Information Inequality Problem over Set Functions

Miika Hannula
University of Helsinki, Finland

Abstract

Information inequalities appear in many database applications such as query output size bounds, query containment, and implication between data dependencies. Recently Khamis et al. [14] proposed to study the algorithmic aspects of information inequalities, including the information inequality problem: decide whether a linear inequality over entropies of random variables is valid. While the decidability of this problem is a major open question, applications often involve only inequalities that adhere to specific syntactic forms linked to useful semantic invariance properties. This paper studies the information inequality problem in different syntactic and semantic scenarios that arise from database applications. Focusing on the boundary between tractability and intractability, we show that the information inequality problem is coNP-complete if restricted to normal polymatroids, and in polynomial time if relaxed to monotone functions. We also examine syntactic restrictions related to query output size bounds, and provide an alternative proof, through monotone functions, for the polynomial-time computability of the entropic bound over simple sets of degree constraints.

2012 ACM Subject Classification Theory of computation → Database theory

Keywords and phrases entropy, information theory, worst-case output size, computational complexity

Digital Object Identifier 10.4230/LIPIcs.ICDT.2024.19

Related Version *Full Version*: `https://arxiv.org/abs/2309.11818` [9]

Funding *Miika Hannula*: The author has been supported by the ERC grant 101020762.

1 Introduction

Information inequalities are linear constraints on entropies of random variables. Often referred to as the laws of information, these inequalities describe what is not possible in information theory. More than three decades ago, Pippenger asked whether all such laws follow from the *polymatroidal axioms* [25], depicted in Fig. 1. The polymatroidal axioms are also known to be equivalent to the non-negativity of Shannon's information measures, which consist of entropy, conditional entropy, mutual information, and conditional mutual information. The inequality constraints derivable from the polymatroidal axioms are hence called *Shannon* inequalities. Pippenger's question was famously answered in the negative by Zhang and Yeung who were the first to find a non-Shannon information inequality that is valid over entropies [28]. Zhang and Yeung's proof was based on a novel innovation, identified as the *copy lemma* in [4], which still today remains essentially the only tool to establish novel non-Shannon inequalities [8].

Constraints on entropies are known to have many applications in database theory. Lee [19, 20] observed already in the 80s that database constraints can alternatively be expressed as equalities over information measures. More recently, the implication problem for data dependencies has been connected to validity of information inequalities [13], information theory has been used to analyze normal forms in relational and XML data models [1], and query containment for conjunctive queries under bag semantics – a notoriously difficult problem to study – has been proven to be equivalent in certain special cases to checking information inequalities involving maximum [15]. Perhaps the most fruitful application has been the use of information inequalities to obtain tight output size bounds for database queries [2, 6, 7, 11, 16, 17], and the subsequent development of *worst-case optimal join algorithms* that run in time proportional to these bounds [16, 17, 23, 24].

Recently Khamis et al. [14] initiated the study of the algorithmic properties of information inequalities. The most central problem, called the information inequality problem, is to decide whether a given information inequality is valid over all entropic functions. The decidability of this problem is a major open question in the foundations of information theory. It was shown in [14] that checking the validity of monotone Boolean combinations of information inequalities (including the aforementioned *max-inequalities*) is co-recursively enumerable (co-r.e.). Since the implication problem for conditional independence implication is undecidable [18, 21], validity for general Boolean combinations of information inequalities is known to be undecidable. While the focus of [14] was on generalizations of the information inequality problem, this paper shifts attention to simplifications of the problem. Many applications, such as implication problems or query output size bounds, are related to information inequalities that adhere to specific syntactic forms. These syntactic forms are also often linked to semantic invariance properties which render the associated problems computable and sometimes even tractable. Identifying factors that make the information inequality problem either easy or hard is thus a task that can prove beneficial in multiple application scenarios.

