

Decision Problems in Information Theory

Mahmoud Abo Khamis

relationalAI, Berkeley, CA, USA

Phokion G. Kolaitis

UC Santa Cruz and IBM Research - Almaden, Santa Cruz, CA, USA

Hung Q. Ngo

relationalAI, Berkeley, CA, USA

Dan Suciu

University of Washington, Seattle, WA, USA

Abstract

Constraints on entropies are considered to be the laws of information theory. Even though the pursuit of their discovery has been a central theme of research in information theory, the algorithmic aspects of constraints on entropies remain largely unexplored. Here, we initiate an investigation of decision problems about constraints on entropies by placing several different such problems into levels of the arithmetical hierarchy. We establish the following results on checking the validity over all almost-entropic functions: first, validity of a Boolean information constraint arising from a monotone Boolean formula is co-recursively enumerable; second, validity of “tight” conditional information constraints is in Π_3^0 . Furthermore, under some restrictions, validity of conditional information constraints “with slack” is in Σ_2^0 , and validity of information inequality constraints involving max is Turing equivalent to validity of information inequality constraints (with no max involved). We also prove that the classical implication problem for conditional independence statements is co-recursively enumerable.

2012 ACM Subject Classification Mathematics of computing → Information theory; Theory of computation → Computability; Theory of computation → Complexity classes

Keywords and phrases Information theory, decision problems, arithmetical hierarchy, entropic functions

1 Introduction

The study of constraints on entropies is a central topic of research in information theory. In fact, more than 30 years ago, Pippenger [38] asserted that constraints on entropies are the “*laws of information theory*” and asked whether the *polymatroidal axioms* form the complete laws of information theory, i.e., whether every constraint on entropies can be derived from the polymatroidal axioms. These axioms consist of the following three types of constraints: (1) $H(\emptyset) = 0$, (2) $H(X) \leq H(X \cup Y)$ (monotonicity), and (3) $H(X) + H(Y) \geq H(X \cap Y) + H(X \cup Y)$ (submodularity). It is known that the polymatroidal axioms are equivalent to Shannon’s basic inequalities, that is, to the non-negativity of the entropy, conditional entropy, mutual information, and conditional mutual information [45]. In a celebrated result published in 1998, Zhang and Yeung [50] answered Pippenger’s question negatively by finding a linear inequality that is satisfied by all entropic functions, but cannot be derived from the polymatroidal axioms.

Zhang and Yeung’s result became the catalyst for the discovery of other information laws that are not captured by the polymatroidal axioms (e.g., [24, 33]). In particular, we now know that there are more elaborate laws, such as conditional inequalities, or inequalities expressed using max, which find equally important applications in a variety of areas. For example, implications between conditional independence statements of discrete random variables can be expressed as conditional information inequalities. In another example, we have recently

shown that conjunctive query containment under bag semantics is at least as hard as checking information inequalities using max [1]. Despite the extensive research on various kinds of information inequalities, to the best of our knowledge nothing is known about the algorithmic aspects of the associated decision problem: check whether a given information law is valid.

In this paper, we initiate a study of algorithmic problems that arise naturally in information theory, and establish several results. To this effect, we introduce a generalized form of information inequalities, which we call *Boolean information constraints*, consisting of Boolean combinations of linear information inequalities, and define their associated decision problems. Since it is still an open problem whether linear information inequalities, which are the simplest kind of information laws, are decidable, we focus on placing these decision problems in the arithmetical hierarchy, also known as the Kleene-Mostowski hierarchy [39]. The arithmetic hierarchy has been studied by mathematical logicians since the late 1940s; moreover, it directly influenced the introduction and study of the polynomial-time hierarchy by Stockmeyer [41]. The first level of the arithmetical hierarchy consists of the collection Σ_1^0 of all recursively enumerable sets and the collection Π_1^0 of the complements of all recursively enumerable sets. The higher levels Σ_n^0 and Π_n^0 , $n \geq 2$, are defined using existential and universal quantification over lower levels. We prove a number of results, including the following.

- (1) Checking the validity of a Boolean information constraint arising from a monotone Boolean formula (in particular, a max information inequality) is in Π_1^0 (Theorem 6).
- (2) Checking the validity of a conditional information inequality whose antecedents are “tight” is in Π_3^0 (Corollary 10). “Tight” inequalities are defined in Section 4.2.2, and include conditional independence assertions between random variables.
- (3) Checking the validity of a conditional information inequality whose antecedents have “slack” and are group-balanced is in Σ_2^0 (Corollary 13).
- (4) Checking the validity of a group-balanced, max information inequality is Turing equivalent to checking the validity of an information inequality (Corollary 16).

While the decidability of linear information inequalities (the simplest kind considered in this paper) remains open, a separate important question is whether more complex Boolean information constraints are any harder. For example, some conditional inequalities, or some max-inequalities can be proven from a simple linear inequality, hence they do not appear to be any harder. However, Kaced and Romashchenko [24] proved that there exist conditional inequalities that are *essentially conditional*, which means that they do not follow from a linear inequality. (We give an example in Equation (9).) We prove here that any conditional information inequality with slack is *essentially unconditioned* (Corollary 9; see also Equation(17)), and that any max-inequality also follows from a single linear inequality (Theorem 15).

A subtle complication involving these results is whether by “validity” it is meant that the given Boolean information constraint holds for the set of all entropic vectors over n variables, denoted by Γ_n^* , or for its topological closure, denoted by $\bar{\Gamma}_n^*$. It is well known that these two spaces differ for all $n \geq 3$. With the exception of (1) above, which holds for both Γ_n^* and $\bar{\Gamma}_n^*$, our results are only for $\bar{\Gamma}_n^*$. A problem of special interest is the implication between conditional independence statements of discrete random variables, and this amounts to checking the Γ_n^* -validity of a tight conditional information inequality; it is known that this problem is not finitely axiomatizable [42], and its decidability remains open. Our result (2) above does not apply here because it is a statement about $\bar{\Gamma}_n^*$ -validity. However, we prove that the implication problem for conditional independence statements is in Π_1^0 (Theorem 7).

2 Background and notations

Throughout this paper, vectors and tuples are denoted by bold-faced letters, and random variables are capitalized. We write $\mathbf{x} \cdot \mathbf{y} \stackrel{\text{def}}{=} \sum_i x_i y_i$ for the dot product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$. For a given set $S \subseteq \mathbb{R}^m$, S is *convex* if $\mathbf{x}, \mathbf{y} \in S$ and $\theta \in [0, 1]$ implies $\theta\mathbf{x} + (1 - \theta)\mathbf{y} \in S$; S is called a *cone* if $\mathbf{x} \in S$ and $\theta \geq 0$ implies $\theta\mathbf{x} \in S$; the topological closure of S is denoted by \overline{S} ; and, finally, $S^* \stackrel{\text{def}}{=} \{\mathbf{y} \mid \forall \mathbf{x} \in S, \mathbf{x} \cdot \mathbf{y} \geq 0\}$ denotes the *dual cone* of S . It is known that S^* is always a closed, convex cone. We provide more background in Appendix A.

For a random variable X with a fixed finite domain D and a probability mass function (pmf) p , its (binary) *entropy* is defined by

$$H(X) \stackrel{\text{def}}{=} - \sum_{x \in D} p(x) \cdot \log p(x) \quad (1)$$

In this paper all logarithms are in base 2.

Fix a joint distribution over n finite random variables $\mathbf{V} \stackrel{\text{def}}{=} \{X_1, \dots, X_n\}$. For each $\alpha \subseteq [n]$, let \mathbf{X}_α denote the random (vector-valued) variable $(X_i : i \in \alpha)$. Define the set function $h : 2^{[n]} \rightarrow \mathbb{R}_+$ by setting $h(\alpha) \stackrel{\text{def}}{=} H(\mathbf{X}_\alpha)$, for all $\alpha \subseteq [n]$. With some abuse, we blur the distinction between the set $[n]$ and the set of variables $\mathbf{V} = \{X_1, \dots, X_n\}$, and write $h(\mathbf{X}_\alpha)$ or $h(\alpha)$ interchangeably. We call the function h an *entropic function*, and also identify it with a vector $\mathbf{h} \stackrel{\text{def}}{=} (h(\alpha))_{\alpha \subseteq [n]} \in \mathbb{R}_+^{2^n}$, which is called an *entropic vector*. Note that most texts and papers on this topic drop the component $h(\emptyset)$, which is always 0, leading to entropic vectors in $\mathbb{R}_+^{2^n-1}$. We prefer to keep the \emptyset -coordinate to simplify notations. The implicit assumption $h(\emptyset) = 0$ is used through the rest of the paper.

The set of entropic functions/vectors is denoted by $\Gamma_n^* \subseteq \mathbb{R}_+^{2^n}$. Its topological closure, denoted by $\overline{\Gamma}_n^*$, is the set of *almost entropic* vectors (or functions). It is known [45] that $\Gamma_n^* \subsetneq \overline{\Gamma}_n^*$ for $n \geq 3$. In general, Γ_n^* is neither a cone nor convex, but its topological closure $\overline{\Gamma}_n^*$ is a closed convex cone [45].

Every entropic function h satisfies the following *basic Shannon inequalities*:

$$h(\mathbf{Y} \cup \mathbf{X}) \geq h(\mathbf{X}) \quad h(\mathbf{X}) + h(\mathbf{Y}) \geq h(\mathbf{X} \cup \mathbf{Y}) + h(\mathbf{X} \cap \mathbf{Y})$$

called *monotonicity* and *submodularity* respectively. Any inequality obtained by taking a positive linear combination of Shannon inequalities is called a *Shannon-type inequality*.

Throughout this paper we will abbreviate the union $\mathbf{X} \cup \mathbf{Y}$ of two sets of variables as \mathbf{XY} . The quantities $h(\mathbf{Y}|\mathbf{X}) \stackrel{\text{def}}{=} h(\mathbf{XY}) - h(\mathbf{X})$ and $I_h(\mathbf{Y}; \mathbf{Z}|\mathbf{X}) \stackrel{\text{def}}{=} h(\mathbf{XY}) + h(\mathbf{XZ}) - h(\mathbf{XYZ}) - h(\mathbf{X})$ are called the *conditional entropy* and the *conditional mutual information* respectively. It can be easily checked that $h(\mathbf{Y}|\mathbf{X}) \geq 0$ and $I_h(\mathbf{Y}; \mathbf{Z}|\mathbf{X}) \geq 0$ are Shannon-type inequalities.

► **Remark 1.** The established notation Γ_n^* [46, 49, 11] for the set of entropic vectors is unfortunate, because the star in this context does **not** represent the dual cone. We will continue to denote by Γ_n^* the set of entropic vectors (which is not a cone!), and use explicit parentheses, as in $(\Gamma_n^*)^*$, to represent the dual cone.

3 Boolean information Constraints

Most of this paper considers the following problem: given a Boolean combination of information inequalities, check whether it is valid. In Appendix G, we will briefly discuss the dual problem, namely, recognizing whether a given vector \mathbf{h} is an entropic vector (or an almost entropic vector).

A *Boolean function* is a function $F : \{0, 1\}^m \rightarrow \{0, 1\}$. We often denote its inputs with variables $Z_1, \dots, Z_m \in \{0, 1\}$, and write $F(Z_1, \dots, Z_m)$ for the value of the Boolean function.

3.1 Problem Definition

A vector $\mathbf{c} \in \mathbb{R}^{2^n}$ defines the following (linear) *information inequality*:

$$\mathbf{c} \cdot \mathbf{h} = \sum_{\alpha \in [n]} c_\alpha h(X_\alpha) \geq 0. \quad (2)$$

The information inequality is said to be *valid* if it holds for all vectors $\mathbf{h} \in \Gamma_n^*$; equivalently, \mathbf{c} is in the dual cone, $\mathbf{c} \in (\Gamma_n^*)^*$. By continuity, an information inequality holds $\forall \mathbf{h} \in \Gamma_n^*$ iff it holds $\forall \mathbf{h} \in \bar{\Gamma}_n^*$. In 1986, Pippenger [38] defined the “*laws of information theory*” as the set of all information inequalities, and asked whether all of them are Shannon-type inequalities. This was answered negatively by Zhang and Yeung in 1998 [50]. We know today that several applications require more elaborate laws, such as max-inequalities and conditional inequalities. Inspired by these new laws, we define the following generalization.

► **Definition 2.** To each Boolean function F with m inputs, and every m vectors $\mathbf{c}_j \in \mathbb{R}^{2^n}$, $j \in [m]$, we associate the following Boolean information constraint:

$$F(\mathbf{c}_1 \cdot \mathbf{h} \geq 0, \dots, \mathbf{c}_m \cdot \mathbf{h} \geq 0). \quad (3)$$

For a set $S \subseteq \mathbb{R}^{2^n}$, a Boolean information constraint is said to be *S-valid* if it holds for all $\mathbf{h} \in S$. Thus, we will distinguish between Γ_n^* -validity and $\bar{\Gamma}_n^*$ -validity. Unlike in the case of information inequality, these two notions of validity no longer coincide for arbitrary Boolean information constraints in general, as we explain in what follows.

► **Definition 3.** Let F be a Boolean function. The entropic Boolean information constraint problem parameterized by F , denoted by $\text{EBIC}(F)$, is the following: given m integer vectors $\mathbf{c}_j \in \mathbb{Z}^{2^n}$, where $j \in [m]$, check whether the constraint (3) holds for all entropic functions $\mathbf{h} \in \Gamma_n^*$. In the almost-entropic version, denoted by $\text{AEBIC}(F)$, we replace Γ_n^* by $\bar{\Gamma}_n^*$.

The inputs $\mathbf{c}_j, j \in [m]$, to these problems are integer vectors, so that $\text{EBIC}(F)$ and $\text{AEBIC}(F)$ are meaningful computational problems; equivalently, the inputs can be rational vectors $\mathbf{c}_j \in \mathbb{Q}^{2^n}$, $j \in [m]$.

