Decision Problems in Information Theory

Mahmoud Abo Khamis
relationalAI, Berkeley, CA, USA

Phokion G. Kolaitis
University of California, Santa Cruz, CA, USA
IBM Research – Almaden, 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

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.106

Category Track B: Automata, Logic, Semantics, and Theory of Programming

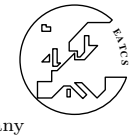
Related Version A full version of the paper is available at `https://arxiv.org/abs/2004.08783`.

Funding *Phokion G. Kolaitis*: Research partially supported by NSF Grant IIS-1814152.
Dan Suciu: Research partially supported by NSF IIS-1907997, NSF III-1703281, NSF III-1614738, and NSF AiTF-1535565.

Acknowledgements We thank Miika Hannula for several useful pointers to earlier work on the implication problem for conditional independence.

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 [40] 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) $H(\varnothing) = 0$, (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 [46]. In a celebrated result published in 1998, Zhang and Yeung [51] 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., [25, 34]). 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 [41]. The arithmetical 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 [43]. 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 7).
(2) Checking the validity of a conditional information inequality whose antecedents are "tight" is in Π_3^0 (Corollary 11). "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 14).
(4) Checking the validity of a group-balanced, max information inequality is Turing equivalent to checking the validity of an information inequality (Corollary 17).

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 [25] 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 10; see also Equation(19)), and that any max-inequality also follows from a single linear inequality (Theorem 16).

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 $\overline{\Gamma}_n^*$. 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 $\overline{\Gamma}_n^*$, our results are only for $\overline{\Gamma}_n^*$. 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 [44], and its decidability remains open. Our result (2) above does not apply here because it is a statement about $\overline{\Gamma}_n^*$-validity. However, we prove that the implication problem for conditional independence statements is in Π_1^0 (Theorem 8).

2 Background and Notations

Throughout this paper, vectors and tuples are denoted by bold-faced letters, and random variables are capitalized. We write $\boldsymbol{x} \cdot \boldsymbol{y} \stackrel{\text{def}}{=} \sum_i x_i y_i$ for the dot product of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$. For a given set $S \subseteq \mathbb{R}^m$, S is *convex* if $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\theta \in [0,1]$ implies $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y} \in S$; S is called a *cone* if $\boldsymbol{x} \in S$ and $\theta \geq 0$ implies $\theta\boldsymbol{x} \in S$; the topological closure of S is denoted by $\overline{S}$; and, finally, $S^* \stackrel{\text{def}}{=} \{\boldsymbol{y} \mid \forall \boldsymbol{x} \in S, \boldsymbol{x} \cdot \boldsymbol{y} \geq 0\}$ denotes the *dual cone* of S. It is known that S^* is always a closed, convex cone. We provide more background in the full version [2].

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

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

In this paper all logarithms are in base 2.

Fix a joint distribution over n finite random variables $\boldsymbol{V} \stackrel{\text{def}}{=} \{X_1, \ldots, X_n\}$. For each $\alpha \subseteq [n]$, let $\boldsymbol{X}_\alpha$ denote the random (vector-valued) variable $(X_i : i \in \alpha)$. Define the set function $h : 2^{[n]} \to \mathbb{R}_+$ by setting $h(\alpha) \stackrel{\text{def}}{=} H(\boldsymbol{X}_\alpha)$, for all $\alpha \subseteq [n]$. With some abuse, we blur the distinction between the set $[n]$ and the set of variables $\boldsymbol{V} = \{X_1, \ldots, X_n\}$, and write $H(\boldsymbol{X}_\alpha)$, $h(\boldsymbol{X}_\alpha)$, or $h(\alpha)$ interchangeably. We call the function h an *entropic function*, and also identify it with a vector $\boldsymbol{h} \stackrel{\text{def}}{=} (h(\alpha))_{\alpha \subseteq [n]} \in \mathbb{R}_+^{2^n}$, which is called an *entropic vector*. Note that most texts and papers on this topic drop the component $h(\varnothing)$, which is always 0, leading to entropic vectors in $\mathbb{R}^{2^n-1}$. We prefer to keep the $\varnothing$-coordinate to simplify notations. The implicit assumption $h(\varnothing) = 0$ 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 [46] 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 [46].

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

$$h(\boldsymbol{Y} \cup \boldsymbol{X}) \geq h(\boldsymbol{X}) \qquad\qquad h(\boldsymbol{X}) + h(\boldsymbol{Y}) \geq h(\boldsymbol{X} \cup \boldsymbol{Y}) + h(\boldsymbol{X} \cap \boldsymbol{Y})$$

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

▶ Remark 1. The established notation Γ_n^* [47, 50, 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. However in Section 5 we briefly discuss the dual problem, namely, recognizing whether a given vector $\boldsymbol{h}$ is an entropic vector (or an almost entropic vector).

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

3.1 Problem Definition

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

$$\boldsymbol{c} \cdot \boldsymbol{h} = \sum_{\alpha \subseteq [n]} c_\alpha h(X_\alpha) \geq 0. \tag{2}$$

The information inequality is said to be *valid* if it holds for all vectors $\boldsymbol{h} \in \Gamma_n^*$; equivalently, $\boldsymbol{c}$ is in the dual cone, $\boldsymbol{c} \in (\Gamma_n^*)^*$. By continuity, an information inequality holds $\forall \boldsymbol{h} \in \Gamma_n^*$ iff it holds $\forall \boldsymbol{h} \in \overline{\Gamma}_n^*$. In 1986, Pippenger [40] 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 [51]. 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 $\boldsymbol{c}_j \in \mathbb{R}^{2^n}, j \in [m]$, we associate the following* Boolean information constraint*:*

$$F(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0, \ldots, \boldsymbol{c}_m \cdot \boldsymbol{h} \geq 0). \tag{3}$$

