the solution set of p_i in the form $\{p_i = P_i, i \in \mathcal{N}_{t_2}\}$, where P_i is a linear function in $t_2 - d_2$ variables of P and d_2 is the rank of the linear system Q .

Step 10. If $P_i \in \mathbb{R}_{<0}$ (the set of negative real numbers) for some $i \in \mathcal{N}_{t_2}$, declare ‘Not Provable’.

Step 11. Otherwise, let S_P be the set $\{P_i, i \in \mathcal{N}_{t_2}\}$, and let $\tilde{S}_P = S_P \setminus \mathbb{R}$. Write $\tilde{S}_P = \{\tilde{P}_i, i \in \mathcal{N}_{t_3}\}$.

If \tilde{S}_P is empty, the objective information inequality \bar{F} is proved. Otherwise go to Step 12.

Step 12. **Problem P₃**:

$$\begin{aligned} &\min(0) \\ \text{s.t. } &\tilde{P}_i \geq 0, i = 1, \dots, t_3. \end{aligned}$$

If the above LP has a solution, the objective information inequality $\bar{F} \geq 0$ is proved. Otherwise, declare ‘Not Provable’.

Remark V.1. Let $N_v(P_1)$, $N_v(P_2)$ and $N_v(P_3)$ be the number of variables in Problems P_1 , P_2 and P_3 respectively. Let $N_c(P_1)$, $N_c(P_2)$ and $N_c(P_3)$ be the number of constraints in Problems P_1 , P_2 and P_3 respectively. It is clear that $N_v(P_1) \geq N_v(P_2) \geq N_v(P_3)$, and $N_c(P_1) \geq N_c(P_2) \geq N_c(P_3)$. The reduction of the number of variables and the number of constraints is in general significant. Since most of the computation in the procedure is attributed to solving the LP in Problem P_3 , compared with the approach in Theorem II.2 where a much larger LP needs to be solved, the efficiency can be significantly improved. Example VI.1 illustrates this point.

B. Procedure II: Proving Information Identities

Input:

Objective information identity: $\bar{F} = 0$.

Additional constraints: $\bar{C}_i = 0$, $i = 1, \dots, r_1$; $\bar{C}_j \geq 0$, $j = r_1 + 1, \dots, r_2$.

Element information inequalities: $\bar{C}_k \geq 0$, $k = r_2 + 1, \dots, r_3$.

Here, \bar{F} , \bar{C}_i , \bar{C}_j , and \bar{C}_k are linear combination of information measures.

Output: A proof of $\bar{F} = 0$ if feasible.

Step 1. Construct the s -variable set S_n and the associated s -variable sequence S_n .

Step 2. Transform \bar{F} , \bar{C}_i , \bar{C}_j and \bar{C}_k to linear polynomials F , C_i , C_j and C_k in S_n respectively.

// We need to solve

// **Problem P₁**: Determine whether $F = 0$ is implied by

$$\begin{aligned} C_i &= 0, \quad i = 1, \dots, r_1, \\ C_j &\geq 0, \quad j = r_1 + 1, \dots, r_2, \\ C_k &\geq 0, \quad k = r_2 + 1, \dots, r_3. \end{aligned}$$

Step 3. Apply Algorithm 1 to reduce $\{C_l, l \in \mathcal{N}_{r_3} \setminus \mathcal{N}_{r_1}\}$ by $\{C_l = 0, l \in \mathcal{N}_{r_1}\}$ to obtain the Jordan normal form of $\{C_l, l \in \mathcal{N}_{r_1}\}$, denoted by B , and the remainder set, denoted by $\mathbf{C}_1 = \{g_i, i \in \mathcal{N}_r\}$.

Step 4. Apply Algorithm 4 to obtain the reduced minimal characterization set of $\mathcal{R}(\mathbf{C}_1)$, denoted by $S_M = \tilde{E} \cup S_{r'}$.

Step 5. Let $G = \tilde{\mathcal{R}}^{-1}(\tilde{E}) \cup B$ and compute the Jordan normal form of G , denoted by $\mathcal{B} = \{C_i, i \in \mathcal{N}_{l_1}\}$.

// The original problem P_1 has been transformed into

// **Problem P₂**: Determine whether $F = 0$ is implied by $\tilde{R}(\mathcal{B})$.

Step 6. Reduce F by $\tilde{R}(\mathcal{B})$ to obtain remainder F_1 . If $F_1 \equiv 0$, then the objective identity $\bar{F} = 0$ is proved.

Otherwise, declare ‘Not Provable’.