Let F be a Boolean function. F can be written as a conjunction of clauses $F = C_1 \wedge C_2 \wedge \dots$, where each clause is a disjunction of literals. Equivalently, a clause C has this form:

$$(Z'_1 \wedge \dots \wedge Z'_k) \Rightarrow (Z_1 \vee \dots \vee Z_\ell) \quad (4)$$

Checking $\text{EBIC}(F)$ is equivalent to checking $\text{EBIC}(C)$, for each clause of F (and similarly for $\text{AEBIC}(F)$); therefore and without loss of generality, we will assume in the rest of the paper that F consists of a single clause (4) and study the problem along these dimensions:

Conditional and Unconditional Constraints When $k = 0$ (i.e., when the antecedent is empty), the formula F is *monotone*, and we call the corresponding Boolean information constraint *unconditional*. If, furthermore, $\ell = 1$, then we say that F defines a *simple inequality*; if $\ell > 1$, then we say that F defines a *max-inequality*. When $k > 0$, the formula F is *non-monotone*, and we call the corresponding constraint *conditional*.

Note that the case when $\ell = 0$ is not interesting because F is not valid, since the zero-vector $\mathbf{h} = \mathbf{0}$ violates the constraint.

Problem	Abbreviation		Simple Example
	Entropic	Almost-entropic	
Boolean information constraint	EBIC(F)	AEBIC(F)	$h(XY) \leq \frac{2}{3}h(XYZ) \Rightarrow \max(h(YX), h(XZ)) \geq \frac{2}{3}h(XYZ)$
Information Inequality	IIP		$h(XY) + h(YZ) + h(XZ) \geq 2h(XYZ)$
Max-Information Inequality	MaxIIP		$\max(h(XY), h(YZ), h(XZ)) \geq \frac{2}{3}h(XYZ)$
Conditional Information Inequalities	ECIIP	AECIIP	$((h(XY) \leq \frac{2}{3}h(XYZ)) \wedge (h(YZ) \leq \frac{2}{3}h(XYZ))) \Rightarrow h(XZ) \geq \frac{2}{3}h(XYZ)$
Conditional Independence	CI	(no name)	$(I(X; Y) = 0 \wedge I(X; Z Y) = 0) \Rightarrow I(X; Z) = 0$

■ **Figure 1** Notations for various Boolean Information Constraint Problems.

3.2 Examples and applications

This section presents examples and applications of Boolean Function Information Constraints and their associated decision problems. A summary of the notations is in Fig. 1, and more applications are described in Appendix B.

3.2.1 Information Inequalities

We start with the simplest form of a Boolean information constraint, namely, the linear information inequality in Eq. (2), which arise from the single-variable Boolean formula Z_1 . We will call the corresponding decision problem the *information-inequality problem*, denoted by IIP: given a vector of integers \mathbf{c} , check whether Eq. (2) is Γ_n^* -valid or, equivalently, $\bar{\Gamma}_n^*$ -valid. Pippenger’s question from 1986 was essentially a question about decidability. Shannon-type inequalities are decidable in exponential time using linear programming methods, and software packages have been developed for this purpose [45, Chapter 13] (it is not known, however, if there is a matching lower bound in the complexity of this problem). Thus, if every information inequality were a Shannon-type inequality, then information inequalities would be decidable. However, Zhang and Yeung’s gave the first example of a non-Shannon-type information inequality [50]. Later, Matúš [33] proved that, when $n \geq 4$ variables, there exists infinitely many inequivalent non-Shannon entropic inequalities. More precisely, he proved that the following is a non-Shannon inequality, for every $k \geq 1$:

$$I_h(C; D|A) + \frac{k+3}{2}I_h(C; D|B) + I_h(A; B) + \frac{k-1}{2}I_h(B; C|D) + \frac{1}{k}I_h(B; D|C) \geq I_h(C; D) \quad (5)$$

This ruined any hope of proving decidability of information inequalities by listing a finite set of axioms. To date, the study of non-Shannon-type inequalities is an active area of research [48, 30, 47], and the question whether IIP is decidable remains open.

Hammer et al. [23], showed that, up to logarithmic precision, information inequalities are equivalent to linear inequalities in Kolmogorov complexity (see also [20, Theorem 3.5]).

3.2.2 Max information inequalities

Next, we consider constraints defined by a disjunction of linear inequalities, in other words $(\mathbf{c}_1 \cdot \mathbf{h} \geq 0) \vee \dots \vee (\mathbf{c}_m \cdot \mathbf{h} \geq 0)$, where $\mathbf{c}_j \in \mathbb{R}^{2^n}$. This is equivalent to:

$$\max(\mathbf{c}_1 \cdot \mathbf{h}, \mathbf{c}_2 \cdot \mathbf{h}, \dots, \mathbf{c}_m \cdot \mathbf{h}) \geq 0 \quad (6)$$

and, for that reason, we call them *Max information inequalities* and denote the corresponding decision problem by **MaxIIP**. As before, Γ_n^* -validity and $\bar{\Gamma}_n^*$ -validity coincide.

Application to Constraint Satisfaction and Database Theory Given two finite structures \mathbf{A} and \mathbf{B} , we write $\text{HOM}(\mathbf{A}, \mathbf{B})$ for the set of homomorphisms from \mathbf{A} to \mathbf{B} . We say that \mathbf{B} *dominates* structure \mathbf{A} , denote by $\mathbf{A} \leq \mathbf{B}$, if for every finite structure \mathbf{C} , we have that $|\text{HOM}(\mathbf{A}, \mathbf{C})| \leq |\text{HOM}(\mathbf{B}, \mathbf{C})|$. The *homomorphism domination problem* asks whether $\mathbf{A} \leq \mathbf{B}$, given \mathbf{A} and \mathbf{B} . In database theory this problem is known as the *query containment problem under bag semantics* [13]. In that setting we are given two Boolean conjunctive queries Q_1, Q_2 , which we interpret using bag semantics, i.e., given a database D , the answer $Q_1(D)$ is the number of homomorphisms $Q_1 \rightarrow D$ [27]. Q_1 is *contained in* Q_2 under bag semantics if $Q_1(D) \leq Q_2(D)$ for every database D . It is open whether the homomorphism domination problem is decidable.

Kopparty and Rossman [28] described a **MaxIIP** problem that yields a sufficient condition for homomorphism domination. In recent work [1] we proved that, when \mathbf{B} is acyclic, then that condition is also necessary, and, moreover, the domination problem for acyclic \mathbf{B} is Turing-equivalent to **MaxIIP**. Hence, any result on the complexity of **MaxIIP** immediately carries over to the homomorphism domination problem for acyclic \mathbf{B} , and vice versa.

We illustrate here Kopparty and Rossman's **MaxIIP** condition on a simple example. Consider the following two Boolean conjunctive queries: $Q_1() = R(u, v) \wedge R(v, w) \wedge R(w, u)$, $Q_2() = R(x, y) \wedge R(x, z)$; interpreted using bag semantics, Q_1 returns the number of triangles and Q_2 the number of V-shaped subgraphs. Kopparty and Rossman proved that $Q_1 \leq Q_2$ follows from the following max-inequality:

$$\max\{2h(XY) - h(X) - h(XYZ), 2h(YZ) - h(Y) - h(XYZ), 2h(XZ) - h(Z) - h(XYZ)\} \geq 0 \quad (7)$$

3.2.3 Conditional Information Inequalities

A *conditional information inequality* has the form:

$$(c_1 \cdot \mathbf{h} \geq 0 \wedge \dots \wedge c_k \cdot \mathbf{h} \geq 0) \Rightarrow c_0 \cdot \mathbf{h} \geq 0 \quad (8)$$

Here we need to distinguish between Γ_n^* -validity and $\bar{\Gamma}_n^*$ -validity, and denote by **ECIIP** and **AECIIP** the corresponding decision problems. Notice that, without loss of generality, we can allow equality in the antecedent, because $c_i \cdot \mathbf{h} = 0$ is equivalent to $c_i \cdot \mathbf{h} \geq 0 \wedge -c_i \cdot \mathbf{h} \geq 0$.

Suppose that there exists $\lambda_1 \geq 0, \dots, \lambda_m \geq 0$ such that the inequality $c_0 \cdot \mathbf{h} - (\sum_i \lambda_i c_i \cdot \mathbf{h}) \geq 0$ is valid; then Eq. (8) is, obviously, also valid. Kaced and Romashchenko [24] called Eq. (8) an *essentially conditioned inequality* if no such λ_i 's exists, and discovered several valid conditional inequalities that are essentially conditioned.

Application to Conditional Independence Fix three joint random variables X, Y, Z . A *conditional independence* (CI) statement is a statement of the form $\phi = (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X})$, and it asserts that \mathbf{Y} and \mathbf{Z} are independent conditioned on \mathbf{X} . A *CI implication* is a statement $\varphi_1 \wedge \dots \wedge \varphi_k \Rightarrow \varphi_0$, where $\varphi_i, i \in \{0, \dots, k\}$ are CI statements. The *CI implication problem* is: given an implication, check if it is valid for all discrete probability distributions. Since $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}) \Leftrightarrow I_h(\mathbf{Y}; \mathbf{Z} \mid \mathbf{X}) = 0 \Leftrightarrow -I_h(\mathbf{Y}; \mathbf{Z} \mid \mathbf{X}) \geq 0$, the CI implication problem is a special case of **ECIIP**.

The CI implication problem has been studied extensively in the literature [29, 42, 18, 26]. Pearl and Paz [37] gave a sound, but incomplete, set of *graphoid axioms*, Studený [42] proved that no finite axiomatization exists, while Geiger and Pearl [18] gave a complete

axiomatization for two restricted classes, called saturated, and marginal CIs. See [16, 21, 36] for some recent work on the CI implication problem.

Results in [24] imply that the following CI implication is essentially conditioned (see [26]):

$$I_h(C; D|A) = I_h(C; D|B) = I_h(A; B) = I_h(B; C|D) = 0 \implies I_h(C; D) = 0 \quad (9)$$

3.2.4 Other Applications

In Appendix B, we describe several other important applications of Boolean information inequalities.

4 Placing EBIC and AEBIC in the Arithmetical Hierarchy

What is the complexity of $\text{EBIC}(F)$ / $\text{AEBIC}(F)$? Is it even decidable? As we have seen there are numerous applications of the Boolean Information Constraint problem, hence any positive or negative answer, even for special cases, would shed light on these applications. While their (un)decidability is currently open, in this paper we provide several upper bounds on their complexity, by placing them in the arithmetical hierarchy.

We briefly review some concepts from computability theory. In this setting it is standard to assume objects are encoded as natural numbers. A set $A \subseteq \mathbb{N}^k$, for $k \geq 1$, is *Turing computable*, or *decidable*, if there exists a Turing machine that, given $x \in \mathbb{N}^k$ decides whether $x \in A$. A set A is *Turing reducible* to B if there exists a Turing machine with an oracle for B that can decide membership in A . The *arithmetical hierarchy* consists of the classes of sets Σ_n^0 and Π_n^0 defined inductively as follows. The class Σ_n^0 consists of all sets of the form $\{x \mid \exists y_1 \forall y_2 \exists y_3 \dots Q y_n R(x, y_1, \dots, y_n)\}$, where R is an $(n+1)$ -ary decidable predicate, $Q = \exists$ if n is odd, and $Q = \forall$ if n is even. In a dual manner, the class Π_n^0 consists of sets of the form $\{x \mid \forall y_1 \exists y_2 \forall y_3 \dots Q y_n R(x, y_1, \dots, y_n)\}$. Then $\Sigma_0^0 = \Pi_0^0$ are the decidable sets, while Σ_1^0 consists of the *recursively enumerable* sets, and Π_1^0 consists of the *co-recursively enumerable* sets. It is known that these classes are closed under union and intersection, but not under complements, and that they form a strict hierarchy, $\Sigma_n^0, \Pi_n^0 \subsetneq (\Sigma_{n+1}^0 \cap \Pi_{n+1}^0)$. For more background, we refer to [39]. Our goal is to place the problems $\text{EBIC}(F)$, $\text{AEBIC}(F)$, and their variants to concrete levels of the arithmetical hierarchy.

4.1 Unconditional Boolean information constraints

We start by discussing unconditional Boolean information constraints, or, equivalently, a Boolean information constraint defined by a monotone Boolean formula F . The results here are rather simple; we include them only as a warmup for the less obvious results in later sections. Based on our discussion in Sections 3.2.1 and 3.2.2, we have the following result.

► **Theorem 4.** *If F is monotone, then $\text{EBIC}(F)$ and $\text{AEBIC}(F)$ are equivalent problems.*

Next, we prove that these problems are co-recursively enumerable, by using the following folklore fact. A *representable set of n random variables* is a finite relation Ω with N rows and $n+1$ columns X_1, \dots, X_n, p , where column p contains rational probabilities in $[0, 1] \cap \mathbb{Q}$ that sum to 1. Thus, Ω defines n random variables with finite domain and probability mass given by rational numbers. We denote \mathbf{h}^Ω its entropic vector. By continuity of Eq.(1), we obtain:

► **Proposition 5.** *For every entropic vector $\mathbf{h} \in \Gamma_n^*$ and every $\varepsilon > 0$, there exists a representable space Ω such that $\|\mathbf{h} - \mathbf{h}^\Omega\| < \varepsilon$.*

The group-characterization proven by Chan and Yeung [12] implies a much stronger version of the proposition; we do not need that stronger version in this paper.

