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

Shannon's information measures refer to entropy, mutual information and their conditional versions. We need to prove various information inequalities and identities involving these information measures, for example, in converse coding theorems. However, proving an information inequality or identity with more than three 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 ITIP was developed [3]. Subsequently, variants of ITIP that expand its functions in different directions have been developed [4] [7] [5] [6] [8].

Using the LP-based approach, to prove an information identity $f = 0$, two LPs need to be solved, one for $f \geq 0$ and the other for $f \leq 0$.

Instead of transforming the problem of proving information inequalities into a general LP to be solved numerically, we develop procedures that can be implemented by symbolic computation. The reader is referred to [9, Chs. 13-15] for the background of this work, and to [15] for the the proofs omitted here.

Laigang Guo is with LMCS (MOE), School of Mathematical Sciences, Beijing Normal University, Beijing, China. Email: lgguo@bnu.edu.cn

Raymond W. Yeung is with Institute of Network Coding and Department of Information Engineering, The Chinese University of Hong Kong, Kong Kong, China. Email: whyeung@ie.cuhk.edu.hk

Xiao-Shan Gao is with KLMM, ISS, and AMSS of Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing, China. Email: xgao@mmrc.iss.ac.cn

The work of X.-S. Gao and L. Guo are partially supported by NSFC 11688101 and Fundamental Research Funds for the Central Universities (2021NTST32), respectively.

II. Information inequality preliminaries

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, $H(X|Y) \geq 0$ and $I(X;Y) \geq 0$ together imply $H(X) = H(X|Y) + I(X;Y) \geq 0$.

Throughout this paper, all random variables are discrete. Unless otherwise specified, all information expressions involve some or all of the random variables $X_1, X_2, \ldots, X_n$. Denote the set $\{1, 2, \ldots, n\}$ by $\mathcal{N}_n$.

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 these two elemental forms, called the *elemental inequalities*, form a proper subset of the basic inequalities. In [1], the minimality of the elemental inequalities is also proved. The total number of elemental inequalities is equal to $m \triangleq n + \binom{n}{2} 2^{n-1}$.

The elemental inequalities are called *unconstrained* information inequalities because they hold for all joint distributions of the random variables. On the other hand, information inequalities (identities) that hold under linear equality constraints on Shannon's information measures are called *constrained* information inequalities (identities).

Information inequalities that are implied by the basic inequalities are called *Shannon-type* inequalities. Most of the information inequalities that are known belong to this type. However, *non-Shannon-type* inequalities do exist, e.g., [10].

All Shannon's information measures can be expressed as a linear combination of joint entropies. For the random variables $X_1, X_2, \ldots, X_n$, 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 $\mathbb{R}^{2^n-1}$. Likewise, the linear equality constrains 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 proving a (linear) Shannon-type inequality can be formulated as a linear program (LP), which is described next.

978-1-6654-2159-1/22/$31.00 ©2022 IEEE

Let $\mathbf{h}$ be the column (2^n-1)-vector of the joint entropies of $X_1, X_2, \ldots, X_n$. The set of elemental inequalities can be written as $G\mathbf{h} \geq 0$, where G is an $m \times (2^n-1)$ matrix. 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, on which ITIP and its variants are based, 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*

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

is zero.

III. Linear inequalities and related algorithms

Let $\mathbf{x} = [x_1, x_2, \ldots, x_n]$, 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 polynomial set $P_f = \{f_i, i \in \mathcal{N}_m\}$ and the corresponding inequality set $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$, and a polynomial set $P_{\tilde{f}} = \{\tilde{f}_i, i \in \mathcal{N}_{\tilde{m}}\}$ and the corresponding equality set $E_{\tilde{f}} = \{\tilde{f}_i = 0, i \in \mathcal{N}_{\tilde{m}}\}$, we write $S_f = \mathcal{R}(P_f)$, $P_f = \mathcal{R}^{-1}(S_f)$, $E_{\tilde{f}} = \widetilde{\mathcal{R}}(P_{\tilde{f}})$ and $P_{\tilde{f}} = \widetilde{\mathcal{R}}^{-1}(E_{\tilde{f}})$. Also, we write $(f_i \geq 0) \in S_f$ to mean that the inequality $f_i \geq 0$ is in S_f.

Let $N_{>0} = \{1, 2, \ldots\}$, and $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$ be the sets of positive and nonnegative real numbers, respectively.