For a set $S \subseteq \mathbb{R}^{2^n}$, a Boolean information constraint is said to be S*-valid* if it holds for all $\boldsymbol{h} \in S$. Thus, we will distinguish between Γ_n^*-validity and $\overline{\Gamma}_n^*$-validity. Unlike in the case of information inequalities, 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* EBIC(F)*, is the following: given m integer vectors $\boldsymbol{c}_j \in \mathbb{Z}^{2^n}$, where $j \in [m]$, check whether the constraint* (3) *holds for all entropic functions $\boldsymbol{h} \in \Gamma_n^*$. In the* almost-entropic *version, denoted by* AEBIC(F)*, we replace Γ_n^* by $\overline{\Gamma}_n^*$.*

The inputs $\boldsymbol{c}_j, j \in [m]$, to these problems are required to be integer vectors in order for EBIC(F) and AEBIC(F) to be meaningful computational problems. Equivalently, one can require the inputs to be rational vectors $\boldsymbol{c}_j \in \mathbb{Q}^{2^n}, j \in [m]$.

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

$$(Z'_1 \wedge \cdots \wedge Z'_k) \Rightarrow (Z_1 \vee \cdots \vee Z_\ell) \tag{4}$$

Problem	Abbreviation		Simple Example
	Entropic	Almost-entropic	
Boolean information constraint	EBIC(F)	AEBIC(F)	$h(XY) \le \frac{2}{3}h(XYZ) \Rightarrow \max(h(YZ), h(XZ)) \ge \frac{2}{3}h(XYZ)$
Information Inequality	IIP		$h(XY) + h(YZ) + h(XZ) \ge 2h(XYZ)$
Max-Information Inequality	MaxIIP		$\max(h(XY), h(YZ), h(XZ)) \ge \frac{2}{3}h(XYZ)$
Conditional Information Inequality	ECIIP	AECIIP	$((h(XY) \le \frac{2}{3}h(XYZ)) \wedge (h(YZ) \le \frac{2}{3}h(XYZ))) \Rightarrow h(XZ) \ge \frac{2}{3}h(XYZ)$
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.

where $Z'_1, \ldots, Z'_k, Z_1, \ldots, Z_\ell$ are distinct Boolean variables. Checking EBIC(F) is equivalent to checking EBIC(C), for each clause of F (and similarly for 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*. When $k > 0$, the formula F is *non-monotone*, and we call the corresponding constraint *conditional*.

Simple and Max Constraints. When $k = 0$ and $\ell = 1$, then we say that F defines a *simple* inequality; when $k = 0$ and $\ell > 1$, then we say that F defines a max-*inequality*. The case when $\ell = 0$ and $k > 0$ is not interesting because F is not valid, since the zero-vector $\boldsymbol{h} = \mathbf{0}$ violates the constraint.

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.

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 arises 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 $\boldsymbol{c}$, check whether Eq. (2) is Γ_n^*-valid or, equivalently, $\overline{\Gamma}_n^*$-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 [46, 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 [51]. Later, Matúš [34] proved that, when $n \ge 4$ variables, there exists infinitely many inequivalent non-Shannon entropic inequalities. More precisely, he proved that the following is a non-Shannon inequality, for every $k \ge 1$:

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

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 [49, 31, 48], 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 $(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0) \vee \cdots \vee (\boldsymbol{c}_m \cdot \boldsymbol{h} \geq 0)$, where $\boldsymbol{c}_j \in \mathbb{R}^{2^n}$. This is equivalent to:

$$\max(\boldsymbol{c}_1 \cdot \boldsymbol{h}, \boldsymbol{c}_2 \cdot \boldsymbol{h}, \ldots, \boldsymbol{c}_m \cdot \boldsymbol{h}) \geq 0 \tag{6}$$

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

Application to Constraint Satisfaction and Database Theory. Given two finite structures $\boldsymbol{A}$ and $\boldsymbol{B}$, we write $\mathsf{HOM}(\boldsymbol{A}, \boldsymbol{B})$ for the set of homomorphisms from $\boldsymbol{A}$ to $\boldsymbol{B}$. We say that $\boldsymbol{B}$ *dominates* structure $\boldsymbol{A}$, denote by $\boldsymbol{A} \preceq \boldsymbol{B}$, if for every finite structure $\boldsymbol{C}$, we have that $|\mathsf{HOM}(\boldsymbol{A}, \boldsymbol{C})| \leq |\mathsf{HOM}(\boldsymbol{B}, \boldsymbol{C})|$. The *homomorphism domination problem* asks whether $\boldsymbol{A} \preceq \boldsymbol{B}$, given $\boldsymbol{A}$ and $\boldsymbol{B}$. 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 $Q_1 \to D$ [28]. 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 [29] described a MaxIIP problem that yields a sufficient condition for homomorphism domination. In recent work [1] we proved that, when $\boldsymbol{B}$ is acyclic, then that condition is also necessary, and, moreover, the domination problem for acyclic $\boldsymbol{B}$ is Turing-equivalent to MaxIIP. Hence, any result on the complexity of MaxIIP immediately carries over to the homomorphism domination problem for acyclic $\boldsymbol{B}$, 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 $Q_1 \preceq Q_2$ follows from the following max-inequality:

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

3.2.3 Conditional Information Inequalities

A *conditional information inequality* has the form:

$$(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0 \wedge \cdots \wedge \boldsymbol{c}_k \cdot \boldsymbol{h} \geq 0) \Rightarrow \boldsymbol{c}_0 \cdot \boldsymbol{h} \geq 0 \tag{8}$$

Here we need to distinguish between Γ_n^*-validity and $\overline{\Gamma}_n^*$-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 $\boldsymbol{c}_i \cdot \boldsymbol{h} = 0$ is equivalent to $\boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0 \wedge -\boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0$.