► **Theorem 6.** *Let F be a monotone Boolean formula. Then $\text{EBIC}(F)$ (and, hence, $\text{AEBIC}(F)$) is in Π_1^0 , i.e., it is co-recursively enumerable.*

Proof. Fix $F = Z_1 \vee \dots \vee Z_m$ and $\mathbf{c}_i \in \mathbb{Z}^{2^n}$, $i \in [m]$. We need to check:

$$\forall \mathbf{h} \in \Gamma_n^* : \quad \mathbf{c}_1 \cdot \mathbf{h} \geq 0 \vee \dots \vee \mathbf{c}_m \cdot \mathbf{h} \geq 0 \quad (10)$$

We claim that (10) is equivalent to:

$$\forall \Omega \quad \mathbf{c}_1 \cdot \mathbf{h}^\Omega \geq 0 \vee \dots \vee \mathbf{c}_m \cdot \mathbf{h}^\Omega \geq 0 \quad (11)$$

Obviously (10) implies (11), and the opposite follows from Prop. 5: if (10) fails on some entropic vector \mathbf{h} , then it also fails on some representable \mathbf{h}^Ω close enough to \mathbf{h} . Finally, (11) is in Π_1^0 because, the property after $\forall \Omega$ is decidable, by expanding the definition of entropy (1) in each condition $\mathbf{c}_i \cdot \mathbf{h}^\Omega \geq 0$, and writing the latter as $\sum_j a_j \log b_j \geq 0$, or, equivalently, $\prod_j (b_j)^{a_j} \geq 1$, where a_j, b_j are rational numbers, which is decidable. ◀

4.2 Conditional Boolean Information Constraints

We now consider non-monotone Boolean functions, in other words, conditional information constraints (8). Since Γ_n^* - and $\bar{\Gamma}_n^*$ -validity no longer coincide, we study $\text{EBIC}(F)$ and $\text{AEBIC}(F)$ separately. The results here are non-trivial, and some proofs are in the Appendix.

4.2.1 The entropic case

Recall that the CI implication problem consists of checking whether an implication between statements of the form $(Y \perp\!\!\!\perp Z \mid X)$ holds for all random variables with finite domain. Our first result is about the entropic version of this problem.

► **Theorem 7.** *The entropic CI implication problem (Section 3.2.3) is in Π_1^0 .*

The proof relies on Tarski's Theorem, which states that the theory of reals with $+, *$ is decidable (see [44]). More precisely, there is an algorithm for the following decision problem: given a first-order formula Φ with function symbols $+$ and $*$, is Φ true on the structure $(\mathbb{R}, +, *, 0, 1)$ of the real numbers? For example, Tarski's algorithm determines whether or not the formula¹ $\forall x \exists y \forall z (x^2 + 3y \geq z \wedge (y^3 + yz \leq xy^2))$ is true on $(\mathbb{R}, +, *, 0, 1)$. Our proof amounts to describing a semi-decision procedure for checking that an instance of the CI implication problem is not valid: iterate over all finite probability distributions and, using Tarski's Theorem, check whether there exist probabilities p_1, p_2, \dots (real numbers in $[0, 1]$) for which the CI implication fails. The details are in Appendix C.

Tarski's exponential function problem

We used Tarski's Theorem of the decidability of the first-order theory of the reals $(\mathbb{R}, +, *, 0, 1)$, to prove that the entropic CI implication problem is in Π_1^0 (Theorem 7). The crux of the proof is to notice that each CI statement $(Y \perp\!\!\!\perp Z \mid X)$ can be expressed in this logic as a set of equalities of probabilities, one for each outcome of the random variables X, Y, Z . A

¹ $3y$ is a shorthand for $y + y + y$ and $x \geq y$ is a shorthand for $\exists u (x = y + u^2)$.

major open problem in model theory, originally formulated by Tarski, is whether decidability continues to hold if we augment the structure of the real numbers with the exponential function (see, e.g., [31] for a discussion). Decidability of the first-order theory of the reals with exponentiation would easily imply that the entropic conditional information inequality problem ECIP (not just the entropic conditional independence (CI) implication problem) is in Π_1^0 , because every condition $\mathbf{c} \cdot \mathbf{h} \geq 0$ can be expressed using $+$, $*$ and the exponential function, by simply expanding the definition of entropy in Equation (1).

4.2.2 The almost-entropic case

Suppose the antecedent of (8) includes the condition $\mathbf{c} \cdot \mathbf{h} \geq 0$. Call $\mathbf{c} \in \mathbb{R}^{2^n}$ *tight* if $\mathbf{c} \cdot \mathbf{h} \leq 0$ is $\bar{\Gamma}_n^*$ -valid. When \mathbf{c} is tight, we can rewrite $\mathbf{c} \cdot \mathbf{h} \geq 0$ as $\mathbf{c} \cdot \mathbf{h} = 0$. If \mathbf{c} is not tight, then there exists $\mathbf{h} \in \bar{\Gamma}_n^*$ such that $\mathbf{c} \cdot \mathbf{h} > 0$; in that case we say that \mathbf{c} *has slack*. For example, all conditions occurring in CI implications are tight, because they are of the form $-I_h(Y; Z|X) \geq 0$, and more conveniently written $I_h(Y; Z|X) = 0$, while a condition like $3h(X) - 4h(YZ) \geq 0$ has slack. We extend the definition of slack to a set. We say that the set $\{\mathbf{c}_1, \dots, \mathbf{c}_k\} \subset \mathbb{R}^{2^n}$ has slack if there exists $\mathbf{h} \in \bar{\Gamma}_n^*$ such that $\mathbf{c}_i \cdot \mathbf{h} > 0$ for all $i = 1, k$; notice that this is more restricted than requiring each of \mathbf{c}_i to have slack. We present below results on the complexity of AEBIC(F) in two special cases: when all antecedents are tight, and when the set of antecedents has slack. Both results use the following theorem, which allows us to move one condition $\mathbf{c}_k \cdot \mathbf{h} \geq 0$ from the antecedent to the consequent:

► **Theorem 8.** *The following statements are equivalent:*

$$\forall \mathbf{h} \in \bar{\Gamma}_n^* : \quad \bigwedge_{i \in [k]} \mathbf{c}_i \cdot \mathbf{h} \geq 0 \Rightarrow \mathbf{c} \cdot \mathbf{h} \geq 0 \quad (12)$$

$$\forall \varepsilon > 0, \exists \lambda \geq 0, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \quad \bigwedge_{i \in [k-1]} \mathbf{c}_i \cdot \mathbf{h} \geq 0 \Rightarrow \mathbf{c} \cdot \mathbf{h} + \varepsilon h([n]) \geq \lambda \mathbf{c}_k \cdot \mathbf{h} \quad (13)$$

Moreover, if the set $\{\mathbf{c}_1, \dots, \mathbf{c}_k\}$ has slack, then one can set $\varepsilon = 0$ in Eq.(13).

Proof. We prove here only the implication from (13) to (12); the other direction is non-trivial and is proven in Appendix D using only the properties of closed convex cones. Assume condition (13) holds, and consider any $\mathbf{h} \in \bar{\Gamma}_n^*$ s.t. $\bigwedge_{i \in [k]} \mathbf{c}_i \cdot \mathbf{h} \geq 0$. We prove that $\mathbf{c} \cdot \mathbf{h} \geq 0$. For any $\varepsilon > 0$, condition (13) states that there exists $\lambda > 0$ such that $\mathbf{c} \cdot \mathbf{h} + \varepsilon h([n]) \geq \lambda \mathbf{c}_k \cdot \mathbf{h}$ and therefore $\mathbf{c} \cdot \mathbf{h} + \varepsilon h([n]) \geq 0$. Since $\varepsilon > 0$ is arbitrary, we conclude that $\mathbf{c} \cdot \mathbf{h} \geq 0$, as required. ◀

By applying the theorem repeatedly, we can move all antecedents to the consequent:

► **Corollary 9.** *Condition (12) is equivalent to:*

$$\forall \varepsilon > 0, \exists \lambda_1 \geq 0, \dots, \exists \lambda_k \geq 0, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \quad \mathbf{c} \cdot \mathbf{h} + \varepsilon h([n]) \geq \sum_{i \in [k]} \lambda_i \mathbf{c}_i \cdot \mathbf{h} \quad (14)$$

Moreover, if the set $\{\mathbf{c}_1, \dots, \mathbf{c}_k\}$ has slack, then one can set $\varepsilon = 0$ in Eq.(14).

Antecedents Are Tight We consider now the case when all antecedents are tight, a condition that can be verified in Π_1^0 , by Th.6. In that case, condition (12) is equivalent to:

$$\forall p \in \mathbb{N}, \exists q \in \mathbb{N}, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \quad \mathbf{c} \cdot \mathbf{h} + \frac{1}{p} h([n]) \geq q \sum_{i \in [k]} \mathbf{c}_i \cdot \mathbf{h} \quad (15)$$

Indeed, the non-trivial direction (14) \Rightarrow (15) follows by setting $q \stackrel{\text{def}}{=} \lceil \max(\lambda_1, \dots, \lambda_k) \rceil \in \mathbb{N}$ and noting that \mathbf{c}_i is tight, hence $\mathbf{c}_i \cdot \mathbf{h} \leq 0$ and therefore $\lambda_i \mathbf{c}_i \cdot \mathbf{h} \geq q \mathbf{c}_i \cdot \mathbf{h}$.

► **Corollary 10.** *Consider a conditional inequality (8). If all antecedents are tight, then the corresponding decision problem AECIP is in Π_3^0*

Proof. Condition (15) is of the form $\forall p \exists q \forall \mathbf{h}$. Replace \mathbf{h} with a representable entropic vector \mathbf{h}^Ω , as in the proof of Theorem 6, and it becomes $\forall p \exists q \forall \mathbf{h}^\Omega$, placing it in Π_3^0 . ◀

An immediate consequence is that checking $\bar{\Gamma}_n^*$ -validity of a CI implication (Sec. 3.2.3) is in Π_3^0 because all antecedents are tight. Note that, although every CI implication is tight, Corollary 10 does *not* apply to Γ_n^* -validity. This is why we used Tarski's Theorem to prove that Γ_n^* -validity of a CI implication is in Π_1^0 (Theorem 7).

It follows from our discussion that *any* tight conditional implication follows from a family of (unconditional) inequalities (15). For example, consider the CI implication (9) (Sec. 3.2.3). Write it in the form (12) by replacing each antecedent $I_h(C; D | A) = 0$ with $-I_h(C; D | A) \geq 0$, and we conclude from (15) that the following holds:

$$q(I_h(C; D | A) + I_h(C; D | B) + I_h(A; B) + I_h(B; C | D)) + \frac{1}{p}h(ABCD) \geq I_h(C; D) \quad (16)$$

Thus, in order to prove (9), it suffices to prove (16). Matúš's inequality (5) provides precisely the proof of (16) (by setting $k \stackrel{\text{def}}{=} p$, $q \stackrel{\text{def}}{=} \max(\lceil \frac{k+3}{2} \rceil, 1)$, and observing that $I_h(B; D | C) \leq h(ABCD)$).

Antecedents Have Slack Next, we consider the case when the antecedents have slack, which is a recursively enumerable condition. In that case, condition (14) is equivalent to:

$$\exists \lambda_1 \geq 0, \dots, \exists \lambda_k \geq 0, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \quad \mathbf{c} \cdot \mathbf{h} \geq \sum_{i \in [k]} \lambda_i \mathbf{c}_i \cdot \mathbf{h} \quad (17)$$

In other words, we have proven the following result of independent interest: any conditional implication with slack is essentially unconditioned. However, we cannot immediately use (17) to prove complexity bounds for AEBIC(F), because the λ_i 's in (17) are not necessarily rational numbers; we can no longer increase or decrease λ_i because the sign of $\mathbf{c}_i \cdot \mathbf{h}$ is unknown. We prove below that, under a restriction called *group balance*, the λ_i 's can be chosen in \mathbb{Q} , placing the decision problem in Σ_2^0 . Group balance generalizes Chan's notion of a *balanced inequality*, which we review below. In Appendix F we give evidence that some restriction is necessary to ensure the λ_i 's are rationals (Example 28), and also show that every conditional inequality can be strengthened to be group balanced (Prop 29).

A vector $\mathbf{h} \in \mathbb{R}^{2^n}$ is called *modular* if $h(\mathbf{X}) + h(\mathbf{Y}) = h(\mathbf{X} \cup \mathbf{Y}) + h(\mathbf{X} \cap \mathbf{Y})$ for all sets of variables $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$. Every non-negative modular function is entropic [45], and is a non-negative linear combination of the *basic modular functions* $\mathbf{h}^{(1)}, \dots, \mathbf{h}^{(n)}$, where $h^{(j)}(\alpha) \stackrel{\text{def}}{=} 1$ when $j \in \alpha$ and is $h^{(j)}(\alpha) \stackrel{\text{def}}{=} 0$ otherwise. Chan [22] called an inequality $\mathbf{c} \cdot \mathbf{h} \geq 0$ *balanced* if $\mathbf{c} \cdot \mathbf{h}^{(j)} = 0$ for every $j \in [n]$. He proved that *any* valid inequality can be strengthened to a balanced one. More precisely: $\mathbf{c} \cdot \mathbf{h} \geq 0$ is valid iff $\mathbf{c} \cdot \mathbf{h}^{(i)} \geq 0$ for all $i \in [n]$ and $\mathbf{c} \cdot \mathbf{h} - \sum_i (\mathbf{c} \cdot \mathbf{h}^{(i)})h(X_i | X_{[n]-\{i\}}) \geq 0$ is valid; notice that the latter inequality is balanced. For example, $h(XY) + h(XZ) - h(X) - h(XYZ) \geq 0$ is balanced, while $h(XY) - h(X) \geq 0$ is not balanced, and can be strengthened to $h(XY) - h(X) - h(Y|X) \geq 0$. We generalize Chan's definition:

► **Definition 11.** *Call a set $\{\mathbf{d}_1, \dots, \mathbf{d}_k\} \subseteq \mathbb{R}^{2^n}$ group balanced if (a) $\text{rank } \mathbf{A} = k - 1$ where \mathbf{A} is the $k \times n$ matrix $A_{ij} = \mathbf{d}_i \cdot \mathbf{h}^{(j)}$, and (b) there exists a non-negative modular function $\mathbf{h}^{(*)} \neq 0$ such that $\mathbf{d}_i \cdot \mathbf{h}^{(*)} = 0$ for all i .*

If $k = 1$ then $\{\mathbf{d}_1\}$ is group balanced iff \mathbf{d}_1 is balanced, because the matrix \mathbf{A} has a single row $(\mathbf{d}_1 \cdot \mathbf{h}^{(1)} \dots \mathbf{d}_1 \cdot \mathbf{h}^{(n)})$, and its rank is 0 iff all entries are 0. We prove in Appendix F:

► **Theorem 12.** *Consider a group balanced set of n vectors with rational coefficients, $D = \{\mathbf{d}_1, \dots, \mathbf{d}_n\} \subseteq \mathbb{Q}^{2^n}$. Suppose the following condition holds:*

$$\exists \lambda_1 \geq 0, \dots, \exists \lambda_n \geq 0, \sum_{i \in [n]} \lambda_i = 1, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \sum_{i \in [n]} \lambda_i \mathbf{d}_i \cdot \mathbf{h} \geq 0 \quad (18)$$

Then there exists rational $\lambda_1, \dots, \lambda_k \geq 0$ with this property.