This paper examines the information inequality problem with respect to different syntactic restrictions and semantic settings, focusing in particular on the boundary between tractable and intractable cases. We demonstrate that different factors, including the influence of the coefficients and the expressiveness of the information measures, give rise to coNP-completeness with respect to normal polymatroids (the subset of entropic functions associated with a non-negative I-measure [13, 27]), and disagreement between normal polymatroids and entropic functions. Our findings also reveal that when we relax the semantics to monotone functions or restrict it to modular functions (an implicit result in existing literature), the information inequality problem can be solved in polynomial time. Additionally, we demonstrate that this problem becomes polynomial-time solvable when we impose syntactic restrictions linked to cases where computing the entropic query output size bound is known to be in polynomial time. Finally, we identify a syntactic restriction over which monotone and entropic functions agree, leading to an alternative proof for the previously established fact [11] that the entropic bound is polynomial-time computable over simple sets of degree constraints.

2 Preliminaries

We write $[n]$ for the set of integers $\{1,\dots,n\}$. We usually use boldface letters to denote sets. For two sets $\boldsymbol{X}$ and $\boldsymbol{Y}$, we write $\boldsymbol{XY}$ to denote their union. If A is an individual element, we sometimes write A instead of $\{A\}$ to denote the singleton set consisting of A.

2.1 Relational databases

Fix disjoint countably infinite sets Var and Val of variables and values. Each variable $A \in \mathsf{Var}$ is associated with a subset of Val, called the *domain* of A, denoted $\mathrm{Dom}(A)$. For a vector $\boldsymbol{X} = (A_1,\dots,A_n)$ of variables, we write $\mathrm{Dom}(\boldsymbol{X})$ for the Cartesian product $\mathrm{Dom}(A_1) \times \cdots \times \mathrm{Dom}(A_n)$. Given a finite set of variables $\boldsymbol{X}$, an $\boldsymbol{X}$*-tuple* is a mapping $t : \boldsymbol{X} \to \mathsf{Val}$ such that $t(A) \in \mathrm{Dom}(A)$. We write $\mathrm{Tup}(\boldsymbol{X})$ for the set of all $\boldsymbol{X}$-tuples. For $\boldsymbol{Y} \subseteq \boldsymbol{X}$, the *projection* $t[\boldsymbol{Y}]$ *of* t *on* $\boldsymbol{Y}$ is the unique $\boldsymbol{Y}$-tuple that agrees with t on $\boldsymbol{X}$. A *relation* R over $\boldsymbol{X}$ is a subset of $\mathrm{Tup}(\boldsymbol{X})$. The variable set $\boldsymbol{X}$ is also called *(relation) schema of* R. We sometimes write $R(\boldsymbol{X})$ instead of R to emphasize that $\boldsymbol{X}$ is the schema of R. For $\boldsymbol{Y} \subseteq \boldsymbol{X}$, the *projection* of R on $\boldsymbol{Y}$, written $R[\boldsymbol{Y}]$, is the set of all projections $t[\boldsymbol{Y}]$ where $t \in R$. A *database* is a finite collection of relations $D = \{R_1^D(\boldsymbol{X}_1),\dots,R_n^D(\boldsymbol{X}_n)\}$. We assume in this paper that each relation is finite.

2.2 Information theory

Let X be a random variable associated with a finite domain $D = \mathrm{Dom}(X)$ and a probability distribution $p : D \to [0, 1]$, where $\sum_{a \in D} p(a) = 1$. The *entropy* of X is defined as

$$H(X) := -\sum_{x \in D} p(x) \log p(x). \tag{1}$$

Entropy is non-negative and does not exceed the logarithm of the domain size: $0 \le H(X) \le \log |D|$. In particular, $H(X) = 0$ if and only if X is constant (i.e., $p(a) = 1$ for some $a \in D$), and $H(X) = \log |D|$ if and only if X is uniformly distributed (i.e., $p(a) = 1/|D|$ for all $a \in D$).