Suppose that there exist $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$ such that the inequality $\boldsymbol{c}_0 \cdot \boldsymbol{h} - (\sum_i \lambda_i \boldsymbol{c}_i \cdot \boldsymbol{h}) \geq 0$ is valid; then Eq. (8) is, obviously, also valid. Kaced and Romashchenko [25] called Eq. (8) an *essentially conditioned inequality* if no such λ_i's exist, and discovered several valid conditional inequalities that are essentially conditioned.

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

The CI implication problem has been studied extensively in the literature [30, 44, 18, 27]. Pearl and Paz [39] gave a sound, but incomplete, set of *graphoid axioms*, Studený [44] 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, 38] for some recent work on the CI implication problem. The decidability of the CI implication problem remains open to date.

Results in [25] imply that the following CI implication is essentially conditioned (see [27]):

$$I_h(C; D|A) = I_h(C; D|B) = I_h(A; B) = I_h(B; C|D) = 0 \Longrightarrow I_h(C; D) = 0 \tag{9}$$

While a CI implication problem is an instance of an *entropic* conditional inequality, one can also consider the question whether a CI implication statement holds for all *almost entropic* functions; for example the implication (9) holds for all almost entropic functions. Kaced and Romashchenko [25] proved that these two problems differ, by giving examples of CI implications that hold for all entropic functions but fail for almost entropic functions.

3.2.4 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 $(G; G_1, \ldots, G_n)$ is called a *group system* if G is a finite group and $G_1, \ldots, G_n \subseteq G$ 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 $\boldsymbol{c} \subseteq \mathbb{R}^{2^n}$ defines the following *group-theoretic inequality*:

$$\sum_{\alpha \subseteq [n]} c_\alpha \log \frac{|G|}{|G_\alpha|} \geq 0 \tag{10}$$

▶ **Theorem 4** ([12]). *An information inequality* (2) *is* Γ_n^**-valid if and only if the corresponding group-theoretic inequality* (10) *holds for all group systems* $(G, G_1, \ldots, G_n)$*,*

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 (10). We note that the group-theoretic inequalities considered here are different from the word problems in group, see e.g. the survey [35]; the undecidability results for word problems in groups do not carry over to the group-theoretic inequalities and, thus, to information inequalities.

3.2.5 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, 5, 14, 36]. 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 [6] 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, 3, 4]. 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 $XZ \to U$ and $YU \to X$ 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:

$$h(XY) + h(YZ) + h(ZU) + h(X|YU) + h(U|XZ) \geq 2h(XYZU) \tag{11}$$

While the tight upper bound of any conjunctive query can be proven using only Shannon-type inequalities, this no longer holds when the relations used in the query are constrained to satisfy functional dependencies. In that case, the tight upper bound can always be obtained from an information inequality, but Abo Khamis et al. [4] gave an example of a conjunctive query for which the tight upper bound requires a non-Shannon inequality.

3.2.6 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 [42] and Blakley [8]. The setup is as follows. There is a set P of *participants*, a *dealer* $d \notin P$, and an *access structure* $\mathcal{F} \subset 2^P$. 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 [7]). 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 $(X_1, \ldots, X_n)$ satisfying:

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

Intuitively, X_i denotes the share given to the ith 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 [9], 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, 26] 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 $\boldsymbol{h} \in \Gamma_n^*$ 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:

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

4 Placing EBIC and AEBIC in the Arithmetical Hierarchy

What is the complexity of EBIC(F) / 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 as follows. The class Σ_n^0 consists of all sets of the form $\{x \mid \exists y_1 \forall y_2 \exists y_3 \cdots \mathsf{Q} y_n R(x, y_1, \ldots, y_n)\}$, where R is an $(n+1)$-ary decidable predicate, $\mathsf{Q} = \exists$ 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 $\{x \mid \forall y_1 \exists y_2 \forall y_3 \cdots \mathsf{Q} y_n R(x, y_1, \ldots, y_n)\}$. 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 [41]. Our goal is to place the problems EBIC(F), AEBIC(F), and their variants in 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 5.** *If F is monotone, then EBIC(F) and 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 $X_1, \ldots, X_n, p$, 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 $\boldsymbol{h}^\Omega$ its entropic vector. By continuity of Eq.(1), we obtain:

▶ **Proposition 6.** *For every entropic vector $\boldsymbol{h} \in \Gamma_n^*$ and every $\varepsilon > 0$, there exists a representable space Ω such that $\|\boldsymbol{h} - \boldsymbol{h}^\Omega\| < \varepsilon$.*

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 7.** *Let F be a monotone Boolean formula. Then* EBIC(F) *(and, hence,* AEBIC(F)*) is in Π_1^0, i.e., it is co-recursively enumerable.*

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

$$\forall \boldsymbol{h} \in \Gamma_n^* : \qquad \boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0 \vee \cdots \vee \boldsymbol{c}_m \cdot \boldsymbol{h} \geq 0 \tag{12}$$

We claim that (12) is equivalent to:

$$\forall \Omega \qquad \boldsymbol{c}_1 \cdot \boldsymbol{h}^\Omega \geq 0 \vee \cdots \vee \boldsymbol{c}_m \cdot \boldsymbol{h}^\Omega \geq 0 \tag{13}$$

Obviously (12) implies (13), and the opposite follows from Prop. 6: if (12) fails on some entropic vector $\boldsymbol{h}$, then it also fails on some representable $\boldsymbol{h}^\Omega$ close enough to $\boldsymbol{h}$. Finally, (13) is in Π_1^0 because, the property after $\forall\Omega$ is decidable, by expanding the definition of entropy (1) in each condition $\boldsymbol{c}_i \cdot \boldsymbol{h}^\Omega \geq 0$, 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 $\overline{\Gamma}_n^*$-validity no longer coincide, we study EBIC(F) and AEBIC(F) separately. The results here are non-trivial, and some proofs are deferred to [2].