This implies that, if $\mathbf{c}_1, \dots, \mathbf{c}_k$ have slack and $\{\mathbf{c}, -\mathbf{c}_1, \dots, -\mathbf{c}_k\}$ is group balanced, then there exists rational λ_i ' for inequality (17). In particular:

► **Corollary 13.** *Consider a conditional inequality (8). If the antecedents have slack and $\{\mathbf{c}, -\mathbf{c}_1, \dots, -\mathbf{c}_k\}$ is group balanced, then the corresponding decision problem is in Σ_2^0 .*

We end this section by illustrating with an example:

► **Example 14.** Consider the following conditional inequality:

$$h(XYZ) + h(X) \geq 2h(XY) \wedge h(XYZ) + h(Y) \geq 2h(YZ) \Rightarrow 2h(XZ) \geq h(XYZ) + h(Z) \quad (19)$$

The antecedents have slack, because, by setting² $\mathbf{h} \stackrel{\text{def}}{=} 2\mathbf{h}^{(X)} + \mathbf{h}^{(Z)}$, both antecedents become strict inequalities: $h(XYZ) + h(X) - 2h(XY) = 3 + 2 - 4 > 0$ and $h(XYZ) + h(Y) - 2h(YZ) = 3 + 0 - 2 > 0$. To check validity, we prove in Example 17 the following inequality:

$$(2h(XY) - h(XYZ) - h(X)) + (2h(YZ) - h(XYZ) - h(Y)) + (2h(XZ) - h(XYZ) - h(Z)) \geq 0$$

and this immediately implies (19).

Consider now the following set $D = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$, where the vectors $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ represent the expressions $2h(XY) - h(XYZ) - h(X)$, $2h(YZ) - h(XYZ) - h(Y)$, and $2h(XZ) - h(XYZ) - h(Z)$ respectively. We prove that D is group balanced. To check condition (a) of Def. 11

we verify that the matrix \mathbf{A} has rank 2; in our example the matrix is $\mathbf{A} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$

and its rank is 2 as required. To check condition (b), we define $\mathbf{h}^{(*)} = \mathbf{h}^{(X)} + \mathbf{h}^{(Y)} + \mathbf{h}^{(Z)}$ and verify that $\mathbf{d}_1 \cdot \mathbf{h}^{(*)} = \mathbf{d}_2 \cdot \mathbf{h}^{(*)} = \mathbf{d}_3 \cdot \mathbf{h}^{(*)} = 4 - 3 - 1 = 0$. Thus, D is group balanced.

4.3 Discussion on the Decidability of MaxIIP

A proof of the decidability of MaxIIP would immediately imply that the domination problem $\mathbf{A} \leq \mathbf{B}$ for acyclic structures \mathbf{B} is also decidable [1]. It is currently open whether MaxIIP is decidable, or even if the special case IIP is decidable. But what can we say about the domination problem if IIP were decidable? Theorem 6 only says that both problems are in Π_1^0 , and does not tell us anything about MaxIIP if IIP were decidable. We prove here that, the decidability of IIP implies the decidability of group-balanced MaxIIP. We start with a result of general interest, which holds even for conditional Max-Information constraints.

² Where $\mathbf{h}^{(X)}$ denotes the basic modular function at X , i.e. $h^{(X)}(X) = 1$, $h^{(X)}(Y) = h^{(X)}(Z) = 0$.

► **Theorem 15.** *The following two statements are equivalent:*

$$\forall \mathbf{h} \in \bar{\Gamma}_n^* : \bigwedge_{i \in [k]} \mathbf{c}_i \cdot \mathbf{h} \geq 0 \Rightarrow \bigvee_{j \in [m]} \mathbf{d}_j \cdot \mathbf{h} \geq 0 \quad (20)$$

$$\exists \lambda_1, \dots, \lambda_m \geq 0, \sum_j \lambda_j = 1, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \bigwedge_{i \in [k]} \mathbf{c}_i \cdot \mathbf{h} \geq 0 \Rightarrow \sum_{j \in [m]} \lambda_j \mathbf{d}_j \cdot \mathbf{h} \geq 0 \quad (21)$$

The theorem says that every max-inequality is essentially a linear inequality. The proof of (21) \Rightarrow (20) is immediate; we prove the reverse in Appendix E. As before, we don't know whether these coefficients λ_i can be chosen to be rational numbers in general, but by Theorem 12 this is the case when $\{\mathbf{c}_1, \dots, \mathbf{c}_k\}$ is group-balanced, and this implies:

► **Corollary 16.** *The MaxIIP problem where the inequalities $\mathbf{c}_1, \dots, \mathbf{c}_n$ are group balanced is Turing equivalent to the IIP problem.*

Proof. We describe a Turing reduction from MaxIIP to IIP. Consider a MaxIIP problem, $\bigvee_{j \in [m]} (\mathbf{c}_j \cdot \mathbf{h} \geq 0)$. We run two computations in parallel. The first computation iterates over all representable spaces Ω , and checks whether $\bigwedge_j (\mathbf{c}_j \cdot \mathbf{h}^\Omega < 0)$; if we find such a space then we stop and we return *false*. If the inequality is invalid then this computation will eventually terminate because in that case there exists a representable counterexample Ω . The second computation iterates over all m -tuples of natural numbers $(\lambda_1, \dots, \lambda_m) \in \mathbb{N}^m$ and checks $\forall \mathbf{h} \in \bar{\Gamma}_n^*, \sum_j \lambda_j \mathbf{c}_j \cdot \mathbf{h} \geq 0$ by using the oracle for IIP: if it finds such λ_j 's, then it stops and returns *true*. If the inequality is valid then this computation will eventually terminate, by Theorems 15 and 12. ◀

We illustrate with an example.

► **Example 17.** Consider Kopparty and Rossman's inequality (7), which can be stated as $\max(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3) \geq 0$, where $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ define the three expressions in (7). To prove that it is valid, it suffices to prove that their sum is ≥ 0 ; we show this briefly here³:

$$\begin{aligned} & (2h(XY) - h(X)) + (2h(YZ) - h(Y)) + (2h(XZ) - h(Z)) - 3h(XYZ) \\ &= (h(XY) + h(YZ) + h(XZ)) + (h(XY) - h(X)) + (h(YZ) - h(Y)) + (h(XZ) - h(Z)) - 3h(XYZ) \\ &\geq (h(XY) + h(YZ) + h(XZ)) + (h(XYZ) - h(XZ)) + (h(XYZ) - h(XY)) + \\ &\quad (h(XYZ) - h(YZ)) - 3h(XYZ) = 0 \end{aligned}$$

Theorem 15 proves that *any* max-inequality necessarily follows from such a linear inequality, we just have to find the right λ_i 's. In this example, the set $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ is group balanced (as we showed in Example 14), therefore there exists rational λ_i 's; indeed, our choice here is $\lambda_1 = \lambda_2 = \lambda_3 = 1$.

5 Discussion

Finite, infinite, continuous random variables. In this paper, all random variables have a finite domain. There are two alternative choices: discrete random variables (possibly infinite), and continuous random variables. The literature on entropic functions has mostly alternated between defining entropic functions over finite r.v., or over discrete infinite r.v. with finite entropy. For example discrete (possibly infinite) random variables are

³ We apply submodularity: $h(XY) - h(X) \geq h(XYZ) - h(XZ)$ etc.

considered by Zhang and Yeung, [49], by Chan and Yeung [12], and by Chan [22], while random variables with finite domains are considered by Matúš [32, 33] and by Kaced and Romashchenko [24]. The reason for this inconsistency is that for information inequalities the distinction doesn't matter: every entropy of a set of discrete random variables can be approximated arbitrarily well by the entropy of a set of random variables with finite domain, and Prop. 5 extends immediately to discrete random variables⁴. However, the distinction is significant for conditional inequalities, and here the choice in the literature is always for finite domains. For example, the implication problem for conditional independence, i.e. the graphoid axioms, is stated for finite probability spaces by Geiger and Pearl [18], while Kaced and Romashchenko [24] also use finite distributions to prove the existence of conditional inequalities that hold over entropic but fail for almost-entropic functions. One could also consider continuous distributions, whose entropy is $\int p(x) \log(1/p(x)) dx$, where p is the probability density function. Chan [22] showed that an information inequality holds for all continuous distributions iff it is balanced and it holds for all discrete distributions. For example, $h(X) \geq 0$ is not balanced, hence it fails in the continuous, because the entropy of the uniform distribution in the interval $[0, c]$ is $\log c$, which is < 0 when $c < 1$.

Strict vs. non-strict inequalities. The literature on information inequalities always defines inequalities using ≥ 0 , in which case validity for entropic functions is the same as validity for almost entropic functions. One may wonder what happens if one examines strict inequalities $\mathbf{c} \cdot \mathbf{h} > 0$ instead. Obviously, each such inequality fails on the zero-entropic vector, but we can consider the conditional version $\mathbf{h} \neq 0 \Rightarrow \mathbf{c} \cdot \mathbf{h} > 0$, which we can write formally as $\mathbf{c} \cdot \mathbf{h} \leq 0 \Rightarrow h(\mathbf{V}) \leq 0$. This is a special case of a conditional inequality as discussed in this paper. An interesting question is whether for this special case Γ_n^* -validity and $\bar{\Gamma}_n^*$ -validity coincide; a negative answer would represent a significant extension of Kaced and Romashchenko's result [24].

References

- 1 Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. Bag query containment and information theory. In *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, CA, USA, 2020*. to appear.
- 2 Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu. Computing join queries with functional dependencies. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 327–342, 2016.
- 3 Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu. What do Shannon-type inequalities, submodular width, and disjunctive Datalog have to do with one another? In *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2017*, pages 429–444, 2017.
- 4 Noga Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel J. Math.*, 38(1-2):116–130, 1981. URL: <http://dx.doi.org/gate.lib.buffalo.edu/10.1007/BF02761855>, doi:10.1007/BF02761855.
- 5 Albert Atserias, Martin Grohe, and Daniel Marx. Size bounds and query plans for relational joins. *SIAM J. Comput.*, 42(4):1737–1767, 2013.

⁴ The idea of the proof relies on the fact that every entropy is required to converge, i.e. $h(X_\alpha) = \sum_i p_i \log 1/p_i$, hence there exists a finite subspace of outcomes $\{1, 2, \dots, N\}$ for which the sum is ε -close to $h(X_\alpha)$. The union of these spaces over all $\alpha \subseteq [n]$ suffices to approximate h well enough.