Fix $n \ge 1$ and consider a set of variables $\boldsymbol{X} = \{X_1, \dots, X_n\}$. We will use α for subsets of $[n]$, and write $\boldsymbol{X}_\alpha := \{X_i \mid i \in \alpha\}$. In the following, we list some common classes of vectors $\boldsymbol{h} = (h_\alpha)_{\alpha \subseteq [n]} \in \mathbb{R}^{2^n}$. Note that such vectors $\boldsymbol{h}$ can alternatively be conceived as functions from $\mathcal{P}(\boldsymbol{X})$ to $\mathbb{R}$, called *set functions*. Hence we often write $h(\boldsymbol{X}_\alpha)$ to denote the element h_α of $\boldsymbol{h}$, and from now on refer to $\boldsymbol{h}$ as a *function*. We assume $h(\emptyset) = 0$ for all functions $\boldsymbol{h}$. For a list of functions $\boldsymbol{h}_1, \dots, \boldsymbol{h}_n$, the function $c_1\boldsymbol{h}_1 + \dots + c_n\boldsymbol{h}$ is called a *positive combination* (resp. *non-negative combination*) of $\boldsymbol{h}_1, \dots, \boldsymbol{h}_n$ if $c_i > 0$ (resp. $c_i \ge 0$) for all $i \in [n]$.

Polymatroids. If $\boldsymbol{h}$ satisfies the *polymatroidal axioms* (Fig. 1), it is called a *polymatroid*. The set of polymatroids over n is denoted Γ_n.

1. $h(\emptyset) = 0$
2. $h(\boldsymbol{X} \cup \boldsymbol{Y}) \ge h(\boldsymbol{X})$ (monotonicity)
3. $h(\boldsymbol{X}) + h(\boldsymbol{Y}) \ge h(\boldsymbol{X} \cap \boldsymbol{Y}) + h(\boldsymbol{X} \cup \boldsymbol{Y})$ (submodularity)

Figure 1 Polymatroidal axioms.

Monotone functions. If $\boldsymbol{h}$ satisfies the first two axioms of the polymatroidal axioms, we call it a *monotone function*, and denote the set of monotone functions over n by Mon_n.

Entropic functions. Consider a relation R over a set $\boldsymbol{X} = \{X_i\}_{i=1}^n$ of variables with finite domains, associated with a probability distribution $p : R \to [0, 1]$. Each subset $\boldsymbol{Y} \subseteq \boldsymbol{X}$ can be viewed as a random variable with domain $D = R[\boldsymbol{Y}]$ and probability distribution $p_{\boldsymbol{Y}}(t) = \sum_{t' \in R, t'[\boldsymbol{Y}] = t} p(t')$. In particular, the subset $\boldsymbol{Y}$ is thus associated with an entropy $H(\boldsymbol{Y})$. The function $\boldsymbol{h} = (H(\boldsymbol{X}_\alpha))_{\alpha \subseteq [n]}$ arising from p in this way is called an *entropic function*. Each entropic function is a polymatroid, but in the converse direction there are polymatroids which are not entropic functions. In general entropic functions satisfy many additional constraints which do not follow by the polymatroidal axioms alone. However, it is not known whether there exists any effective procedure to check that a given function is entropic. The *entropic region* $\Gamma_n^* \subseteq \mathbb{R}^{2^n}$ consists of all entropic functions over n. The *almost entropic region* $\overline{\Gamma_n^*}$ is defined as the topological closure of Γ_n^*.

Normal polymatroids and step functions. For $\boldsymbol{U} \subseteq \boldsymbol{X}$, the function

$$s_{\boldsymbol{U}}(\boldsymbol{W}) = \begin{cases} 0 & \text{if } \boldsymbol{W} \subseteq \boldsymbol{U}; \\ 1 & \text{otherwise;} \end{cases} \tag{2}$$

is called a *step function*. We also use the notation $s^{\boldsymbol{V}}$ to denote the step function $s_{\boldsymbol{X}\setminus\boldsymbol{V}}$. Note that the step function $s^{\boldsymbol{V}}$ is the entropic function arising from the uniform distribution of two tuples t and t' such that $t(A) \neq t'(A)$ if and only if $A \in \boldsymbol{V}$. The set of all step functions over n is denoted S_n. A *normal polymatroid* is a positive combination of step functions, and the set of all normal polymatroids over n is denoted N_n.