4.2.1 The Entropic Case

Our result for EBIC(F) is restricted to the CI implication problem. Recall from Sec. 3.2.3 that this problem consists of checking whether an implication between statements of the form $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{Z} \mid \boldsymbol{X})$ holds for all random variables with finite domain, and this is equivalent to checking whether a certain conditional inequality holds for all entropic functions. We prove that this problem is in Π_1^0 by using Tarski's theorem of the decidability of the theory of reals with $+, *$ [45].

▶ **Theorem 8.** *The CI implication problem (Section 3.2.3) is in Π_1^0.*

Proof. 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 whether[1] $\Phi \equiv \forall x \exists y \forall z (x^2 + 3y \geq z \wedge (y^3 + yz \leq xy^2))$ is true. We will write $(\mathbb{R}, +, *) \vDash \Phi$ to denote the fact that Φ is true in the model of reals.

Consider a conditional inequality over a set of n joint random variables:

$$I_h(Y_1; Z_1 | X_1) = 0 \wedge \cdots \wedge I_h(Y_k; Z_k | X_k) = 0 \Rightarrow I_h(Y; Z | X) = 0$$

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.

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

- Consider n joint random variables $X_1, \ldots, X_n$ where each has outcomes in the domain $[N]$; thus there are N^n possible outcomes. Let $p_1, \ldots, p_{N^n}$ be real variables representing the probabilities of these outcomes.
- Construct a formula Δ stating "there exist probabilities $p_1, \ldots, p_{N^n}$ 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: $p(X_iY_iZ_i)p(X_i) = p(X_iY_i)p(X_iZ_i)$.
  - Replace each such statement with a conjunction of statements of the form $p(X_i = x, Y_i = y, Z_i = z) \cdot p(X_i = x) = p(X_i = x, Y_i = y) \cdot p(X_i = x, Z_i = z)$, for all combinations of values x, y, z. If X_i, Y_i, Z_i have in total k random variables, then there are N^k combinations of values x, y, z, thus we create a conjunction of N^k equality statements.
  - Each marginal probability is a sum of atomic probabilities, for example $p(X_i = x, Y_i = y) = p_{k_1} + p_{k_2} + \cdots$ where $p_{k_1}, p_{k_2}, \ldots$ are the probabilities of all outcomes that have $X_i = x$ and $Y_i = y$. Thus, the equality statement in the previous step becomes the following formula: $(p_{i_1} + p_{i_2} + \cdots)(p_{j_1} + p_{j_2} + \cdots) = (p_{k_1} + p_{k_2} + \cdots)(p_{\ell_1} + p_{\ell_2} + \cdots)$. 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 $\Phi = \Phi_1 \wedge \cdots \wedge \Phi_k$. Let Ψ be the similar formula for the consequent: thus, Ψ asserts $I_h(Y; Z|X) = 0$.
  - Finally, construct the formula $\Delta \stackrel{\text{def}}{=} \exists p_1, \ldots, \exists p_{N^n}, (\Phi \wedge \neg\Psi)$.
- Check whether $(\mathbb{R}, +, *) \models \Delta$. By Tarski's theorem this step is decidable.
- If Δ is true, then return *false*; otherwise, continue with $N + 1$. ◀

Tarski's exponential function problem

One may attempt to extend the proof above from the CI implication problem to arbitrary conditional inequalities (8). To check if a conditional inequality is valid for all entropic functions, we can repeat the argument above: iterate over all domain sizes $N = 1, 2, 3, \ldots$, and check if there exists probabilities $p_1, \ldots, p_{N^n}$ that falsify the implication $(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0 \wedge \cdots \wedge \boldsymbol{c}_k \cdot \boldsymbol{h} \geq 0) \Rightarrow \boldsymbol{c}_0 \cdot \boldsymbol{h} \geq 0$. The problem is that in order to express $\boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0$ we need to express the vector $\boldsymbol{h}$ in terms of the probabilities $p_1, \ldots, p_{N^n}$. To apply directly the definition of entropy in (1) we need to use the log function, or, alternatively, the exponential function, and this takes us outside the scope of Tarski's theorem. A major open problem in model theory, originally formulated also by Tarski, is whether decidability continues to hold if we augment the structure of the real numbers with the exponential function (see, e.g., [32] for a discussion). Decidability of the first-order theory of the reals with exponentiation would easily imply that the entropic conditional information inequality problem ECIIP (not just the entropic conditional independence (CI) implication problem) is in Π_1^0, because every condition $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$ 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 $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$. Call $\boldsymbol{c} \in \mathbb{R}^{2^n}$ *tight* if $\boldsymbol{c} \cdot \boldsymbol{h} \leq 0$ is $\overline{\Gamma}_n^*$-valid. When $\boldsymbol{c}$ is tight, we can rewrite $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$ as $\boldsymbol{c} \cdot \boldsymbol{h} = 0$. If $\boldsymbol{c}$ is not tight, then there exists $\boldsymbol{h} \in \overline{\Gamma}_n^*$ such that $\boldsymbol{c} \cdot \boldsymbol{h} > 0$; in that case we say that $\boldsymbol{c}$ *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 $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\} \subset \mathbb{R}^{2^n}$ has slack if there exists $\boldsymbol{h} \in \overline{\Gamma}_n^*$ such that $\boldsymbol{c}_i \cdot \boldsymbol{h} > 0$ for all $i = 1, k$; notice that this is more restricted than requiring each of $\boldsymbol{c}_i$ 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 $\boldsymbol{c}_k \cdot \boldsymbol{h} \geq 0$ from the antecedent to the consequent:

▶ **Theorem 9.** *The following statements are equivalent:*

$$\forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \bigwedge_{i \in [k]} \boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0 \Rightarrow \boldsymbol{c} \cdot \boldsymbol{h} \geq 0 \tag{14}$$

$$\forall \varepsilon > 0, \exists \lambda \geq 0, \forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \bigwedge_{i \in [k-1]} \boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0 \Rightarrow \boldsymbol{c} \cdot \boldsymbol{h} + \varepsilon h([n]) \geq \lambda \boldsymbol{c}_k \cdot \boldsymbol{h} \tag{15}$$

Moreover, if the set $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$ has slack, then one can set $\varepsilon = 0$ in Eq.(15).

Proof. We prove here only the implication from (15) to (14); the other direction is non-trivial and is proven in the full version [2] using only the properties of closed convex cones. Assume condition (15) holds, and consider any $\boldsymbol{h} \in \overline{\Gamma}_n^*$ s.t. $\bigwedge_{i \in [k]} \boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0$. We prove that $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$. For any $\varepsilon > 0$, condition (15) states that there exists $\lambda > 0$ such that $\boldsymbol{c} \cdot \boldsymbol{h} + \varepsilon h([n]) \geq \lambda \boldsymbol{c}_k \cdot \boldsymbol{h}$ and therefore $\boldsymbol{c} \cdot \boldsymbol{h} + \varepsilon h([n]) \geq 0$. Since $\varepsilon > 0$ is arbitrary, we conclude that $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$, as required. ◀

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

▶ **Corollary 10.** *Condition* (14) *is equivalent to:*

$$\forall \varepsilon > 0, \exists \lambda_1 \geq 0, \cdots, \exists \lambda_k \geq 0, \forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \boldsymbol{c} \cdot \boldsymbol{h} + \varepsilon h([n]) \geq \sum_{i \in [k]} \lambda_i \boldsymbol{c}_i \cdot \boldsymbol{h} \tag{16}$$

Moreover, if the set $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$ has slack, then one can set $\varepsilon = 0$ in Eq.(16).

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.7. In that case, condition (14) is equivalent to:

$$\forall p \in \mathbb{N}, \exists q \in \mathbb{N}, \forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \boldsymbol{c} \cdot \boldsymbol{h} + \frac{1}{p} h([n]) \geq q \sum_{i \in [k]} \boldsymbol{c}_i \cdot \boldsymbol{h} \tag{17}$$

Indeed, the non-trivial direction (16)⇒ (17) follows by setting $q \stackrel{\text{def}}{=} \lceil \max(\lambda_1, \ldots, \lambda_k) \rceil \in \mathbb{N}$ and noting that $\boldsymbol{c}_i$ is tight, hence $\boldsymbol{c}_i \cdot \boldsymbol{h} \leq 0$ and therefore $\lambda_i \boldsymbol{c}_i \cdot \boldsymbol{h} \geq q \boldsymbol{c}_i \cdot \boldsymbol{h}$.

▶ **Corollary 11.** *Consider a conditional inequality* (8)*. If all antecedents are tight, then the corresponding decision problem AECIIP is in* Π_3^0

Proof. Based on our discussion, the inequality (8) is equivalent to condition (17), which is of the form $\forall p \exists q \forall \boldsymbol{h}$. Replace $\boldsymbol{h}$ with a representable entropic vector $\boldsymbol{h}^\Omega$, as in the proof of Theorem 7, and it becomes $\forall p \exists q \forall \boldsymbol{h}^\Omega$, placing it in Π_3^0. ◀

Recall that the implication problem for CI is a special case of a conditional inequality with tight antecedents. We have seen in Theorem 8 that the *entropic* version of the CI implication problem is in Π_1^0; Corollary 11 proves that the *almost entropic* version is in Π_3^0.

Consider any conditional inequality (8) where the antecedents are tight. If this inequality holds for all almost entropic functions, then it can be proven by proving a family of (unconditional) inequalities (17). In fact, some conditional inequalities in the literature have been

proven precisely in this way. For example, consider the CI implication (9) (Sec. 3.2.3), and replace each antecedent $I_h(\boldsymbol{Y};\boldsymbol{Z}|\boldsymbol{X}) = 0$ with $-I_h(\boldsymbol{Y};\boldsymbol{Z}|\boldsymbol{X}) \geq 0$. By Eq. (17), the following condition holds: $\forall p \in \mathbb{N}, \exists q \in \mathbb{N}$ such that

$$q(I_h(C;D \mid A) + I_h(C;D \mid B) + I_h(A;B) + I_h(B;C \mid D)) + \frac{1}{p}h(ABCD) \geq I_h(C;D) \tag{18}$$

Thus, in order to prove (9), it suffices to prove (18). Matúš's inequality (5) provides precisely the proof of (18) (by setting $k \stackrel{\text{def}}{=} p$, $q \stackrel{\text{def}}{=} \max(\lceil \frac{k+3}{2} \rceil, 1)$, and observing that $I_h(B;D \mid C) \leq h(ABCD)$).