- 6 Amos Beimel. Secret-sharing schemes: A survey. In Yeow Meng Chee, Zhenbo Guo, San Ling, Fengjing Shao, Yuansheng Tang, Huaxiong Wang, and Chaoping Xing, editors, *Coding and Cryptology - Third International Workshop, IWCC 2011, Qingdao, China, May 30-June 3, 2011. Proceedings*, volume 6639 of *Lecture Notes in Computer Science*, pages 11–46. Springer, 2011. URL: https://doi.org/10.1007/978-3-642-20901-7_2, doi:10.1007/978-3-642-20901-7_2.
- 7 G. R. Blakley. Safeguarding cryptographic keys. In *Managing Requirements Knowledge, International Workshop on*, page 313, Los Alamitos, CA, USA, jun 1979. IEEE Computer Society. URL: <https://doi.ieeecomputersociety.org/10.1109/AFIPS.1979.98>, doi:10.1109/AFIPS.1979.98.
- 8 Carlo Blundo, Alfredo De Santis, and Ugo Vaccaro. On secret sharing schemes. *Inf. Process. Lett.*, 65(1):25–32, 1998. URL: [https://doi.org/10.1016/S0020-0190\(97\)00194-4](https://doi.org/10.1016/S0020-0190(97)00194-4), doi:10.1016/S0020-0190(97)00194-4.
- 9 Stephen P. Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, 2014. URL: <https://web.stanford.edu/%7Eboyd/cvxbook/>, doi:10.1017/CBO9780511804441.
- 10 Renato M. Capocelli, Alfredo De Santis, Luisa Gargano, and Ugo Vaccaro. On the size of shares for secret sharing schemes. *J. Cryptology*, 6(3):157–167, 1993. URL: <https://doi.org/10.1007/BF00198463>, doi:10.1007/BF00198463.
- 11 Terence H. Chan. Group characterizable entropy functions. In *IEEE International Symposium on Information Theory, ISIT 2007, Nice, France, June 24-29, 2007*, pages 506–510. IEEE, 2007.
- 12 Terence H. Chan and Raymond W. Yeung. On a relation between information inequalities and group theory. *IEEE Transactions on Information Theory*, 48(7):1992–1995, 2002.
- 13 Surajit Chaudhuri and Moshe Y. Vardi. Optimization of *Real* conjunctive queries. In Catriel Beeri, editor, *Proceedings of the Twelfth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 25-28, 1993, Washington, DC, USA*, pages 59–70. ACM Press, 1993.
- 14 F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Combin. Theory Ser. A*, 43(1):23–37, 1986. URL: [http://dx.doi.org.gate.lib.buffalo.edu/10.1016/0097-3165\(86\)90019-1](http://dx.doi.org.gate.lib.buffalo.edu/10.1016/0097-3165(86)90019-1), doi:10.1016/0097-3165(86)90019-1.
- 15 László Csirmaz. The size of a share must be large. *J. Cryptology*, 10(4):223–231, 1997. URL: <https://doi.org/10.1007/s001459900029>, doi:10.1007/s001459900029.
- 16 Peter R. de Waal and Linda C. van der Gaag. Stable independance and complexity of representation. In David Maxwell Chickering and Joseph Y. Halpern, editors, *UAI '04, Proceedings of the 20th Conference in Uncertainty in Artificial Intelligence, Banff, Canada, July 7-11, 2004*, pages 112–119. AUAI Press, 2004. URL: https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1165&proceeding_id=20.
- 17 Ehud Friedgut and Jeff Kahn. On the number of copies of one hypergraph in another. *Israel J. Math.*, 105:251–256, 1998. URL: <http://dx.doi.org.gate.lib.buffalo.edu/10.1007/BF02780332>, doi:10.1007/BF02780332.
- 18 Dan Geiger and Judea Pearl. Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics*, 21(4):2001–2021, 1993.
- 19 Georg Gottlob, Stephanie Tien Lee, Gregory Valiant, and Paul Valiant. Size and treewidth bounds for conjunctive queries. *J. ACM*, 59(3):16:1–16:35, 2012.
- 20 Peter Grunwald and Paul Vitányi. Shannon information and Kolmogorov complexity. *arXiv preprint cs/0410002*, 2004.
- 21 Marc Gyssens, Mathias Niepert, and Dirk Van Gucht. On the completeness of the semigraphoid axioms for deriving arbitrary from saturated conditional independence statements. *Inf. Process. Lett.*, 114(11):628–633, 2014. URL: <https://doi.org/10.1016/j.ipl.2014.05.010>, doi:10.1016/j.ipl.2014.05.010.

- 22 Terence H. Chan. Balanced information inequalities. *Information Theory, IEEE Transactions on*, 49:3261 – 3267, 01 2004. doi:10.1109/TIT.2003.820037.
- 23 Daniel Hammer, Andrei Romashchenko, Alexander Shen, and Nikolai Vereshchagin. Inequalities for Shannon entropy and Kolmogorov complexity. *Journal of Computer and System Sciences*, 60(2):442–464, 2000.
- 24 Tarik Kaced and Andrei E. Romashchenko. Conditional information inequalities for entropic and almost entropic points. *IEEE Trans. Information Theory*, 59(11):7149–7167, 2013. URL: <https://doi.org/10.1109/TIT.2013.2274614>, doi:10.1109/TIT.2013.2274614.
- 25 Ehud D. Karnin, J. W. Greene, and Martin E. Hellman. On secret sharing systems. *IEEE Trans. Information Theory*, 29(1):35–41, 1983. URL: <https://doi.org/10.1109/TIT.1983.1056621>, doi:10.1109/TIT.1983.1056621.
- 26 Batya Kenig and Dan Suciu. Integrity constraints revisited: From exact to approximate implication. *CoRR*, abs/1812.09987, 2018. URL: <http://arxiv.org/abs/1812.09987>, arXiv:1812.09987.
- 27 Phokion G. Kolaitis and Moshe Y. Vardi. Conjunctive-query containment and constraint satisfaction. In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington, USA*, pages 205–213, 1998. URL: <https://doi.org/10.1145/275487.275511>, doi:10.1145/275487.275511.
- 28 Swastik Kopparty and Benjamin Rossman. The homomorphism domination exponent. *European Journal of Combinatorics*, 32(7):1097 – 1114, 2011. Homomorphisms and Limits.
- 29 Tony T. Lee. An information-theoretic analysis of relational databases - part I: data dependencies and information metric. *IEEE Trans. Software Eng.*, 13(10):1049–1061, 1987.
- 30 Konstantin Makarychev, Yury Makarychev, Andrei Romashchenko, and Nikolai Vereshchagin. A new class of non-Shannon-type inequalities for entropies. *Commun. Inf. Syst.*, 2(2):147–165, 2002. URL: <https://doi-org.gate.lib.buffalo.edu/10.4310/CIS.2002.v2.n2.a3>, doi:10.4310/CIS.2002.v2.n2.a3.
- 31 David Marker. Model theory and exponentiation, 1996.
- 32 Frantisek Matúš. Probabilistic conditional independence structures and matroid theory: Background, 1994.
- 33 Frantisek Matúš. Infinitely many information inequalities. In *IEEE International Symposium on Information Theory, ISIT 2007, Nice, France, June 24-29, 2007*, pages 41–44, 2007.
- 34 C.F. Miller. Decision problems for groups — survey and reflections. In G. Baumslag and C.F. Miller, editors, *Algorithms and Classification in Combinatorial Group Theory. Mathematical Sciences Research Institute Publications*, volume 23. Springer, NY, 1992.
- 35 Hung Q. Ngo. Worst-case optimal join algorithms: Techniques, results, and open problems. In Jan Van den Bussche and Marcelo Arenas, editors, *Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Houston, TX, USA, June 10-15, 2018*, pages 111–124. ACM, 2018. URL: <https://doi.org/10.1145/3196959.3196990>, doi:10.1145/3196959.3196990.
- 36 Mathias Niepert, Marc Gyssens, Bassem Sayrafi, and Dirk Van Gucht. On the conditional independence implication problem: A lattice-theoretic approach. *Artif. Intell.*, 202:29–51, 2013. URL: <https://doi.org/10.1016/j.artint.2013.06.005>, doi:10.1016/j.artint.2013.06.005.
- 37 Judea Pearl and Azaria Paz. Graphoids: Graph-based logic for reasoning about relevance relations or when would x tell you more about y if you already know z? In *ECAI*, pages 357–363, 1986.
- 38 Nicholas Pippenger. What are the laws of information theory. In *1986 Special Problems on Communication and Computation Conference*, pages 3–5, 1986.
- 39 Hartley Rogers and H Rogers. *Theory of recursive functions and effective computability*, volume 5. McGraw-Hill New York, 1967.
- 40 Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979. URL: <http://doi.acm.org/10.1145/359168.359176>, doi:10.1145/359168.359176.

- 41 Larry J Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.
- 42 Milan Studený. Conditional independence relations have no finite complete characterization. In *11th Prague Conf. Information Theory, Statistical Decision Foundation and Random Processes*, pages 377–396. Norwell, MA, 1990.
- 43 Milan Studený. Convex cones in finite-dimensional real vector spaces. *Kybernetika*, 29(2):180–200, 1993.
- 44 Alfred Tarski. A decision method for elementary algebra and geometry. In *Quantifier elimination and cylindrical algebraic decomposition*, pages 24–84. Springer, 1998.
- 45 Raymond W. Yeung. *A first course in information theory*. Information Technology: Transmission, Processing and Storage. Kluwer Academic/Plenum Publishers, New York, 2002. With a foreword by Toby Berger, With 1 CD-ROM. URL: <https://doi-org.gate.lib.buffalo.edu/10.1007/978-1-4419-8608-5>, doi:10.1007/978-1-4419-8608-5.
- 46 Raymond W. Yeung. *Information Theory and Network Coding*. Springer Publishing Company, Incorporated, 1 edition, 2008.
- 47 Raymond W. Yeung and Zhen Zhang. A class of non-Shannon-type information inequalities and their applications. *Commun. Inf. Syst.*, 1(1):87–100, 2001. URL: <https://doi-org.gate.lib.buffalo.edu/10.4310/CIS.2001.v1.n1.a6>, doi:10.4310/CIS.2001.v1.n1.a6.
- 48 Zhen Zhang. On a new non-Shannon type information inequality. *Commun. Inf. Syst.*, 3(1):47–60, 2003. URL: <https://doi-org.gate.lib.buffalo.edu/10.4310/CIS.2003.v3.n1.a4>, doi:10.4310/CIS.2003.v3.n1.a4.
- 49 Zhen Zhang and Raymond W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Information Theory*, 43(6):1982–1986, 1997.
- 50 Zhen Zhang and Raymond W Yeung. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory*, 44(4):1440–1452, 1998.

A

 Background on Cones

We will employ basic facts about closed, convex cones [9, 43], which we review briefly in this section.

Fix a set $S \subseteq \mathbb{R}^m$. We denote by \bar{S} its topological closure. The set S is *convex* if it is closed under taking convex combination, i.e. $\mathbf{x}, \mathbf{y} \in S$ and $\theta \in [0, 1]$ imply $\theta\mathbf{x} + (1 - \theta)\mathbf{y} \in S$. The set S is a (Euclidean) *cone* if it is closed under taking non-negative multiple, i.e. $\mathbf{x} \in S$ and $\theta \geq 0$ imply $\theta\mathbf{x} \in S$. We use $\text{cone}(S)$ to denote its *conic hull*, i.e. the set of conic combinations $\sum_i \theta_i \mathbf{x}_i$ where $\theta_i \geq 0$ and $\mathbf{x}_i \in S$. It is easy to see that conic hulls are convex.

For any set $K \subseteq \mathbb{R}^m$, its *dual cone* K^* is defined by

$$K^* := \{\mathbf{y} \mid \mathbf{x} \cdot \mathbf{y} \geq 0, \forall \mathbf{x} \in K\} \quad (22)$$

If $\mathbf{x} \in \mathbb{R}^m$ then we denote $\mathbf{x}^* = \{\mathbf{y} \mid \mathbf{x} \cdot \mathbf{y} \geq 0\}$ for short.

It is not hard to see that K^* is always a closed, convex cone (regardless of whether K is closed or convex or even a cone). For any two sets $K, L \subseteq \mathbb{R}^m$ it holds that $K \subseteq L^*$ iff $K^* \supseteq L$ (i.e. $(-)^*$ forms an antitone Galois connection). It is also known that, taking duality twice, K^{**} is the closure of the smallest convex cone containing K , i.e. $K^{**} = \overline{\text{cone}(K)}$; in particular, $K^{**} = K$ iff K is a closed convex cone.

A cone K is called *pointed* if $\mathbf{x} \in K$ and $-\mathbf{x} \in K$ imply $\mathbf{x} = \mathbf{0}$; in other words, K is pointed if it contains no line. If K has a non-empty interior, then its dual K^* is pointed. If K is a cone and its closure is pointed, then K^* has a non-empty interior. For any sets $K, L \subseteq \mathbb{R}^m$, it is easy to see that $(K \cup L)^* = K^* \cap L^*$.

B Other Applications of the Boolean Information Constraint Problem

We continue here our list of applications of the Boolean Information Constraint Problem from Sec. 3.2.

B.1 Group-Theoretic Inequalities

There turns out to be a way to “rephrase” IIP as a decision problem in group theory; This was a wonderful result by Chan and Yeung [12] (see also [11]). A tuple $(G; G_1, \dots, G_n)$ is called a *group system* if G is a finite group and $G_1, \dots, G_n \subseteq G$ are n subgroups. For any $\alpha \subseteq [n]$, define $G_\alpha := \bigcap_{i \in \alpha} G_i$; implicitly, we set $G_\emptyset := G$. A vector $\mathbf{c} \subseteq \mathbb{R}^{2^n}$ defines the following *group-theoretic inequality*:

$$\sum_{\alpha \subseteq [n]} c_\alpha \log \frac{|G|}{|G_\alpha|} \geq 0 \quad (23)$$

► **Theorem 18** ([12]). *An information-inequality (2) is Γ_n^* -valid if and only if the corresponding group-theoretic inequality (23) holds for all group systems (G, G_1, \dots, G_n) ,*

In particular, a positive or negative answer to the decidability problem for IIP immediately carries over to the validity problem of group-theoretic inequalities of the form (23). We note that the group-theoretic inequalities considered here are different from the word problems in group, see e.g. the survey [34]; the undecidability results for word problems in groups do not carry over to the group-theoretic inequalities and, thus, to information inequalities.

B.2 Application to Relational Query Evaluation

The problem of bounding the number of copies of a graph inside of another graph has a long and interesting history [17, 4, 14, 35]. The subgraph homomorphism problem is a special case of the relational query evaluation problem, in which case we want to find an upper bound on the output size of a full conjunctive query. Using the entropy argument from [14], *Shearer’s lemma* in particular, Atserias, Grohe, and Marx [5] established a tight upper bound on the answer to a full conjunctive query over a database. Note that Shearer’s lemma is a Shannon-type inequality. Their result was extended to include functional dependencies and more generally degree constraints in a series of recent work in database theory [19, 2, 3]. All these results can be cast as applications of Shannon-type inequalities. For a simple example, let $R(X, Y), S(Y, Z), T(Z, U)$ be three binary relations (tables), each with N tuples, then their join $R(X, Y) \bowtie S(Y, Z) \bowtie T(Z, U)$ can be as large as N^2 tuples. However, if we further know that the functional dependencies $XZ \rightarrow U$ and $YU \rightarrow X$ hold in the output, then one can prove that the output size is $\leq N^{3/2}$, by using the following Shannon-type information inequality:

$$h(XY) + h(YZ) + h(ZU) + h(X|YU) + h(U|XZ) \geq 2h(XYZU) \quad (24)$$

B.3 Application to Secret Sharing

An interesting application of conditional information inequalities is secret sharing, which is a classic problem in cryptography, independently introduced by Shamir [40] and Blakley [7]. The setup is as follows. There is a set P of *participants*, a *dealer* $d \notin P$, and an *access structure* $\mathcal{F} \subseteq 2^P$. The access structure is closed under taking superset: $A \in \mathcal{F}$ and $A \subseteq B$

implies $B \in \mathcal{F}$. The dealer has a secret s , from some finite set K , which she would like to share in such a way that every set $F \in \mathcal{F}$ of participants can recover the secret s , but every set $F \notin \mathcal{F}$ knows *nothing* about s . The dealer shares her secret by using a *secret sharing scheme*, in which she gives each participant $p \in P$ a *share* $s_p \in K_p$, where K_p is some finite domain. The scheme is designed in such a way that from the tuple $(s_p)_{p \in F}$ one can recover s if $F \in \mathcal{F}$, and conversely one cannot infer any information about s if $F \notin \mathcal{F}$.