Modular functions. A polymatroid $\boldsymbol{h}$ is called *modular* if the submodularity inequality (Fig. 1) is an equality: $h(\boldsymbol{X}) + h(\boldsymbol{Y}) = h(\boldsymbol{X} \cap \boldsymbol{Y}) + h(\boldsymbol{X} \cup \boldsymbol{Y})$. Alternatively, a function $\boldsymbol{h}$ is modular if it is non-negative and such that $h(\boldsymbol{X}) = \sum_{Y \in \boldsymbol{X}} h(Y)$. The set of modular functions over n is denoted Mod_n. Modular functions can alternatively be defined in terms of *basic modular functions*, which are step functions $s^{\{A\}}$ defined in terms of singleton sets $\{A\}$. We denote the set of all basic modular functions over n by B_n. A function $\boldsymbol{f}$ is a modular function if and only if it is a positive combination of basic modular functions.

By continuity of Eq. (1), and since there are no restrictions on domain sizes, $c\boldsymbol{h}$ is entropic for $c > 0$ and step functions $\boldsymbol{h}$. Furthermore, if $\boldsymbol{h}$ and $\boldsymbol{h}'$ are entropic functions defined by probability distributions p and p' over some relation R, the distribution $p''(t \otimes t') := p(t)p'(t')$ on the direct product $\{t \otimes t' \mid t, t' \in R\}$, $(t \otimes t')(X) := (t(X), t'(X))$, defines $\boldsymbol{h} + \boldsymbol{h}'$. This shows that the entropic region is closed under multiplication by positive integers, even though in general it is not closed under positive scalar multiplication [27]; in other words, Γ_n^* is not a *cone*. We conclude that both modular and normal polymatroids are entropic. In fact, Kenig and Suciu [13] have shown that the normal polymatroids are exactly those entropic functions that have a non-negative I-measure [27]. The introduced set functions are related to one another in the following way:

$$\begin{array}{ccccccccccc}
\mathrm{B}_n & \subseteq & \mathrm{S}_n & & & & & & & & \\
\cap\!\!| & & \cap\!\!| & & & & & & & & \\
\mathrm{Mod}_n & \subseteq & \mathrm{N}_n & \subseteq & \Gamma_n^* & \subseteq & \overline{\Gamma_n^*} & \subseteq & \Gamma_n & \subseteq & \mathrm{Mon}_n.
\end{array}$$

If $n \geq 4$, then all the above subset relations are strict.

We will repeatedly refer to the following Shannon's information measures (over some $\boldsymbol{h}$).

- Conditional entropy: $h(\boldsymbol{Y} \mid \boldsymbol{X}) := h(\boldsymbol{XY}) - h(\boldsymbol{X})$.
- Mutual information: $I_{\boldsymbol{h}}(\boldsymbol{X};\boldsymbol{Y}) := h(\boldsymbol{X}) + h(\boldsymbol{Y}) - h(\boldsymbol{XY})$.
- Conditional mutual information: $I_{\boldsymbol{h}}(\boldsymbol{Y};\boldsymbol{Z} \mid \boldsymbol{X}) := h(\boldsymbol{XY}) + h(\boldsymbol{XZ}) - h(\boldsymbol{X}) - h(\boldsymbol{XYZ})$.

We may drop the subscript $\boldsymbol{h}$ if it is clear from the context.