// As explained in Procedure I, F_1 involves only the free variables in the Jordan normal form \mathcal{B} . Therefore,

// if $F_1 \not\equiv 0$, the free variables can be chosen such that F_1 is evaluated to a nonzero value.

Remark V.2. An information identity $F = 0$ is equivalent to the two information inequalities $F \geq 0$ and $F \leq 0$. In the previous approach, in order to prove $F = 0$, $F \geq 0$ and $F \leq 0$ are proved separately by solving two LPs. In Procedure II, we transform the proof into a Gauss elimination problem, which greatly reduces the computational complexity.

Remark V.3. Procedures I and II can be implemented on the computer by Maple for symbolic computation. Therefore, they can give explicit proofs of information inequalities and identities.

VI. ILLUSTRATIVE EXAMPLES

In this section, we give two examples to illustrate Procedures I and II. The computation is performed by Maple.

A. Information Inequality under Equality Constraints

Example VI.1. $I(X_i; X_4) = 0$, $i = 1, 2, 3$ and $H(X_4|X_i, X_j) = 0$, $1 \leq i < j \leq 3 \Rightarrow H(X_i) \geq H(X_4)$.

Proof. By symmetry of the problem, we only need to prove $H(X_1) \geq H(X_4)$. The proof is given according to Procedure I.

Input:

Objective information inequality: $\bar{F} = H(X_1) - H(X_4) \geq 0$.

Equality Constraints: $\bar{C}_1 = I(X_1; X_4) = 0$, $\bar{C}_2 = I(X_2; X_4) = 0$, $\bar{C}_3 = I(X_3; X_4) = 0$,

$$\bar{C}_4 = H(X_4|X_1, X_2) = 0, \bar{C}_5 = H(X_4|X_1, X_3) = 0, \bar{C}_6 = H(X_4|X_2, X_3) = 0.$$

28 element information inequalities: $\bar{C}_k \geq 0, k \in \mathcal{N}_{34} \setminus \mathcal{N}_6$.

Step 1. The s -variables set contains 15 elements. The s -variable sequence $\mathcal{S}_4 = [s_{1,2,3,4}, s_{1,1,3,4}, s_{1,2,1,4}, s_{1,2,3,1}, s_{2,2,3,4}, s_{1,1,1,4}, s_{1,1,3,1}, s_{1,2,1,1}, s_{2,2,2,4}, s_{2,2,3,2}, s_{3,3,3,4}, s_{1,1,1,1}, s_{2,2,2,2}, s_{3,3,3,3}, s_{4,4,4,4}]$.

Step 2. We have $F = s_{1,1,1,1} + s_{1,1,3,1} + s_{1,2,1,1} + s_{1,2,3,1} - s_{2,2,2,4} - s_{2,2,3,4} - s_{3,3,3,4} - s_{4,4,4,4}$, $C_1 = s_{1,1,1,4} + s_{1,1,3,4} + s_{1,2,1,4} + s_{1,2,3,4}$, $C_2 = s_{1,2,1,4} + s_{1,2,3,4} + s_{2,2,2,4} + s_{2,2,3,4}$, $C_3 = s_{1,1,3,4} + s_{1,2,3,4} + s_{2,2,3,4} + s_{3,3,3,4}$, $C_4 = s_{3,3,3,4} + s_{4,4,4,4}$, $C_5 = s_{2,2,2,4} + s_{4,4,4,4}$, $C_6 = s_{1,1,1,4} + s_{4,4,4,4}$, and 28 linear polynomials $C_k, k \in \mathcal{N}_{34} \setminus \mathcal{N}_6$ are obtained from the 28 element information inequalities.

Step 3. Compute the Gauss-Jordan normal form of $\{C_i, i \in \mathcal{N}_6\}$, $B = \{s_{3,3,3,4} + s_{4,4,4,4}, s_{2,2,2,4} + s_{4,4,4,4}, s_{1,1,1,4} + s_{4,4,4,4}, s_{1,2,1,4} - s_{2,2,3,4}, s_{1,1,3,4} - s_{2,2,3,4}, s_{1,2,3,4} + 2s_{2,2,3,4} - s_{4,4,4,4}\}$. Use Algorithm 1 to reduce $\{C_l, l \in \mathcal{N}_{34} \setminus \mathcal{N}_6\}$ by $\tilde{R}(B)$ to obtain the remainder set $\mathbf{C}_1 = \{g_i, i \in \mathcal{N}_{18}\}$.