One way to formalize secret sharing uses information theory (for other formalisms, see [6]). We identify the participants P with the set $[n-1]$, and the dealer with the number n . A secret sharing scheme on P with access structure $\mathcal{F} \subseteq 2^P$ is a joint distribution on n discrete random variables (X_1, \dots, X_n) satisfying:

- (i) $H(X_n) > 0$
- (ii) $H(X_n \mid \mathbf{X}_F) = 0$ if $F \in \mathcal{F}$
- (iii) $H(X_n \mid \mathbf{X}_F) = H(X_n)$ if $F \notin \mathcal{F}$; equivalently, $I_H(X_n; \mathbf{X}_F) = 0$.

Intuitively, X_i denotes the share given to the i th participant, and X_n is the unknown secret. It can be shown, without loss of generality, that (i) can be replaced by the assumption that the marginal distribution on X_n is uniform [8], which encodes the fact that the scheme does not reveal any information about the secret X_n . Condition (ii) means one can recover the secret from the shares of qualified participants, while condition (iii) guarantees the complete opposite. A key challenge in designing a good secret sharing scheme is to reduce the total size of the shares. The only known [15, 10, 25] way to prove a *lower bound* on share sizes is to lower bound the *information ratio* $\frac{\max_{p \in P} H(X_p)}{H(X_n)}$. In order to prove that some number ℓ is a lower bound on the information ratio we need to check that $\max_{i \in [n-1]} \{h(X_i) - \ell \cdot h(X_n)\} \geq 0$ holds for all entropic functions $\mathbf{h} \in \Gamma_n^*$ satisfying the extra conditions (i), (ii), and (iii) above. Equivalently, ℓ is a lower bound on the information ratio if and only if the following Boolean information constraint is Γ_n^* -valid:

$$\bigwedge_{F \in \mathcal{F}} (h(X_n \mid \mathbf{X}_F) = 0) \wedge \bigwedge_{F \notin \mathcal{F}} (I_h(X_n; \mathbf{X}_F) = 0) \implies (h(X_n) = 0) \vee \left[\bigvee_{i \in [n-1]} (h(X_i) \geq \ell \cdot h(X_n)) \right]$$

C Proof of Theorem 7

Proof. (of Theorem 7) Tarski has proven that the theory of reals with $+, *$ is decidable. More precisely, given a formula Φ in FO with symbols $+$ and $*$ it is decidable whether that formula is true in the model of real numbers $(\mathbb{R}, +, *)$; for example, it is decidable whether⁵ $\Phi \equiv \forall x \exists y \forall z (x^2 + 3y \geq z \wedge (y^3 + yz \leq xy^2))$ is true. We will write $(\mathbb{R}, +, *) \models \Phi$ to denote the fact that Φ is true in the model of reals.

Consider a conditional inequality:

$$I_h(Y_1; Z_1 \mid X_1) = 0 \wedge \dots \wedge I_h(Y_k; Z_k \mid X_k) = 0 \implies I_h(Y; Z \mid X) = 0$$

The following algorithm returns *false* if the inequality fails on some entropic function h , and runs forever if the inequality holds for all h , proving that the problem is in Π_1^0 :

- Iterate over all $N \geq 0$. For each N do the following steps.
- Consider n joint random variables X_1, \dots, X_n where each has outcomes in the domain $[N]$; thus there are N^n possible outcomes.

⁵ $3y$ is a shorthand for $y + y + y$ and $x \geq y$ is a shorthand for $\exists u (x = y + u^2)$.

- Construct a formula Δ stating “there exist probabilities p_1, \dots, p_{N^n} for these outcomes, whose entropy fails the conditional inequality”. More precisely, the formula consists of the following:
 - Convert each conditional independence statement in the antecedent $I_h(Y_i; Z_i | X_i) = 0$ into its equivalent statement on probabilities: $p(X_i Y_i Z_i) p(X_i) = p(X_i Y_i) p(X_i Z_i)$.
 - Replace each such statement with a conjunction of statements of the form $p(X_i = x, Y_i = y, Z_i = z) p(X_i = x) = p(X_i = x, Y_i = y) \cdot p(X_i = x, Z_i = z)$, for all combinations of values x, y, z .
 - Each marginal probability is a sum of atomic probabilities, for example $p(X_i = x, Y_i = y) = p_{k_1} + p_{k_2} + \dots$ where p_{k_1}, p_{k_2}, \dots are the probabilities of all outcomes that have $X_i = x$ and $Y_i = y$. Thus, the identity from the previous step becomes the following formula: $(p_{i_1} + p_{i_2} + \dots)(p_{j_1} + p_{j_2} + \dots) = (p_{k_1} + p_{k_2} + \dots)(p_{\ell_1} + p_{\ell_2} + \dots)$. There is one such formula for every combination of values x, y, z ; denote Φ_i the conjunction of all these formulas. Thus, Φ_i asserts $I_h(Y_i; Z_i | X_i) = 0$.
 - Let $\Phi = \Phi_1 \wedge \dots \wedge \Phi_k$. Let Ψ be the similar formula for the consequent: thus, Ψ asserts $I_h(Y; Z | X) = 0$.
 - Finally, construct the formula $\Delta \stackrel{\text{def}}{=} \forall p_1, \dots, \forall p_{N^n}, (\Phi \wedge \neg \Psi)$.
- Check whether $(\mathbb{R}, +, *) \models \Delta$. By Tarski’s theorem this step is decidable.
- Return *false* if the formula is false. Otherwise, continue with $N + 1$.

◀

D Proof of Theorem 8

We need the following non-obvious lemma due to Studený:

► **Lemma 19.** [43, Lemma 1] Let $L \subseteq \mathbb{R}^m$ be a closed, convex cone, and $\mathbf{y} \in \mathbb{R}^m$ be a vector s.t. $\mathbf{y} \notin L$. Then $\text{cone}(L \cup \{\mathbf{y}\})$ is closed.

► **Example 20.** The condition $\mathbf{y} \notin L$ is necessary, as illustrated by the following example, also from [43]. Let

$$L = \{(a, b, c) \mid a, b, c \in \mathbb{R}, a \geq 0, c \geq 0, ac \geq b^2\} \quad (25)$$

One can check that this is a closed, convex cone⁶. Notice that $c = 0$ implies $b = 0$. Let $\mathbf{y} = (-1, 0, 0)$, then $\text{cone}(L \cup \{\mathbf{y}\}) = \{(a, b, c) \mid a, b, c \in \mathbb{R}, c \geq 0 \wedge (c = 0 \Rightarrow b = 0)\}$. Then the sequence of vectors $\mathbf{z}_n = (0, 1, 1/n)$ is in $\text{cone}(L \cup \{\mathbf{y}\})$, but their limit $(0, 1, 0)$ is not, proving that $\text{cone}(L \cup \{\mathbf{y}\})$ is not closed.

Fix a cone $K \subseteq \mathbb{R}^m$. We say that a set of vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_k\} \subseteq \mathbb{R}^m$ has *slack* w.r.t. K if there exists $\mathbf{x} \in K$ such that $\mathbf{y}_i \cdot \mathbf{x} > 0$ for $i = 1, k$. We prove the following lemma, which is an extension of a result in [26]:

► **Lemma 21.** Let $K \subseteq \mathbb{R}^m$ be a closed convex cone. Then the following statements are equivalent:

$$\forall \mathbf{x} \in K : \bigwedge_{i \in [k]} (\mathbf{x} \cdot \mathbf{y}_i \geq 0) \Rightarrow (\mathbf{x} \cdot \mathbf{y} \geq 0) \quad (26)$$

$$\forall \varepsilon > 0, \exists \lambda \geq 0, \forall \mathbf{x} \in K : \bigwedge_{i \in [k-1]} (\mathbf{x} \cdot \mathbf{y}_i \geq 0) \Rightarrow (\mathbf{x} \cdot \mathbf{y} + \varepsilon \|\mathbf{x}\|_2 \geq \lambda \mathbf{x} \cdot \mathbf{y}_k) \quad (27)$$

Moreover, if $\{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ has slack w.r.t. K then we can set $\varepsilon = 0$ in Eq.(27).

⁶ L is isomorphic to the cone S^2 of positive semi-definite symmetric 2×2 matrices.

Proof. The implication (27) \Rightarrow (26) is immediate, hence we prove (26) \Rightarrow (27). For every i , the statement $\mathbf{x} \cdot \mathbf{y}_i \geq 0$ is equivalent to $\mathbf{x} \in \mathbf{y}_i^*$. Denote $L \stackrel{\text{def}}{=} K \cap \bigcap_{i \in [k-1]} \mathbf{y}_i^*$ and notice that this is a closed, convex cone. We start by proving that Condition (26) is equivalent to $\mathbf{y} \in \overline{\text{cone}(L^* \cup \{\mathbf{y}_k\})}$:

$$\begin{aligned}
\text{Condition (26)} &\Leftrightarrow L \cap \mathbf{y}_k^* \subseteq \mathbf{y}^* \\
&\Leftrightarrow L^{**} \cap \mathbf{y}_k^* \subseteq \mathbf{y}^* && L \text{ is closed, convex, hence } L^{**} = L \\
&\Leftrightarrow (L^* \cup \{\mathbf{y}_k\})^* \subseteq \mathbf{y}^* && A^* \cap B^* = (A \cup B)^* \\
&\Leftrightarrow \mathbf{y} \in (L^* \cup \{\mathbf{y}_k\})^{**} && A \subseteq B^* \text{ iff } A^* \supseteq B \\
&\Leftrightarrow \mathbf{y} \in \overline{\text{cone}(L^* \cup \mathbf{y}_k)} && A^{**} = \overline{\text{cone}(A)}
\end{aligned}$$

Consider first the case when the set $\{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ has slack in K , and let $\mathbf{x}_0 \in K$ be such that $\mathbf{x}_0 \cdot \mathbf{y}_i > 0$ for all $i = 1, k$; in particular, $\mathbf{x}_0 \in L$ and $\mathbf{x}_0 \cdot \mathbf{y}_k > 0$. This implies that $\mathbf{y}_k \notin -L^*$, hence, by Lemma 19, $\text{cone}(L^* \cup \{\mathbf{y}_k\})$ is closed. Therefore we have that $\mathbf{y} \in \text{cone}(L^* \cup \{\mathbf{y}_k\})$, hence there exists $\mathbf{z} \in L^*$ and $\lambda \geq 0$ such that $\mathbf{y} = \mathbf{z} + \lambda \mathbf{y}_k$. To prove condition (27), it suffices to show that $\forall \mathbf{x} \in L$, $\mathbf{x} \cdot \mathbf{y} \geq \lambda \mathbf{x} \cdot \mathbf{y}_k$. This follows from the fact that $\mathbf{x} \cdot \mathbf{z} \geq 0$, which implies that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{z} + \lambda \mathbf{x} \cdot \mathbf{y}_k \geq \lambda \mathbf{x} \cdot \mathbf{y}_k$ as required.

Consider now the general case, when $\mathbf{y} \in \overline{\text{cone}(L^* \cup \{\mathbf{y}_k\})}$. Then, for every $\varepsilon > 0$ there exists $\mathbf{y}' \in \text{cone}(L^* \cup \{\mathbf{y}_k\})$ such that, denoting $\boldsymbol{\delta} \stackrel{\text{def}}{=} \mathbf{y}' - \mathbf{y}$, we have $\|\boldsymbol{\delta}\|_2 < \varepsilon$. Applying the argument above to \mathbf{y}' instead of \mathbf{y} , we obtain $\mathbf{x} \cdot \mathbf{y}' \geq \lambda \mathbf{x} \cdot \mathbf{y}_k$. On the other hand, $\mathbf{x} \cdot \mathbf{y}' = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \boldsymbol{\delta} \leq \mathbf{x} \cdot \mathbf{y} + \|\mathbf{x}\|_2 \cdot \|\boldsymbol{\delta}\|_2 \leq \mathbf{x} \cdot \mathbf{y} + \varepsilon \|\mathbf{x}\|_2$. \blacktriangleleft

► **Example 22.** We show that the error term $\varepsilon \|\mathbf{x}\|_2$ in Condition (27) is necessary in general. For that, consider the cone L in Eq. (25). It satisfies the condition $\forall (a, b, c) \in L : a \leq 0 \Rightarrow b \leq 0$. Indeed, $a \leq 0$ is equivalent to $a = 0$, thus $b^2 \leq ac = 0$ implying $b = 0$, in particular $b \leq 0$. By writing the implication as $-a \geq 0 \Rightarrow -b \geq 0$, Lemma 21 says that $\forall \varepsilon > 0, \exists \lambda \geq 0$ such that $\forall (a, b, c) \in L : -b + \varepsilon \|(a, b, c)\|_2 \geq \lambda(-a)$ or, equivalently $b \leq \varepsilon \|(a, b, c)\|_2 + \lambda a$. When $\varepsilon = 0$ then this condition fails for any choice of $\lambda \geq 0$, for example it fails on the vector $(1, 1 + \lambda, (1 + \lambda)^2) \in L$. On the other hand, if $\varepsilon > 0$, then the condition holds, for example we can choose $\lambda = 1/\varepsilon$ and we obtain $\varepsilon \|(a, b, c)\|_2 + \lambda a \geq \varepsilon c + \lambda a \geq 2\sqrt{ac} \geq 2|b| \geq b$.