Definition III.1. *A linear polynomial F in $\mathbf{x}$ is called a conic combination of polynomials f_j in $\mathbf{x}$, $j = 1, \ldots, k$, if $F = \sum_{j=1}^k r_j f_j$ with $r_j \in \mathbb{R}_{\geq 0}$.*

Definition III.2. *The inequalities $f_1 \geq 0, f_2 \geq 0, \ldots, f_k \geq 0$ imply the inequality $f \geq 0$ if: $\mathbf{x}$ satisfies $f_1 \geq 0, f_2 \geq 0, \ldots, f_k \geq 0$ implies $\mathbf{x}$ satisfies $f \geq 0$.*

Definition III.3. *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.4. *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 $j \in \mathcal{N}_{m_1}$.*

Lemma III.1. *Given $h_1, \ldots, h_k, h \in \mathbb{R}_h[\mathbf{y}]$, $h_1 \geq 0, ..., h_k \geq 0$ imply $h \geq 0$ if and only if h is a conic combination of $h_1, \ldots, h_k$.*

This lemma, which generalizes [1, Theorem 2], is a consequence of Farkas' lemma [12] [13].

Definition III.5. *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.2. *Let $S_f = \{f_i(\mathbf{x}) \geq 0, i \in \mathcal{N}_m\}$ be an inequality set. Then f_k is an implied equality of S_f if and only if*

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

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

Let $E_{\tilde{f}}$ be the set of all implied equalities of S_f. Evidently, $\widetilde{\mathcal{R}}^{-1}(E_{\tilde{f}}) \subseteq \mathcal{R}^{-1}(S_f)$.

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

In the rest of the section, we will develop a few 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 IV for proving information inequalities and identities.

A. Dimension reduction of an inequality 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.

Proposition III.2. *Under the variable order $x_1 \prec x_2 \prec \cdots \prec x_n$, the linear equation system $E_{\tilde{f}}$ can be reduced by Gauss-Jordan elimination to the unique form*

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

where $k_1 < k_2 < \cdots < k_{\tilde{n}}$, 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 $\{x_j$, for $k_i < j < k_{i+1}, i \in \mathcal{N}_{\tilde{n}}\}$, with $k_{i+1} = n+1$ by convention. Furthermore, $\sum_{i\in\mathcal{N}_{\tilde{n}}} |U_i| = n - \tilde{n}$.

We call the equality set $\widetilde{E}$ the *Jordan normal form* of $E_{\tilde{f}}$. Likewise, we call the polynomial set $\widetilde{R}^{-1}(\widetilde{E})$ the Jordan normal form of $\widetilde{R}^{-1}(E_{\tilde{f}})$. 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* (the remainder if R_f is a singleton).

Algorithm 1 Dimension Reduction

Input: S_f, $E_{\tilde{f}}$.
Output: The remainder set R_f.
1: Compute $\widetilde{E}$ with $E_{\tilde{f}}$ by Proposition III.2.
2: Substitute x_{k_i} by U_i in P_f to obtain a set R.
3: Let $R_f = R\backslash\{0\}$.
4: **return** $\mathcal{R}(R_f)$.

Example III.1. *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 $P_{\tilde{f}} = \widetilde{\mathcal{R}}^{-1}(E_{\tilde{f}}) = \{\tilde{f}_1, \tilde{f}_2, \tilde{f}_3\}$.

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 $\widetilde{E} = \{x_{k_1} - U_1 = 0, x_{k_2} - U_2 = 0\}$, where $k_1 = 1,\ k_2 = 3,\ U_1 = -x_2,\ U_2 = 0$.

Using the equality constraints in $\widetilde{E}$, we substitute $x_1 = -x_2$ and $x_3 = 0$ into $P_f = \{f_1, f_2\}$ to obtain $R = \{0, x_2\}$. Hence $R_f = R\backslash\{0\} = \{x_2\}$. 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.