Antecedents Have Slack. Next, we consider the case when the antecedents have slack, which is a recursively enumerable condition. In that case, condition (16) is equivalent to:

$$\exists \lambda_1 \geq 0, \cdots, \exists \lambda_k \geq 0, \forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \boldsymbol{c} \cdot \boldsymbol{h} \geq \sum_{i \in [k]} \lambda_i \boldsymbol{c}_i \cdot \boldsymbol{h} \tag{19}$$

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 (19) to prove complexity bounds for $\mathsf{AEBIC}(F)$, because the λ_i's in (19) are not necessarily rational numbers. When we derived Eq. (17) we used the fact that the antecedents are tight, hence $\boldsymbol{c}_i \cdot \boldsymbol{h} \leq 0$, hence we could replace the λ_i's with some natural number q larger than all of them. But now, the sign of $\boldsymbol{c}_i \cdot \boldsymbol{h}$ 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 the full version [2] we give evidence that some restriction is necessary to ensure the λ_i's are rationals, and also show that every conditional inequality can be strengthened to be group balanced.

A vector $\boldsymbol{h} \in \mathbb{R}^{2^n}$ is called *modular* if $h(\boldsymbol{X}) + h(\boldsymbol{Y}) = h(\boldsymbol{X} \cup \boldsymbol{Y}) + h(\boldsymbol{X} \cap \boldsymbol{Y})$ forall sets of variables $\boldsymbol{X}, \boldsymbol{Y} \subseteq \boldsymbol{V}$. Every non-negative modular function is entropic [46], and is a non-negative linear combination of the *basic modular functions* $\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(n)}$, 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 $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$ *balanced* if $\boldsymbol{c} \cdot \boldsymbol{h}^{(j)} = 0$ for every $j \in [n]$. He proved that *any* valid inequality can be strengthened to a balanced one. More precisely: $\boldsymbol{c} \cdot \boldsymbol{h} \geq 0$ is valid iff $\boldsymbol{c} \cdot \boldsymbol{h}^{(i)} \geq 0$ for all $i \in [n]$ and $\boldsymbol{c} \cdot \boldsymbol{h} - \sum_i (\boldsymbol{c} \cdot \boldsymbol{h}^{(i)}) h(X_i \mid X_{[n]-\{i\}}) \geq 0$ 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 12.** *Call a set* $\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_k\} \subseteq \mathbb{R}^{2^n}$ group balanced *if (a)* $\text{rank}(\boldsymbol{A}) = k - 1$ *where* $\boldsymbol{A}$ *is the* $k \times n$ *matrix* $A_{ij} = \boldsymbol{d}_i \cdot \boldsymbol{h}^{(j)}$*, and (b) there exists a non-negative modular function* $\boldsymbol{h}^{(*)} \neq 0$ *such that* $\boldsymbol{d}_i \cdot \boldsymbol{h}^{(*)} = 0$ *for all* i*.*

If $k = 1$ then $\{\boldsymbol{d}_1\}$ is group balanced iff $\boldsymbol{d}_1$ is balanced, because the matrix $\boldsymbol{A}$ has a single row $\left(\boldsymbol{d} \cdot \boldsymbol{h}^{(1)} \cdots \boldsymbol{d} \cdot \boldsymbol{h}^{(n)}\right)$, and its rank is 0 iff all entries are 0. We prove in [2]:

▶ **Theorem 13.** *Consider a group balanced set of* n *vectors with rational coefficients,* $D = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n\} \subseteq \mathbb{Q}^{2^n}$*. Suppose the following condition holds:*

$$\exists \lambda_1 \geq 0, \cdots, \exists \lambda_n \geq 0, \sum_{i \in [n]} \lambda_i = 1, \forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \sum_{i \in [n]} \lambda_i \boldsymbol{d}_i \cdot \boldsymbol{h} \geq 0 \tag{20}$$

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

This implies that, if $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$ have slack and $\{\boldsymbol{c}, -\boldsymbol{c}_1, \ldots, -\boldsymbol{c}_k\}$ is group balanced, then there exist rational λ_i's for inequality (19). In particular:

▶ **Corollary 14.** *Consider a conditional inequality* (8)*. If the antecedents have slack and* $\{\boldsymbol{c}, -\boldsymbol{c}_1, \ldots, -\boldsymbol{c}_k\}$ *is group balanced, then the corresponding decision problem is in* Σ_2^0*.*

We end this section by illustrating with an example:

▶ **Example 15.** Consider the following conditional inequality:

$$h(XYZ) + h(X) \geq 2h(XY) \wedge h(XYZ) + h(Y) \geq 2h(YZ) \quad \Rightarrow \quad 2h(XZ) \geq h(XYZ) + h(Z) \tag{21}$$

The antecedents have slack, because, by setting[2] $\boldsymbol{h} \stackrel{\text{def}}{=} 2\boldsymbol{h}^{(X)} + \boldsymbol{h}^{(Z)}$, 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 18 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 (21).

Consider now the following set $D = \{\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3\}$, where the vectors $\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3$ 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. 12 we verify that the matrix $\boldsymbol{A}$ has rank 2; in our example the matrix is $\boldsymbol{A} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and its rank is 2 as required. To check condition (b), we define $\boldsymbol{h}^{(*)} = h^{(X)} + h^{(Y)} + h^{(Z)}$ and verify that $\boldsymbol{d}_1 \cdot \boldsymbol{h}^{(*)} = \boldsymbol{d}_2 \cdot \boldsymbol{h}^{(*)} = \boldsymbol{d}_3 \cdot \boldsymbol{h}^{(*)} = 4 - 3 - 1 = 0$. 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 $\boldsymbol{A} \preceq \boldsymbol{B}$ for acyclic structures $\boldsymbol{B}$ 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 7 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.