We can now prove Theorem 8.

Proof. (of Theorem 8) It remains to prove that (12) implies (13). This follows from Lemma 21 applied to $K \stackrel{\text{def}}{=} \overline{\Gamma_n^*}$, which is a closed, convex cone, see [45]. By renaming the vectors $\mathbf{x}, \mathbf{y}_i, \mathbf{y}$ in Lemma 21 to $\mathbf{h}, \mathbf{c}_i, \mathbf{c}$ in Theorem 8, and using $\varepsilon/2^n$ instead of ε , we use the Lemma to argue that (12) implies:

$$\forall \varepsilon > 0, \exists \lambda > 0, \forall \mathbf{h} \in \overline{\Gamma_n^*} : \bigwedge_{i \in [k-1]} \mathbf{c}_i \cdot \mathbf{h} \geq 0 \Rightarrow \mathbf{c} \cdot \mathbf{h} + \frac{1}{2^n} \varepsilon \|\mathbf{h}\|_2 \geq \lambda \mathbf{c}_k \cdot \mathbf{h}$$

Condition (13) follows from the fact that $h([n]) \geq \frac{1}{2^n} \sum_{\alpha \subseteq [n]} h(\alpha) = \frac{1}{2^n} \|\mathbf{h}\|_1 \geq \frac{1}{2^n} \|\mathbf{h}\|_2$. \blacktriangleleft

We end this section by stating an obvious consequence of Lemma 21: by applying it repeatedly, we can move all antecedents to the consequent, and obtain:

► **Corollary 23.** Let $K \subseteq \mathbb{R}^m$ be a closed convex cone, and assume that $\{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ has slack w.r.t. K . Then the following are equivalent:

$$\forall \mathbf{x} \in K : \bigwedge_{i \in [k]} (\mathbf{x} \cdot \mathbf{y}_i \geq 0) \Rightarrow (\mathbf{x} \cdot \mathbf{y} \geq 0) \quad (28)$$

$$\exists \lambda_1, \dots, \lambda_k \geq 0, \forall \mathbf{x} \in K : (\mathbf{x} \cdot \mathbf{y} \geq \sum_{i \in [k]} \lambda_i \mathbf{x} \cdot \mathbf{y}_i) \quad (29)$$

E

 Proof of Theorem 15

We prove the theorem by generalizing it to arbitrary cones, in Th. 25 below. Its proof, in turn, is based on the following lemma.

► **Lemma 24.** *If $K \subseteq \mathbb{R}^m$ is a convex cone such that $K \cap (-\infty, 0)^m = \emptyset$, then $K^* \cap \mathbb{R}_+^m \neq \{0\}$.*

In other words, if a convex cone K satisfies following property

$$\forall \mathbf{x} \in K, \quad \bigvee_{j \in [m]} x_j \geq 0 \quad (30)$$

then there exists \mathbf{y} s.t. $y_j \geq 0$ for all $j \in [m]$, and $\mathbf{x} \cdot \mathbf{y} \geq 0$ for all $\mathbf{x} \in K$.

Proof. Let $\mathbf{e}_1, \dots, \mathbf{e}_m$ be the canonical basis of \mathbb{R}^m , i.e. $(\mathbf{e}_j)_i = \delta_{ij}$, and let $L = \text{cone}(K \cup \{\mathbf{e}_1, \dots, \mathbf{e}_m\})$. We claim that L also satisfies (30). Indeed, every $\mathbf{x}' \in L$ has the form $\mathbf{x}' = \mathbf{x} + \sum_i \theta_i \mathbf{e}_i$ with $\mathbf{x} \in K$ and $\theta_i \geq 0$, $i = 1, m$. If $x'_j < 0$ for all j , then $x_j < 0$ for all j , which is a contradiction because K satisfies property (30). Thus, L is a convex cone, and disjoint from the strictly negative quadrant $(-\infty, 0)^m$; since the latter is an open set, it is also disjoint from \bar{L} . We claim that $L^* \neq \{0\}$. Indeed, otherwise $L^{**} = \{0\}^* = \mathbb{R}^m$, but $L^{**} = \bar{L}$ is disjoint from $(-\infty, 0)^m$, which is a contradiction. Therefore there exists $\mathbf{y} \in L^*$ s.t. $\mathbf{y} \neq 0$. Since $\mathbf{e}_i \in L$, we have $y_i = \mathbf{y} \cdot \mathbf{e}_i \geq 0$, i.e. $\mathbf{y} \in \mathbb{R}_+^m$. Since $K \subseteq L$ we have $L^* \subseteq K^*$, hence $\mathbf{y} \in K^*$ proving the lemma. ◀

We prove Theorem 15 by generalizing it to arbitrary convex cones.

► **Theorem 25.** *Let $S \subseteq \mathbb{R}^m$ be any convex cone, and $\mathbf{y}_1, \dots, \mathbf{y}_k \in \mathbb{R}^m$. Then the following two properties are equivalent:*

$$\forall \mathbf{x} \in S : \max(\mathbf{x} \cdot \mathbf{y}_1, \dots, \mathbf{x} \cdot \mathbf{y}_k) \geq 0 \quad (31)$$

$$\exists \lambda_1 \geq 0, \dots, \lambda_k \geq 0, \sum_i \lambda_i = 1, \forall \mathbf{x} \in S : \sum_{i \in [k]} \lambda_i \mathbf{x} \cdot \mathbf{y}_i \geq 0 \quad (32)$$

Theorem 15 is the special case when $S \stackrel{\text{def}}{=} \bar{\Gamma}_n^* \cap \bigcap_{i \in [k]} \mathcal{C}_i^*$.

Proof. (of Theorem 25). We prove only the implication from (31) to (32); the other direction is immediate. Define:

$$K \stackrel{\text{def}}{=} \{(\mathbf{x} \cdot \mathbf{y}_1, \dots, \mathbf{x} \cdot \mathbf{y}_k) \mid \mathbf{x} \in S\}$$

Since S is a convex cone, it follows that K is also a convex cone, hence we can apply Lemma 24 to K . Condition (20) of the theorem states:

$$\forall \mathbf{z} \in K : \bigvee_{j \in [m]} z_j \geq 0$$

Lemma 24 implies that there exists $\boldsymbol{\lambda} \in \mathbb{R}_+^k \cap K^*$ s.t. $\boldsymbol{\lambda} \neq 0$:

$$\forall \mathbf{z} \in K : \boldsymbol{\lambda} \cdot \mathbf{z} \geq 0$$

Expanding the definition of K , this condition becomes:

$$\forall \mathbf{x} \in S : \sum_{i \in [k]} \lambda_i \mathbf{x} \cdot \mathbf{y}_i \geq 0 \quad (33)$$

We can assume w.l.o.g. that $\sum_i \lambda_i = 1$ (otherwise we normalize it by dividing by $\sum_i \lambda_i$), and this proves (32). ◀

F

 Proof of Theorem 12

We prove Theorem 12 by generalizing it to arbitrary cones. First, we give the obvious generalization of the notion of group balanced. Fix a set of vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{Q}^m$ with rational coordinates.

► **Definition 26.** A set $D = \{\mathbf{y}_1, \dots, \mathbf{y}_k\} \subseteq \mathbb{R}^m$ is called group balanced if (a) $\text{rank } \mathbf{A} = k - 1$ where \mathbf{A} is the matrix $A_{ij} = \mathbf{x}_j \cdot \mathbf{y}_i$, and (b) there exists $\mathbf{x}^{(*)} \in \text{conv}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ such that $\mathbf{x}^{(*)} \cdot \mathbf{y}_i = 0$ for all i .

► **Theorem 27.** Let $K \subseteq \mathbb{R}^m$ be a cone and $D = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \subseteq \mathbb{Q}^m$ be a group balanced set of rational vectors. Suppose that the following condition holds:

$$\exists \lambda_1 \geq 0, \dots, \lambda_k \geq 0, \sum_i \lambda_i = 1, \forall \mathbf{x} \in K : \quad \sum_i \lambda_i \mathbf{x} \cdot \mathbf{y}_i \geq 0 \quad (34)$$

Then there exists rational λ_i 's with this property.

Proof. Denote by $\Lambda \stackrel{\text{def}}{=} \{(\lambda_1, \dots, \lambda_n) \mid \lambda_1 \geq 0, \dots, \lambda_n \geq 0, \forall \mathbf{x} \in K, \sum_i \lambda_i \mathbf{x} \cdot \mathbf{y}_i \geq 0\}$. Then Λ is a convex cone, and is $\neq \{\mathbf{0}\}$ by condition (34). To prove the theorem, we will show $\Lambda = \{t\boldsymbol{\lambda} \mid t \geq 0\}$ for some rational vector $\boldsymbol{\lambda} \in \mathbb{Q}_+^n$. Denote the matrix $\mathbf{A} \stackrel{\text{def}}{=} (\mathbf{y}_i \cdot \mathbf{x}_j)_{ij}$; by assumption its rank is $n - 1$. Def. 26 (b) implies that there exists $\mu_1 \geq 0, \dots, \mu_n \geq 0$ such that, denoting $\mathbf{x}^{(*)} \stackrel{\text{def}}{=} \sum_{j \in [n]} \mu_j \mathbf{x}_j$, it holds that $\mathbf{x}^{(*)} \cdot \mathbf{y}_i = 0$ for all i . Since $\text{rank } \mathbf{A} = n - 1$, we have $\mu_j > 0$ for all j . Consider now any $(\lambda_1, \dots, \lambda_n) \in \Lambda$, and let $\mathbf{y}^{(*)} = \sum_{i \in [n]} \lambda_i \mathbf{y}_i$. We prove that, for every $j \in [n]$, $\mathbf{x}_j \cdot \mathbf{y}^{(*)} = 0$. To prove this, we note that $\mathbf{x}^{(*)} \cdot \mathbf{y}_i = 0$ and $\mathbf{x}_j \cdot \mathbf{y}^{(*)} \geq 0$ (condition (34) applied to \mathbf{x}_j) imply:

$$0 = \sum_{i \in [n]} \lambda_i \mathbf{x}^{(*)} \cdot \mathbf{y}_i = \sum_{i, j \in [n]} \lambda_i \mu_j \mathbf{x}_j \cdot \mathbf{y}_i = \sum_{j \in [n]} \mu_j (\mathbf{x}_j \cdot \mathbf{y}^{(*)}) \geq 0$$

If $\mathbf{x}_j \cdot \mathbf{y}^{(*)} > 0$ for some j then we obtain $0 > 0$, a contradiction, hence $\mathbf{x}_j \cdot \mathbf{y}^{(*)} = 0$ for all j . Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the column vectors of \mathbf{A} , and let $V \stackrel{\text{def}}{=} \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$. We have proven that $\Lambda \subseteq V^\perp$, where V^\perp denotes the orthogonal space⁷. Since $\dim(V) = n - 1$, we have that $\dim(V^\perp) = 1$, in other words $V^\perp = \{t\boldsymbol{\lambda} \mid t \in \mathbb{R}\}$ for some non-zero vector $\boldsymbol{\lambda}$. Moreover, $\boldsymbol{\lambda}$ can be chosen to be a rational vectors, because $\mathbf{v}_1, \dots, \mathbf{v}_n$ have rational coordinates. This proves the claim and the theorem. ◀

Next we give an example showing that some additional condition on the vectors $\mathbf{y}_1, \dots, \mathbf{y}_k$ is necessary to ensure that the values λ_i 's can be chosen rational numbers.

► **Example 28.** We show here an example where the values λ_i cannot be chosen to be rational numbers. We generalize Example 20 as follows. Fix two numbers $\alpha, \gamma \in (0, 1)$ such that $\alpha + \gamma = 1$ and $\frac{\alpha}{\gamma} \notin \mathbb{Q}$. Consider the cone:

$$K \stackrel{\text{def}}{=} \{(a, b, c) \mid a \geq 0, c \geq 0, a^\alpha c^\gamma \geq b\} \quad (35)$$

(The reader may verify that K is a closed, convex cone.) We claim that, for all $(a, b, c) \in K$, either $a \geq b$ or $c \geq b$, in other words K satisfies the max-inequality:

$$\forall (a, b, c) \in K : \quad \max(a - b, c - b) \geq 0$$

⁷ For any set $V \subseteq \mathbb{R}^k$, its orthogonal space is $V^\perp \stackrel{\text{def}}{=} \{\mathbf{u} \mid \forall \mathbf{v} \in V, \mathbf{u} \cdot \mathbf{v} = 0\}$.

This follows from the inequality between the weighted arithmetic mean and weighted geometric mean: $\alpha a + \gamma c \geq a^\alpha c^\gamma \geq b = (\alpha + \gamma)b$, hence, if both $a < b$ and $c < b$ then $\alpha a + \gamma c < (\alpha + \gamma)b = b$ contradiction. By Theorem 25 there exists $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$ such that

$$\forall (a, b, c) \in K : \quad \lambda_1(a - b) + \lambda_2(c - b) \geq 0$$

Equivalently, $\forall (a, b, c) \in K$, $\lambda_1 a + \lambda_2 c \geq b$. We prove that the only possible values are $\lambda_1 = \alpha, \lambda_2 = \gamma$, and thus no rational values exists. For that we set $b = a^\alpha c^{1-\alpha}$ in the inequality above and obtain:

$$\begin{aligned} \forall a, c \geq 0 : \quad & \lambda_1 a + (1 - \lambda_1)c \geq a^\alpha c^{1-\alpha} \\ \forall a \geq 0, c > 0 : \quad & \lambda_1 \frac{a}{c} + 1 - \lambda_1 \geq \left(\frac{a}{c}\right)^\alpha \\ \forall x \geq 0 : \quad & f(x) \stackrel{\text{def}}{=} (1 - x^\alpha) - \lambda_1(1 - x) \geq 0 \end{aligned}$$

Since $f(1) = 0$ and $f(x) \geq 0$ for $x \in \mathbb{R}$, we must have $f'(1) = 0$ by Lagrange's theorem. Thus, $-\alpha + \lambda_1 = 0$ or $\lambda_1 = \alpha$, proving the claim.