An *information inequality* is an expression ϕ of the form

$$c_1 h(\boldsymbol{X}_1) + \cdots + c_k h(\boldsymbol{X}_k) \geq 0, \tag{3}$$

where $c_i \in \mathbb{R}$, and $\boldsymbol{X}_i$ are sets of variables from $\{X_j\}_{j=1}^n$. We sometimes write $\phi(\boldsymbol{X})$ instead of ϕ to emphasize that the set of variables appearing in ϕ is $\boldsymbol{X}$. For $V \subseteq \mathbb{R}^{2^n}$, we say that ϕ is *valid* over V, denoted $V \models \phi$, if it holds true for all functions $\boldsymbol{h} \in V$.

▶ **Example 1.** Suppose X and Y are independent and uniformly either 0 or 1, and let $Z = X + Y \pmod 2$. This joint distribution can be constructed by taking the uniform distribution over the relation R in Tab. 1. Let $\boldsymbol{h}$ be the entropic function arising from this distribution. Let ϕ be an information inequality of the form $I_{\boldsymbol{h}}(X, Y, Z) \geq 0$ where

$$I_{\boldsymbol{h}}(X,Y,Z) := h(XYZ) - h(XY) - h(XZ) - h(YZ) + h(X) + h(Y) + h(Z) \tag{4}$$

is the mutual information of variables X, Y, Z. We observe that ϕ is not true for $\boldsymbol{h}$ because Eq. (4) evaluates to -1. In particular, this means that $\Gamma_3^* \not\models \phi$. On the other hand, ϕ is true if we interpret $\boldsymbol{h}$ as any step function $s_{\boldsymbol{U}}$, $\boldsymbol{U} \subseteq \{X, Y, Z\}$. Since normal polymatroids are positive combinations of step functions, this entails $\mathrm{N}_3 \models \phi$.

Table 1 The relation R representing $X + Y \equiv Z \pmod 2$.

R	X	Y	Z
	0	0	0
	0	1	1
	1	0	1
	1	1	0

This paper focuses on the *information inequality problem* (IIP), introduced in [14], which is to decide whether a given information inequality is valid over Γ_n^*. This problem is co-r.e. [14], as the continuity of the entropy (1) and the density of the rationals in the reals imply that enumeration of all rational distributions will eventually lead to a counterexample of (3), if one exists at all. We introduce the following relativized version of IIP. Fixing sets of functions $F_n \subseteq \mathbb{R}^{2^n}$, $n \geq 1$, and a set $\mathcal{C}$ of information inequalities, the *information inequality problem over F_n w.r.t. $\mathcal{C}$* ($\mathsf{IIP}_{F_n}(\mathcal{C})$) is to determine whether a given information inequality $\phi \in \mathcal{C}$ over n variables is valid over F_n. We leave out F_n (resp. $\mathcal{C}$) if $F_n = \Gamma_n^*$ (resp. $\mathcal{C}$ contains all information inequalities). Note that an inequality ϕ is valid over the entropic region Γ_n^* if and only if it is valid over the almost entropic region $\overline{\Gamma_n^*}$. To see why, $V \models \phi$ is tantamount to $V \subseteq C_\phi$, where $C_\phi = \{\boldsymbol{h} \in \mathbb{R}^{2^n} \mid \boldsymbol{h} \models \phi\}$, and by taking closures on both sides, $\Gamma_n^* \subseteq C_\phi$ entails $\overline{\Gamma_n^*} \subseteq C_\phi$. More generally, validity over Γ_n^* and $\overline{\Gamma_n^*}$ disagrees with respect to Boolean combinations of information inequalities [12, 14]. Since our focus is on the information inequality problem alone, we now drop the almost entropic region $\overline{\Gamma_n^*}$ from discussions.

Before proceeding, we shortly discuss input representation. We assume that the coefficients are rational. Note that in [14] the inputs of IIP and other related problems are vectors $\boldsymbol{c} \in \mathbb{Z}^{2^n}$ representing the coefficients in Eq. (3). In this paper, we consider the input as a sequence $((c_1, \boldsymbol{X}_1), \ldots, (c_k, \boldsymbol{X}_k))$, which is potentially exponentially shorter than the aforementioned coefficient vector $\boldsymbol{c}$. This distinction is not important if one is solely interested in decidability, as is the case in [14]. Since our aim is to chart the tractability boundary for different information inequality problems, we opt for the latter more concise representation. Furthermore, we assume that the coefficients themselves are encoded in binary.