▶ **Theorem 16.** *The following two statements are equivalent:*

$$\forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \bigwedge_{i \in [k]} \boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0 \Rightarrow \bigvee_{j \in [m]} \boldsymbol{d}_j \cdot \boldsymbol{h} \geq 0 \tag{22}$$

$$\exists \lambda_1, \ldots, \lambda_m \geq 0, \sum_j \lambda_j = 1, \forall \boldsymbol{h} \in \overline{\Gamma}_n^* : \qquad \bigwedge_{i \in [k]} \boldsymbol{c}_i \cdot \boldsymbol{h} \geq 0 \Rightarrow \sum_{j \in [m]} \lambda_j \boldsymbol{d}_j \cdot \boldsymbol{h} \geq 0 \tag{23}$$

The theorem says that every max-inequality is essentially a linear inequality. The proof of (23) $\Rightarrow$ (22) is immediate; we prove the reverse in [2]. As before, we don't know whether these coefficients λ_i can be chosen to be rational numbers in general, but by Theorem 13 this is the case when $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$ is group-balanced, and this implies:

▶ **Corollary 17.** *The MaxIIP problem where the inequalities* $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$ *are group balanced is Turing equivalent to the IIP problem.*

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

Proof. We describe a Turing reduction from MaxIIP to IIP. Consider a MaxIIP problem, $\bigvee_{j\in[m]}(\boldsymbol{c}_j \cdot \boldsymbol{h} \geq 0)$. We run two computations in parallel. The first computation iterates over all representable spaces Ω, and checks whether $\bigwedge_j(\boldsymbol{c}_j \cdot \boldsymbol{h}^\Omega < 0)$; 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 $(\lambda_1, \ldots, \lambda_m) \in \mathbb{N}^m$ and checks $\forall \boldsymbol{h} \in \Gamma_n^*, \sum_j \lambda_j \boldsymbol{c}_j \cdot \boldsymbol{h} \geq 0$ 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 16 and 13. ◀

We illustrate with an example.

▶ **Example 18.** Consider Kopparty and Rossman's inequality (7), which can be stated as $\max(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3) \geq 0$, where $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ 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 here[3]:

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

Theorem 16 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 $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ is group balanced (as we showed in Example 15), therefore there exists rational λ_i's; indeed, our choice here is $\lambda_1 = \lambda_2 = \lambda_3 = 1$.

5 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 $\boldsymbol{h}$ and checks if $\boldsymbol{h} \in \Gamma_n^*$. The *almost-entropic-recognizability problem* checks if $\boldsymbol{h} \in \overline{\Gamma}_n^*$. 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 must first address the question of how to represent the input $\boldsymbol{h}$. One possibility is to represent $\boldsymbol{h}$ 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 $\boldsymbol{h}$ 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 the quantity $h(\boldsymbol{X})$ has the form $h(\boldsymbol{X}) = \log \prod_i p_i^{p_i}$. In general, any product $\prod_i m_i^{n_i}$ where $m_i, n_i \in Q$, 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:

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

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

Justified by this observation, we assume that the input to our problem consists of three vectors $(a_{\boldsymbol{X}})_{\boldsymbol{X}\subseteq\boldsymbol{V}}$, $(b_{\boldsymbol{X}})_{\boldsymbol{X}\subseteq\boldsymbol{V}}$, and $(c_{\boldsymbol{X}})_{\boldsymbol{X}\subseteq\boldsymbol{V}}$ in $\mathbb{N}^{2^n}$, with the convention that $h(\boldsymbol{X}) \stackrel{\text{def}}{=} \frac{1}{c_{\boldsymbol{X}}} \log \frac{a_{\boldsymbol{X}}}{b_{\boldsymbol{X}}}$. 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 19** ((Almost-)Entropic Recognizability Problem)**.** *Given natural numbers* $(a_{\boldsymbol{X}})_{\boldsymbol{X}\subseteq\boldsymbol{V}}$, $(b_{\boldsymbol{X}})_{\boldsymbol{X}\subseteq\boldsymbol{V}}$ *and* $(c_{\boldsymbol{X}})_{\boldsymbol{X}\subseteq\boldsymbol{V}}$*, check whether the vector* $h(\boldsymbol{X}) \stackrel{\text{def}}{=} \frac{1}{c_{\boldsymbol{X}}} \log \frac{a_{\boldsymbol{X}}}{b_{\boldsymbol{X}}}$, $\boldsymbol{X} \subseteq \boldsymbol{V}$*, represents an entropic vector, or an almost-entropic vector.*

Our result in this section is (see [2] for a proof):

▶ **Theorem 20.** *The almost entropic recognizability problem is in* Π_2^0.

We end with a brief comment on the complexity of the entropic-recognizability problem: given $\boldsymbol{h}$ (represented as in Def. 19) check if $\boldsymbol{h} \in \Gamma_n^*$. Consider the following restricted form of the problem: check if $\boldsymbol{h}$ 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 their 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 $\mathrm{FO}(\mathbb{R}, 0, 1, +, *)$ 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 would be in Σ_1^0. To see this, consider the following semi-decision problem. Iterate over $N = 1, 2, 3, \ldots$ and for each N check if there exists a probability space whose active domain has size N (thus, there are N^n outcomes, where $n = |\boldsymbol{V}|$ is the number of variables) and whose entropies are precisely those given. This 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.

6 Discussion

CI Implication Problem. The implication problem for Conditional Independence statements has been extensively studied in the literature, but its complexity remains an open problem. It is not even known whether this problem is decidable [18, 37, 38]. Our Theorem 8 appears to be the first upper bound on the complexity of the CI implication problem, placing it in Π_1^0. Hannula et al. [24] prove that, if all random variables are restricted to be binary random variables, then the CI implication problem is in EXPSPACE; the implication problem for binary random variables differs from that for general discrete random variables; see the discussion in [18].