Finally, we prove that the definition of strong balance is natural. Recall Condition (18) of Theorem 12, which we repeat here for readability:

$$\exists \lambda_1 \geq 0, \dots, \lambda_n \geq 0, \sum_{i \in [n]} \lambda_i = 1, \forall \mathbf{h} \in \bar{\Gamma}_n^* : \sum_i \lambda_i \mathbf{d}_i \cdot \mathbf{h} \geq 0 \quad (36)$$

► **Proposition 29.** *For any set of vectors $D = \{\mathbf{d}_1, \dots, \mathbf{d}_n\} \subseteq \mathbb{R}^{2^n}$ there exists a strongly balanced set $D' = \{\mathbf{d}'_1, \dots, \mathbf{d}'_n\} \subseteq \mathbb{R}^{2^n}$ such that D satisfies Condition (36) iff D' satisfies it.*

Proof. (Of Prop. 29) Fix any set $D = \{\mathbf{d}_1, \dots, \mathbf{d}_n\} \in \mathbb{R}^{2^n}$, and denote by \mathbf{A} the $n \times n$ matrix $A_{ij} = \mathbf{d}_i \cdot \mathbf{h}^{(j)}$. We will modify D to ensure that the matrix \mathbf{A} satisfies Def. 11 First, we replace each \mathbf{d}_i by \mathbf{d}'_i obtained using Chan's transformation:

$$\mathbf{d}'_i \cdot \mathbf{h} = \mathbf{d}_i \cdot \mathbf{h} - \sum_j (\mathbf{d}_i \cdot \mathbf{h}^{(j)}) h(X_j | X_{[n]-\{j\}})$$

and denote by $D' \stackrel{\text{def}}{=} \{\mathbf{d}'_1, \dots, \mathbf{d}'_n\}$. First, we claim that D satisfies (36) iff D' does. This follows from Chan's theorem, since, for any values $\lambda_1, \dots, \lambda_n$ that sum to 1, the expression $\mathbf{d}' \stackrel{\text{def}}{=} \sum_i \lambda_i \mathbf{d}'_i$ is precisely Chan's transformation applied to $\mathbf{d} \stackrel{\text{def}}{=} \sum_i \lambda_i \mathbf{d}_i$, hence $\mathbf{d}' \cdot \mathbf{h} \geq 0$ is valid iff $\mathbf{d} \cdot \mathbf{h} \geq 0$ is valid. After this transformation, every \mathbf{d}'_i is balanced, and therefore the matrix \mathbf{A}' associated to the new set D' is identically 0. Next, assume that D' satisfies condition (36), and let $\lambda_1, \dots, \lambda_n$ be the corresponding coefficients. We have $\lambda_i > 0$ for all i because the rank of \mathbf{A} is $n - 1$. Replace each \mathbf{d}'_i by \mathbf{d}''_i where:

$$\mathbf{d}''_i \cdot \mathbf{h} = \mathbf{d}'_i \cdot \mathbf{h} + \frac{1}{\lambda_i} \left(n h(X_i | X_{[n]-\{i\}}) - \sum_{j \in [n]} h(X_j | X_{[n]-\{j\}}) \right)$$

In other words, we add the term $\frac{n-1}{\lambda_i} h(X_i | X_{[n]-\{i\}})$ and subtract all terms $\frac{1}{\lambda_i} h(X_j | X_{[n]-\{j\}})$ for $j \neq i$. This transformation does not affect the expression (36) because $\sum_i \lambda_i \mathbf{d}''_i = \sum_i \lambda_i \mathbf{d}'_i$. The new $n \times n$ matrix \mathbf{A}'' is as follows. The diagonal is $A_{ii} = (n - 1)/\lambda_i$, and all other entries are $A_{ij} = -1/\lambda_i$. To compute its rank, multiply each row i with λ_i , and the new matrix has $n - 1$ on the diagonal and 1 everywhere else, hence the rank is $n - 1$. Finally, we check that there exists $\mathbf{h}^{(*)}$ such that $\mathbf{d}''_i \cdot \mathbf{h}^{(*)} = 0$ for all i , by setting $\mathbf{h}^{(*)} = \sum_j \mathbf{h}^{(j)}$; qthen $\mathbf{d}''_i \cdot \mathbf{h}^{(*)}$ is precisely the sum of the elements in row i of the matrix \mathbf{A}'' , hence it is 0. ◀

G

 The Recognizability Problems

We study here two problems that are the dual of the Boolean information constraint problem. The *entropic-recognizability problem* takes as input a vector \mathbf{h} and checks if $\mathbf{h} \in \Gamma_n^*$. The *almost-entropic-recognizability problem* checks if $\mathbf{h} \in \bar{\Gamma}_n^*$. We will prove that the latter is in Π_2^0 , and leave open the complexity of the former.

Before we define these problems formally, we first must address the question of how to represent the input \mathbf{h} . One possibility is to represent \mathbf{h} as a vector of rational numbers, but this is unsatisfactory, because usually entropies are not rational numbers. Instead, we will allow a more general representation. To justify it, assume first that \mathbf{h} were given by some representable space Ω (Sec. 4.1), where all probabilities are rational numbers. In that case, every term $p_i \log p_i$ in the definition of the entropy can be written as $\log(p_i^{p_i})$, hence then quantity $h(\mathbf{X})$ has the form $h(\mathbf{X}) = \log \prod_i p_i^{p_i}$. In general, any product $\prod_i m_i^{n_i}$ where $m_i, n_i \in \mathbb{Q}$, for $i = 1, n$, can be rewritten as $\left(\frac{a}{b}\right)^{\frac{1}{c}}$, where $a, b, c \in \mathbb{N}$. Indeed, writing $m_i = u_i/v_i$ and $n_i = s_i/t_i$ where $u_i, v_i, s_i, t_i \in \mathbb{N}$, we have:

$$\prod_i \left(\frac{u_i}{v_i}\right)^{\frac{s_i}{t_i}} = \prod_i \left(\frac{u_i^{s_i}}{v_i^{s_i}}\right)^{\frac{1}{t_i}} = \left(\prod_i \frac{u_i^{s_i \cdot \prod_{j \neq i} t_j}}{v_i^{s_i \cdot \prod_{j \neq i} t_j}}\right)^{\frac{1}{\prod_i t_i}} = \left(\frac{a}{b}\right)^{\frac{1}{c}} \quad a, b, c \in \mathbb{N}$$

Justified by this observation, we assume that the input to our problem consists of three vectors $(a_{\mathbf{X}})_{\mathbf{X} \subseteq \mathbf{V}}$, $(b_{\mathbf{X}})_{\mathbf{X} \subseteq \mathbf{V}}$, and $(c_{\mathbf{X}})_{\mathbf{X} \subseteq \mathbf{V}}$ in \mathbb{N}^{2^n} , with the convention that $h(\mathbf{X}) \stackrel{\text{def}}{=} \frac{1}{c_{\mathbf{X}}} \log \frac{a_{\mathbf{X}}}{b_{\mathbf{X}}}$. Thus, we do not assume that these vectors come from a representable space Ω , we only assume their entropies can be represented in this form.

► **Definition 30** ((Almost-)Entropic Recognizability Problem). *Given natural numbers $(a_{\mathbf{X}})_{\mathbf{X} \subseteq \mathbf{V}}$, $(b_{\mathbf{X}})_{\mathbf{X} \subseteq \mathbf{V}}$ and $(c_{\mathbf{X}})_{\mathbf{X} \subseteq \mathbf{V}}$, check whether the vector $h(\mathbf{X}) \stackrel{\text{def}}{=} \frac{1}{c_{\mathbf{X}}} \log \frac{a_{\mathbf{X}}}{b_{\mathbf{X}}}$, $\mathbf{X} \subseteq \mathbf{V}$, represents an entropic vector, or an almost-entropic vector.*

Our result in this section is:

► **Theorem 31.** *The almost entropic recognizability problem is in Π_2^0 .*

To prove the theorem, we need to establish the follows (separating-hyperplane-type) lemma:

► **Lemma 32.** *Let $\mathbf{h} \in \mathbb{R}^{2^n}$ and suppose $\mathbf{h} \notin \bar{\Gamma}_n^*$. Then there exists an information theoretic inequality with integral coefficients that is not satisfied by \mathbf{h} . In other words, $\exists \mathbf{c} \in \mathbb{Z}^{2^n}$, such that $\forall \mathbf{h}_0 \in \Gamma_n^*$, $\mathbf{c} \cdot \mathbf{h}_0 \geq 0$, and $\mathbf{c} \cdot \mathbf{h} < 0$.*

To prove the lemma, we review some background, following Studený [43].

► **Definition 33.** *Given a cone L , define its plane as $pl(L) \stackrel{\text{def}}{=} L \cap (-L)$.*

► **Fact 34.** [43, Fact 9] *If L is a closed, convex cone, then $pl(L)$ is a vector space.*

► **Definition 35.** [43, Def. 6] *A closed convex cone $L \subseteq \mathbb{R}^m$ is called regular if \mathbb{Q} is dense in $pl(L)$, i.e. $\overline{\mathbb{Q}^m \cap pl(L)} = pl(L)$.*

► **Lemma 36.** [43, Prop. 3] *If L is regular, then \mathbb{Q}^m is dense in L^* .*

These lemmas allows us to prove the separating-plane lemma:

Proof. (of Lemma 32) We use Lemma 36. Every pointed cone L is regular, because $pl(L) = \{0\}$. The cone of almost entropic functions $K = \bar{\Gamma}_n^*$ is pointed, because for every $\mathbf{h} \in K$ and every $X \subseteq [n]$, $h(X) \geq 0$; thus, if $\mathbf{h}_1, \mathbf{h}_2 \in \bar{\Gamma}_n^*$ and $\mathbf{h}_1 + \mathbf{h}_2 = 0$, then $\mathbf{h}_1 = \mathbf{h}_2 = 0$. Therefore K is regular, hence \mathbb{Q}^{2^n} is dense in K^* . Consider a vector that is not almost-entropic, $\mathbf{h} \notin K$. Then there exists an information inequality $c_0 \in K^*$ s.t. $c_0 \cdot \mathbf{h} < 0$. Since \mathbb{Q} is dense in K^* , there exists a sequence in $\mathbb{Q}^{2^n} \cap K^*$ that converges to c , and therefore there exists some $c_1 \in \mathbb{Q}^{2^n} \cap K^*$ s.t. $c_1 \cdot \mathbf{h} < 0$. Multiply c_1 with the product of all 2^n denominators of its components, and obtain a vector $c \in \mathbb{Z}^{2^n} \cap K^*$ s.t. $c \cdot \mathbf{h} < 0$. This completes the proof. \blacktriangleleft

And, finally, we use it to place the almost-entropic-recognizability problem in Π_2^0 :

Proof. (of Theorem 31) Given $\mathbf{c} \in \mathbb{Z}^{2^n}$, let $P(\mathbf{c})$ be the following predicate:

$$P(\mathbf{c}) : \quad \forall \mathbf{h} \in \mathbb{R}^{2^n} (\mathbf{h} \in \bar{\Gamma}_n^* \Rightarrow \mathbf{c} \cdot \mathbf{h} \geq 0)$$

Thus, $P(\mathbf{c})$ checks if \mathbf{c} defines a valid information inequality, and this, by Theorem 6, is in Π_1^0 . By Lemma 32 the almost-entropic recognizability problem as follows. Given $\mathbf{h} \in \mathbb{R}^{2^n}$ (represented as in Def. 30):

$$\mathbf{h} \in \bar{\Gamma}_n^* \Leftrightarrow \forall \mathbf{c} (P(\mathbf{c}) \rightarrow \mathbf{c} \cdot \mathbf{h} \geq 0) \Leftrightarrow \forall \mathbf{c} (\neg P(\mathbf{c}) \vee \mathbf{c} \cdot \mathbf{h} \geq 0)$$

which places the problem in Π_2^0 because $\neg P(\mathbf{c})$ is in Σ_1^0 . \blacktriangleleft

We end with a brief comment on the complexity of the entropic-recognizability problem: given \mathbf{h} (represented as in Def. 30) check if $\mathbf{h} \in \bar{\Gamma}_n^*$. Consider the following restricted form of the problem: check if \mathbf{h} is the entropic vector of a representable space Ω (i.e. finite space with rational probabilities). This problem is in Σ_1^0 , because one can iterate over all representable spaces Ω and check that its entropies are those required. However, in the general setting we ask whether *any* finite probability space has these entropies, not necessarily one with rational probabilities. This problem would remain in Σ_1^0 if the theory of reals with exponentiation were decidable. Recall that Tarski's theorem states that the theory of reals $FO(\mathbb{R}, 0, 1, +, *)$ is decidable. A major open problem in model theory is whether the theory remains decidable if we add exponentiation. If that were decidable, then the entropic-recognizability problem were in Σ_1^0 . To see this, consider the following semi-decision problem. Iterate $N = 1, 2, 3, \dots$ and for each N check if there exists a probability space whose active domain has size N (thus, there are N^n outcomes, where $n = |\mathbf{V}|$ is the number of variables) whose entropies are precisely those given. This is statement that can be expressed using the exponential function (which we need in order to express the entropy as $\sum_i p_i \log p_i$). If there exists any finite probability space with the required entropies, then this procedure will find it; otherwise it will run forever, placing the problem in Σ_1^0 .