B. The implied equalities contained in an inequality set

Let $S_f = \{f_i \geq 0, i \in \mathcal{N}_m\}$ be a given inequality set. 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-Jordon 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 $L_k, k = 1, \ldots, m$ be the following linear programming problem:

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

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.2. *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-Jordon 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:*

$$\begin{aligned}
L_1:&\ \max(v_3 + v_4) &&\text{s.t. } v_3 + v_4 \geq 0,\ v_3 \geq 0,\ v_4 \geq 0.\\
L_2:&\ \max(v_4) &&\text{s.t. } v_3 + v_4 \geq 0,\ v_3 \geq 0,\ v_4 \geq 0.\\
L_3:&\ \max(v_3) &&\text{s.t. } v_3 + v_4 \geq 0,\ v_3 \geq 0,\ v_4 \geq 0.\\
L_4:&\ \max(v_4) &&\text{s.t. } v_3 + v_4 \geq 0,\ v_3 \geq 0,\ v_4 \geq 0.\\
L_5:&\ \max(0) &&\text{s.t. } v_3 + v_4 \geq 0,\ v_3 \geq 0,\ v_4 \geq 0.
\end{aligned}$$

- *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 $S'_f = \{f_5 \geq 0\} = \{x_2 + x_3 \geq 0\}$ constrainted by the equality set $E_{\tilde{f}}$. Finally, apply Algorithm 1 with S'_f and $E_{\tilde{f}}$ as inputs to obtain $R_f = \{x_3\}$. 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.2, $f_k = 0$ 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

Definition III.6. *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.

Proposition III.3. *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 $0 \notin \mathcal{R}^{-1}(S_{g'})$.*

The following corollary is immediate from Definition III.6 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 $h_1, \ldots, h_m \in \mathbb{R}_h[\mathbf{x}]$ 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.*

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.*

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

$$\begin{aligned} &\min(0)\\ \text{s.t.}\quad &\mathcal{Q}_{i,k} \geq 0,\ i \in \mathcal{M}\backslash\{k\}. \end{aligned}$$

8: **if** L_k can be solved **then**
9: $P_h := P_h\backslash\{h_k\}$, $\mathcal{M} := \mathcal{M}\backslash\{k\}$.
10: **end if**
11: **end if**
12: **end for**
13: **return** $\mathcal{R}(P_h)$.

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.1, Algorithm 3, called the Minimal Characterization Set Algorithm, finds a minimal characterization set of S_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 $h_i, i \in \mathcal{M}\backslash\{k\}$. 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

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

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

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.7. *The set $S_M = \widetilde{E} \cup S_{r'}$ is called the reduced minimal characterization set of S_f.*

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

The following algorithm finds 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 $\mathcal{R}^{-1}(S_f)\backslash\widetilde{\mathcal{R}}^{-1}(E_{\tilde{f}})$ 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** $S_M = \widetilde{E} \cup S_{r'}$.

By Proposition III.2 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.

IV. Procedures for Proving Information Inequalities and Identities

In [15], we introduce a set of variables called the *s-variables* which are obtained through a linear transformation of the joint entropies of $X_1, X_2, \ldots, X_n$ according to the theory of I-Measure [11] [9, Ch. 3]. The s-variables facilitate the implementation of the procedures to be discussed below.

A. Procedure I: Proving Information Inequalities

Input:
Objective information inequality: $\bar{F} \geq 0$.
Additional constraints: $\bar{C}_i = 0,\ i = 1, \ldots, r_1$; $\bar{C}_j \geq 0,\ j = r_1 + 1, \ldots, r_2$.
Element information inequalities: $\bar{C}_k \geq 0,\ k = r_2 + 1, \ldots, r_3$.
// 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 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 $\mathbf{P_1}$**: Determine whether $F \geq 0$ is implied by

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

Step 3. Apply Algorithm 1 to reduce $\{C_l, l \in \mathcal{N}_{r_3}\backslash\mathcal{N}_{r_1}\}$ 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 $S_M = \widetilde{E} \cup S_{r'}$. Write $S_{r'} = \{\mathbb{C}_j \geq 0, j \in \mathcal{N}_{t_2}\}$.

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

// In the above, the inequality set $\mathcal{R}(\mathbf{C}_1)$ is generated by reducing $\{C_l \geq 0, l \in \mathcal{N}_{r_3} \backslash \mathcal{N}_{r_1}\}$ 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 $\widetilde{E}$.
// Therefore, in $S_{r'}$, only the free variables in the Jordan normal form $\mathcal{B}$ are involved.