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 random variables, or over discrete infinite random variables with finite entropy. For example discrete (possibly infinite) random variables are considered by Zhang and Yeung, [50], by Chan and Yeung [12], and by Chan [22], while random variables with finite domains are considered by Matúš [33, 34] and by Kaced and Romashchenko [25]. 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. 6 extends immediately to discrete random variables[4]. 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 [25] 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 $\boldsymbol{c} \cdot \boldsymbol{h} > 0$ instead. Obviously, each such inequality fails on the zero-entropic vector, but we can consider the conditional version $\boldsymbol{h} \neq 0 \Rightarrow \boldsymbol{c} \cdot \boldsymbol{h} > 0$, which we can write formally as $\boldsymbol{c} \cdot \boldsymbol{h} \leq 0 \Rightarrow h(\boldsymbol{V}) \leq 0$. This 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 $\overline{\Gamma}_n^*$-validity coincide; a negative answer would represent a significant extension of Kaced and Romashchenko's result [25].

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, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. Decision problems in information theory. *CoRR*, abs/2004.08783, 2020. `arXiv:2004.08783`.
3. 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.
4. 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.
5. 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. `doi:10.1007/BF02761855`.
6. Albert Atserias, Martin Grohe, and Dániel Marx. Size bounds and query plans for relational joins. *SIAM J. Comput.*, 42(4):1737–1767, 2013. `doi:10.1137/110859440`.

[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 $\{1, 2, \ldots, N\}$ 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.

7 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. doi:10.1007/978-3-642-20901-7_2.

8 G. R. Blakley. Safeguarding cryptographic keys. In *Managing Requirements Knowledge, International Workshop on*, page 313, Los Alamitos, CA, USA, June 1979. IEEE Computer Society. doi:10.1109/AFIPS.1979.98.

9 Carlo Blundo, Alfredo De Santis, and Ugo Vaccaro. On secret sharing schemes. *Inf. Process. Lett.*, 65(1):25–32, 1998. doi:10.1016/S0020-0190(97)00194-4.

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. 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. URL: http://dl.acm.org/citation.cfm?id=153850.

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. 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. 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. 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. doi:10.1145/2220357.2220363.

20 Peter Grunwald and Paul Vitányi. Shannon information and Kolmogorov complexity. *arXiv preprint*, 2004. arXiv:cs/0410002.

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. doi:10.1016/j.ipl.2014.05.010.

22 Terence H. Chan. Balanced information inequalities. *Information Theory, IEEE Transactions on*, 49:3261–3267, January 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 Miika Hannula, Åsa Hirvonen, Juha Kontinen, Vadim Kulikov, and Jonni Virtema. Facets of distribution identities in probabilistic team semantics. *CoRR*, abs/1812.05873, 2018. arXiv:1812.05873.

25 Tarik Kaced and Andrei E. Romashchenko. Conditional information inequalities for entropic and almost entropic points. *IEEE Trans. Information Theory*, 59(11):7149–7167, 2013. doi:10.1109/TIT.2013.2274614.

26 Ehud D. Karnin, J. W. Greene, and Martin E. Hellman. On secret sharing systems. *IEEE Trans. Information Theory*, 29(1):35–41, 1983. doi:10.1109/TIT.1983.1056621.

27 Batya Kenig and Dan Suciu. Integrity constraints revisited: From exact to approximate implication. *CoRR*, abs/1812.09987, 2018. arXiv:1812.09987.

28 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. doi:10.1145/275487.275511.

29 Swastik Kopparty and Benjamin Rossman. The homomorphism domination exponent. *European Journal of Combinatorics*, 32(7):1097–1114, 2011. Homomorphisms and Limits.

30 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. doi:10.1109/TSE.1987.232847.

31 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. doi:10.4310/CIS.2002.v2.n2.a3.

32 David Marker. Model theory and exponentiation, 1996.

33 Frantisek Matúš. Probabilistic conditional independence structures and matroid theory: Background, 1994.

34 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. doi:10.1109/ISIT.2007.4557201.

35 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.

36 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. doi:10.1145/3196959.3196990.

37 Mathias Niepert, Dirk Van Gucht, and Marc Gyssens. Logical and algorithmic properties of stable conditional independence. *Int. J. Approx. Reason.*, 51(5):531–543, 2010. doi:10.1016/j.ijar.2010.01.011.

38 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. doi:10.1016/j.artint.2013.06.005.

39 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.

40 Nicholas Pippenger. What are the laws of information theory. In *1986 Special Problems on Communication and Computation Conference*, pages 3–5, 1986.

41 Hartley Rogers and H Rogers. *Theory of recursive functions and effective computability*, volume 5. McGraw-Hill New York, 1967.

42 Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979. doi:10.1145/359168.359176.

43 Larry J Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.

44 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.

45 Alfred Tarski. A decision method for elementary algebra and geometry. In *Quantifier elimination and cylindrical algebraic decomposition*, pages 24–84. Springer, 1998.

46 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. doi:10.1007/978-1-4419-8608-5.

47 Raymond W. Yeung. *Information Theory and Network Coding*. Springer Publishing Company, Incorporated, 1 edition, 2008.

48 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. doi:10.4310/CIS.2001.v1.n1.a6.

49 Zhen Zhang. On a new non-Shannon type information inequality. *Commun. Inf. Syst.*, 3(1):47–60, 2003. doi:10.4310/CIS.2003.v3.n1.a4.

50 Zhen Zhang and Raymond W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Information Theory*, 43(6):1982–1986, 1997.

51 Zhen Zhang and Raymond W Yeung. On characterization of entropy function via information inequalities. *IEEE Transactions on Information Theory*, 44(4):1440–1452, 1998.