We begin our analysis from intractable examples, and then move on to discuss tractable cases and their connections to query output bounds.

3 Intractable cases

Kenig and Suciu [13] establish an interesting connection between information inequalities and the implication problem for database dependencies. Fix a relation schema $\boldsymbol{X}$ of n variables. An expression of the form $\sigma = (\boldsymbol{V}; \boldsymbol{W} \mid \boldsymbol{U})$ is called a *conditional independence* (CI). If $\boldsymbol{UVW} = \boldsymbol{X}$, σ is specifically called a *saturated conditional independence* (SCI), and if $\boldsymbol{V} = \boldsymbol{W}$, it is called a *conditional* and shortened as $(\boldsymbol{V} \mid \boldsymbol{U})$. Lee [19] observed that an SCI of the form $(\boldsymbol{V}; \boldsymbol{W} \mid \boldsymbol{U})$ holds true on the uniform distribution of a database relation $R(\boldsymbol{UVW})$ if and only if R satisfies the corresponding *multivalued dependency* (MVD) $\boldsymbol{U} \twoheadrightarrow \boldsymbol{V}$. An analogous correspondence can be drawn between a conditional $(\boldsymbol{V} \mid \boldsymbol{U})$ and the *functional dependency* (FD) $\boldsymbol{U} \to \boldsymbol{V}$. The results in [13] entail that if Σ is a set of SCIs and conditionals, and τ is a conditional, then for any V such that $\mathrm{N}_n \subseteq V \subseteq \Gamma_n$,

$$V \models \sum_{\sigma \in \Sigma} h(\sigma) \geq h(\tau) \iff \Sigma \models \tau, \tag{5}$$

where the right-hand side denotes implication between the corresponding MVDs and FDs over database relations. Whether or not $\sum_{\sigma \in \Sigma} h(\sigma) \geq h(\tau)$ is valid over V can be thus decided in polynomial time, because the implication problem for MVDs and FDs is known to be in polynomial time [3].

There are at least two ways to make the inequality in Eq. (5) harder. One possibility is to allow more complex information measures, much like how one can allow more expressive database dependencies in the implication problem. For instance, once the aforementioned syntactic restrictions are lifted, the implication problem for CIs becomes undecidable in both database theory (where CIs are known as embedded multivalued dependencies) and probability theory [10, 18, 21]. Another possibility, which does not seem to have a counterpart in the implication problem, is to permit coefficients distinct from 1. Next, we consider both of these strategies in isolation, considering first complex information measures.

The mutual information of two random variables generalizes to the multivariate mutual information over a set of random variables $\boldsymbol{S}$. For a general set function $\boldsymbol{h}$, the *multivariate mutual information* is given as

$$I_h(\boldsymbol{S}) = \sum_{\boldsymbol{T} \subseteq \boldsymbol{S}} (-1)^{|\boldsymbol{T}|-1} h(\boldsymbol{T}). \tag{6}$$

Again, we drop the subscript whenever this is possible without confusion. A particular case of the multivariate mutual information is the three-variate one, presented in Eq. (4). Multivariate mutual information is non-negative on step functions, but, as discussed in Example 1, it can be negative on entropic functions. Next we show that solving inequalities containing three-variate mutual informations and conditional entropies can already be coNP-hard, even if each coefficient is exactly one. The result, proven by a reduction from monotone satisfiability, holds for step functions but does not extend to entropic functions.