Step 6. Reduce F by $\widetilde{R}(\mathcal{B})$ 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 P_1 is now transformed into
// **Problem $\mathbf{P_2}$**: Determine whether $F_1 \geq 0$ is implied by the inequalities in $S_{r'}$, i.e.,

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

// Since the equality set $\widetilde{R}(\mathcal{B})$ contains only constraints on the pivot variables in $\mathcal{B}$, it is ignored in formulation of
// Problem P_2. The remaining steps follow Algorithm 3.

Step 7. Let $x_j, j \in \mathcal{N}_{n_1}$ be the variables in Problem P_2. Let $F_2 = F_1 - \sum_{i=1}^{t_2} p_i \mathbb{C}_i$, where
$P = \{p_i, i \in \mathcal{N}_{t_2}\}$ is a set of variables. Set $F_2 \equiv \sum_{j=1}^{n_1} q_j x_j \equiv 0$. Then $Q = \{q_j = 0, j \in \mathcal{N}_{n_1}\}$
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 elmination 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 $\bar{S}_P = S_P \backslash \mathbb{R}$. Write $\bar{S}_P = \{\bar{P}_i,\ i \in \mathcal{N}_{t_3}\}$.

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

Step 12. **Problem $\mathbf{P_3}$**:

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

If the above LP has a solution, the objective information inequality $\bar{F} \geq 0$ is proved.

Otherwise, declare 'Not Provable'.

B. Procedure II: Proving Information Identities

Input:
Objective information identity: $\bar{F} = 0$.
Additional constraints: $\bar{C}_i = 0,\ i = 1, \ldots, r_1$; $\bar{C}_j \geq 0,\ j = r_1 + 1, \ldots, r_2$.
Element information inequalities: $\bar{C}_k \geq 0,\ k = r_2 + 1, \ldots, r_3$.
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 $\mathcal{S}_n$.

Step 2. Transform $\bar{F},\ \bar{C}_i,\ \bar{C}_j$.

// We need to solve
// **Problem $\mathbf{P_1}$**: Determine whether $F = 0$ is implied by

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

Step 3. Apply Algorithm 1 to reduce $\{C_l, l \in \mathcal{N}_{r_3} \backslash \mathcal{N}_{r_1}\}$ 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
$S_M = \widetilde{E} \cup S_{r'}$.

Step 5. Let $G = \widetilde{\mathcal{R}}^{-1}(\widetilde{E}) \cup B$ and compute the Jordan normal form of G, denoted by $\mathcal{B} = \{\mathcal{C}_i, i \in \mathcal{N}_{t_1}\}$.
// The original problem P_1 has been transformed into
// **Problem $\mathbf{P_2}$**: Determine whether $F = 0$ is implied by $\widetilde{R}(\mathcal{B})$.

Step 6. Reduce F by $\widetilde{R}(\mathcal{B})$ 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 IV.1. *In Procedure II, we transform the proof of an information identity into a Gauss elimination problem, which greatly reduces the computational complexity compared with existing methods that need to solve two LPs.*

Remark IV.2. *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.*

V. An illustrative example

We give an example to illustrate Procedure I. The computation is performed by Maple.

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

The inequality above can be proved by applying Procedure I. The details can be found in [15]. Table I shows the advantage of Procedure I 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

VI. Conclusion

We have developed a new method to prove linear information inequalities and identities. Instead of solving an LP directly, we transform the problem into a polynomial reduction problem, significantly improving the computational efficiency.

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," *IEEE BITS the Information Theory Magazine*, 2021.

[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. Pulikkoonattu 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. Gläβle, E. Perron, R. Pulikkoonattu, 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] 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.

[11] R. W. Yeung, "A new outlook on Shannon's information measures," *IEEE Trans. Inform. Theory*, vol. 37, pp. 466-74, May 1991.

[12] J. Farkas, "Uber die Theorie der einfachen Ungleichungen," *J. Reine Angew. Math.*, vol. 124, pp. 1-24, 1902.

[13] D. Achiya, "An elementary proof of Farkas' lemma," *SIAM Review*, vol. 39, no. 3, pp. 503-07, Sep. 1997.

[14] D. C. Lay, S. R. Lay and J. J. McDonald, *Linear Algebra and Its Applications*, 5th Edition. New York: Pearson, 2016.

[15] L. Guo, R. W. Yeung, and X.-S. Gao, "Proving information inequalities and Identities with symbolic computaiton," preprint (full version), DOI: 10.13140/RG.2.2.25207.70562.