Step 4. Use Algorithm 4 to obtain $S_M = \tilde{E} \cup S_{r'}$ and $S_{r'} = \{C_i, i = 1, \dots, 10\}$, where

$$\begin{aligned} \mathbb{C}_1 &= s_{1,1,1,1}, \mathbb{C}_2 = s_{1,1,3,1}, \mathbb{C}_3 = s_{1,2,1,1}, \mathbb{C}_4 = s_{2,2,2,2}, \mathbb{C}_5 = s_{2,2,3,2}, \mathbb{C}_6 = s_{2,2,3,4}, \mathbb{C}_7 = s_{3,3,3,3}, \\ \mathbb{C}_8 &= s_{1,2,3,1} - s_{2,2,3,4} + s_{1,1,3,1}, \mathbb{C}_9 = s_{1,2,3,1} - s_{2,2,3,4} + s_{1,2,1,1}, \mathbb{C}_{10} = s_{1,2,3,1} - s_{2,2,3,4} + s_{2,2,3,2}. \end{aligned}$$

Step 5. Compute the Gauss-Jordan normal form $\mathcal{B} = \{s_{4,4,4,4}, s_{3,3,3,4}, s_{2,2,2,4}, s_{1,1,1,4}, s_{1,2,1,4} - s_{2,2,3,4}, s_{1,1,3,4} - s_{2,2,3,4}, s_{1,2,3,4} + 2s_{2,2,3,4}\}$.

Step 6. Reduce F by $\tilde{R}(\mathcal{B})$ to obtain $F_1 = s_{1,1,1,1} + s_{1,2,1,1} - s_{2,2,3,4} + s_{1,1,3,1} + s_{1,2,3,1}$.

Steps 7-11. We have $t_2 = 10, n_1 = 8, \bar{S}_P = \{p_9 + p_{10}, 1 - p_9, -p_{10}, 1 - p_9 - p_{10}, p_9, p_{10}\}$ and $S_P = \bar{S}_P \cup \{1, 0\}$.

Step 12. Solve the LP in Problem P_3 to complete the proof. Alternatively, we can solve the inequality set $\mathcal{R}(\bar{S}_P)$ to obtain the solution $\{0 \leq p_9 \leq 1, p_{10} = 0\}$. Substituting $p_9 = 0$ and $p_{10} = 0$ to $\{p_i = P_i, i \in \mathcal{N}_{10}\}$ yields $\{p_1 = 1, p_2 = 0, p_3 = 1, p_4 = 0, p_5 = 0, p_6 = 0, p_7 = 0, p_8 = 1, p_9 = 0, p_{10} = 0\}$. Thus an explicit proof is given by $F_1 = \mathbb{C}_1 + \mathbb{C}_3 + \mathbb{C}_8 \geq 0$. \square

Remark VI.1. Table I shows the advantage of Procedure I for Example VI.1 by comparing it with the Direct LP method induced by Theorem II.2.

TABLE I

	Number of variables	Number of equality constraints	Number of Inequality constraints
Direct LP method	15	6	28
LP in Problem P_3	2	0	6

B. Information Identity under Equality Constraints

Example VI.2. $I(X_1; X_2|X_3) = 0, H(X_3) = I(X_2; X_3|X_1) \Rightarrow H(X_1) = H(X_1|X_2, X_3)$.

Proof. The proof is given according to Procedure II.

Input:

Objective information inequality: $\bar{F} = H(X_1) - H(X_1|X_2, X_3) \geq 0$.

Equality Constraints: $\bar{C}_1 = I(X_1; X_2|X_3) = 0, \bar{C}_2 = H(X_3) - I(X_2; X_3|X_1) = 0$.

9 element information inequalities: $\bar{C}_k \geq 0, k \in \mathcal{N}_{11} \setminus \mathcal{N}_2$.

Step 1. The s -variables set contains 7 elements. The s -variable sequence $\mathcal{S}_3 = [s_{1,2,3}, s_{1,1,3}, s_{1,2,1}, s_{2,2,3}, s_{1,1,1}, s_{2,2,2}, s_{3,3,3}]$.

Step 2. We have $F = s_{1,1,3} + s_{1,2,1} + s_{1,2,3}$, $C_1 = s_{1,2,1}$, $C_2 = s_{1,1,3} + s_{1,2,3} + s_{3,3,3}$, $C_3 = s_{1,1,1}$, $C_4 = s_{2,2,2}$, $C_5 = s_{3,3,3}$, $C_6 = s_{1,2,1} + s_{1,2,3}$, $C_7 = s_{1,2,3} + s_{2,2,3}$, $C_8 = s_{1,1,3} + s_{1,2,3}$, $C_9 = s_{1,2,1}$, $C_{10} = s_{1,1,3}$ and $C_{11} = s_{2,2,3}$.