A conjunctive normal form Boolean formula ϕ is called *monotone* if each clause in ϕ contains only negative or only positive literals. The *monotone satisfiability problem* is the problem of deciding whether such a formula ϕ has a satisfying truth assignment. This problem is well known to be NP-complete [5], and it remains NP-complete even if each clause consists of exactly three distinct literals [22]. Let us denote this restriction of the problem by 3DMONSAT. An instance of 3DMONSAT can be represented as a pair $\phi = (\phi^+, \phi^-)$, where ϕ^+ (resp. ϕ^-) is the set of all positive (resp. negative) clauses of ϕ, and each clause is a set of exactly 3 variables.

▶ **Theorem 2.** *The information inequality problem over normal polymatroids is* coNP-*complete.*

Proof. Since normal polymatroids are positive combinations of step functions, and inequalities are preserved under positive combinations, the information inequality problems over step functions and normal polymatroids coincide. The upper bound is thus obvious. For the lower bound, we present a reduction from the complement of 3DMONSAT to the information inequality problem over step functions. Let $\phi = (\phi^+, \phi^-)$ be an instance of 3DMONSAT. Suppose $\boldsymbol{X}$ is the set of variables appearing in ϕ. We may assume without loss of generality that every satisfying assignment must map at least one variable to 1.

Define an information inequality

$$\sum_{\boldsymbol{C} \in \phi^+} h(\boldsymbol{X} \mid \boldsymbol{C}) + \sum_{\boldsymbol{C} \in \phi^-} I(\boldsymbol{C}) \geq h(\boldsymbol{X}), \tag{7}$$

where $I(\boldsymbol{C})$ is the three-variate mutual information (4) over the variables of $\boldsymbol{C}$, and $h(\boldsymbol{X} \mid \boldsymbol{C})$ is the conditional entropy of $\boldsymbol{X}$ given $\boldsymbol{C}$.

Each subset $\boldsymbol{Y} \subseteq \boldsymbol{X}$ determines a unique step function $s^{\boldsymbol{Y}}$ (Eq. 2) and a unique Boolean assignment

$$m_{\boldsymbol{Y}}(A) = \begin{cases} 1 & \text{if } A \in \boldsymbol{Y}, \\ 0 & \text{otherwise.} \end{cases}$$

We claim that $m_{\boldsymbol{Y}}$ satisfies ϕ if and only if Eq. (7) is false for $h = s^{\boldsymbol{Y}}$.

Assume first that $m_{\boldsymbol{Y}}$ satisfies ϕ. By our assumption some variable is mapped to 1, which means that $\boldsymbol{Y}$ is non-empty. In particular, $s^{\boldsymbol{Y}}(\boldsymbol{X}) = 1$. For any positive clause $\boldsymbol{C} \in \phi^+$, we have $\boldsymbol{C} \cap \boldsymbol{Y} \neq \emptyset$, and consequently $s^{\boldsymbol{Y}}(\boldsymbol{X} \mid \boldsymbol{C}) = s^{\boldsymbol{Y}}(\boldsymbol{X}) - s^{\boldsymbol{Y}}(\boldsymbol{C}) = 0$. For any negative clause $\boldsymbol{C} \in \phi^-$, we have $\boldsymbol{C} \not\subseteq \boldsymbol{Y}$, in which case it is straightforward to verify that $I(\boldsymbol{C}) = 0$. We conclude that Eq. (7) is false for $h = s^{\boldsymbol{Y}}$.

Assume then that $m_{\boldsymbol{Y}}$ does not satisfy ϕ. If $s^{\boldsymbol{Y}}(\boldsymbol{X}) = 0$, then Eq. (7) is trivially true for $h = s^{\boldsymbol{Y}}$ by the non-negativity of the conditional entropy and the multivariate mutual information on step functions. Suppose then $s^{\boldsymbol{Y}}(\boldsymbol{X}) \neq 0$, in which case $s^{\boldsymbol{Y}}(\boldsymbol{X}) = 1$. Assuming $m_{\boldsymbol{Y}}$ does not satisfy some $\boldsymbol{C} \in \phi^+$, we have $\boldsymbol{C} \cap \boldsymbol{Y} = \emptyset$ implying $s^{\boldsymbol{Y}}(\boldsymbol{X} \mid \boldsymbol{C}) = 0$. Assuming $m_{\boldsymbol{Y}}$ does not satisfy some $\boldsymbol{C} \in \phi^-$, we have $\boldsymbol{C} \subseteq \boldsymbol{Y}$ implying $I(\boldsymbol{C}) = 1$. We conclude that, for $h = s^{\boldsymbol{Y}}$, the left-hand side of Eq. (7) is at least 1, and thus the inequality holds. This concludes the proof of the claim.