Step 3. Compute the Gauss-Jordan normal form $B = \{s_{1,2,1}, s_{1,1,3} + s_{1,2,3} + s_{3,3,3}\}$. Use Algorithm 1 to reduce $\{C_l, l \in \mathcal{N}_{11} \setminus \mathcal{N}_2\}$ by $\tilde{R}(B)$ to obtain the remainder set $\mathbf{C}_1 = \{g_i, i \in \mathcal{N}_8\}$, where $g_1 = s_{1,1,1}, g_2 = s_{2,2,2}, g_3 = s_{3,3,3}, g_4 = -s_{1,1,3} - s_{3,3,3}, g_5 = -s_{1,1,3} - s_{3,3,3} + s_{2,2,3}, g_6 = -s_{3,3,3}, g_7 = s_{1,1,3}, g_8 = s_{2,2,3}$.

Step 4. Use Algorithm 4 to obtain $S_M = \tilde{E} \cup S_{r'}$, where $\tilde{E} = \{s_{1,1,3} = 0, s_{3,3,3} = 0\}$.

Step 5. Compute the Gauss-Jordan normal form $\mathcal{B} = \{s_{1,2,3}, s_{1,1,3}, s_{1,2,1}, s_{3,3,3}\}$.

Step 6. Reduce F by \mathcal{B} to obtain $F_1 \equiv 0$. Thus the information identity is proved. \square

VII. CONCLUSION AND DISCUSSION

In this paper, we develop a new method to prove linear information inequalities and identities. Instead of solving an LP, we transform the problem into a polynomial reduction problem. For the proof of information inequalities, compared with existing methods (ITIP and its variations), our method takes advantage of the algebraic structure of the problem and greatly reduces the computational complexity. For the proof of information identities, we give a simple direct proof method which is much more efficient than existing methods.

ACKNOWLEDGMENT

This work is partially supported by NSFC 11688101, and Fundamental Research Funds for the Central Universities (2021NTST32).

REFERENCES

- [1] R. W. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1924-1934, Nov. 1997.
- [2] R. W. Yeung and C. T. Li, "Machine-Proving of Entropy Inequalities," to appear in *IEEE BITS the Information Theory Magazine*.
- [3] R. W. Yeung and Y.-O. Yan (1996), Information Theoretic Inequality Prover (ITIP), MATLAB Program Software Package. [Online]. Available: <http://home.ie.cuhk.edu.hk/~ITIP>
- [4] R. Pulikoonattu and S. Diggavi (2006), Xitip, ITIP-Based C Program Software Package. [Online]. Available: <http://xitip.epfl.ch>
- [5] L. Csirmaz (2016), A MINimal Information Theoretic Inequality Prover (Minitip). [Online]. Available: <https://github.com/lcsirmaz/minitip>.
- [6] C. T. Li (2020), Python Symbolic Information Theoretic Inequality Prover (psitip). [Online]. Available: <https://github.com/cheuktingli/>
- [7] N. Rathenakar, S. Diggavi, T. Gläßle, E. Perron, R. Pulikoonattu, R. W. Yeung, and Y.-O. Yan (2020), Online X-Information Theoretic Inequalities Prover (oXitip). [Online]. Available: <http://www.oxitip.com>
- [8] S.-W. Ho, L. Ling, C. W. Tan, and R. W. Yeung, "Proving and disproving information inequalities: Theory and scalable algorithms," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5522-536, Sep. 2020.
- [9] R. W. Yeung, *Information Theory and Network Coding*. New York, NY, USA: Springer, 2008.
- [10] R. W. Yeung, A. Al-Bashabsheh, C. Chen, Q. Chen, and P. Moulin, "On information-theoretic characterizations of Markov random fields and subfields," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1493-1511, 2018.
- [11] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. Inf. Theory*, vol. 44, pp. 1440-1452, July 1998.
- [12] T. Chan, S. Thakor, and A. Grant, "Minimal characterization of Shannon-type inequalities under functional dependence and full conditional independence structures," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4041-4051, Jul. 2019.
- [13] R. W. Yeung, "A new outlook on Shannon's information measures," *IEEE Trans. Inform. Theory*, vol. 37, pp. 466-74, May 1991.
- [14] J. Farkas, "Über die Theorie der einfachen Ungleichungen," *J. Reine Angew. Math.*, vol. 124, pp. 1-24, 1902.
- [15] D. Acharya, "An elementary proof of Farkas' lemma," *SIAM Review*, vol. 39, no. 3, pp. 503-07, 1997.
- [16] D. A. Cox, J. Little and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer Science & Business Media, 2013.
- [17] D. C. Lay, *Linear Algebra and Its Applications*, 5th Edition. Pearson, 2016.