The claim implies that ϕ is not satisfiable if and only if Eq. (7) is valid over step functions. The theorem statement follows, since the reduction is clearly polynomial. ◀

Observe that Eq. (7) is syntactically similar to the inequality in Eq. (5) in that each coefficient is exactly one. The difference comes from allowing three-variate mutual information, whereas the inequality in Eq. (5) allows only specific forms of conditional mutual information. The above proof moreover establishes *strong* coNP-completeness, because the problem remains coNP-complete even under unary encoding of the coefficients.

Alternatively, the preceding theorem can be proven by reducing 3-colorability to inequalities that allow the coefficients to grow while using only conditionals. Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a graph consisting of a vertex set $\boldsymbol{V}$ and a set of undirected edges $\boldsymbol{E}$. For each node $A \in \boldsymbol{V}$, introduce variables A_r, A_g, A_b representing possible colors of A. Assume that the graph contains n vertices. We define

$$\sum_{\substack{c \in \{r,g,b\} \\ A \in \boldsymbol{V}}} h(A_c) + \sum_{\substack{c,d \in \{r,g,b\} \\ c \neq d \\ A \in \boldsymbol{V}}} (2n+1) h(\boldsymbol{V} \mid A_c A_d) + \sum_{\substack{c \in \{r,g,b\} \\ \{A,B\} \in \boldsymbol{E}}} (2n+1) h(\boldsymbol{V} \mid A_c B_c) \geq (2n+1) h(\boldsymbol{V}). \tag{8}$$

Then G is three-colorable if and only if Eq. (8) is not valid over step functions (Appendix A). This way of proving Theorem 2 establishes also strong coNP-completeness, since each coefficient is bounded by a polynomial in the input size. It is necessary to allow coefficients other than 1 in Eq. (8). Otherwise, the equivalence (5) holds, meaning that the validity problem is equivalent to the implication problem for FDs, which is in polynomial time.

▶ **Example 3.** Eq. (8) behaves differently for step functions and entropic functions, even though both functions are non-negative on all the occurring information measures; in contrast, the proof of Theorem 2 relied on three-variate mutual information which is only guaranteed to be non-negative for step functions but can be negative for entropic functions. For a concrete example, suppose G is the complete graph of four vertices A, B, C, D. Since G is not three-colorable, Eq. (8) is valid over step functions. For the entropic function arising from the uniform distribution of Tab. 2, however, Eq. (8) is false.

Table 2 Three-tuple counterexample.

A_r	B_g	C_b	A_g	B_b	C_r	A_b	B_r	C_g	D_r	D_g	D_b
0	0	0	0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0	1	1	1
1	1	1	0	0	0	0	0	0	2	2	2

4 Tractable cases

We have seen that intractable inequalities arise from (i) complex information measures even if coefficients are restricted to 1, (ii) more simple information measures if coefficients are allowed to grow, and (iii) inequalities where the negative coefficients are associated with sets of size at most two. In this section we consider restrictions that give rise to inequalities solvable in polynomial time. We are specifically interested in inequalities of the form $\sum_{\sigma\in\Sigma} w_i h(\sigma) \geq h(\boldsymbol{X})$ where $w_i \geq 0$, and Σ is a set of conditionals. Such inequalities make appearance when information theory is applied to obtain tight upper bounds for query output sizes. Since inequalities of the form (8) are intractable, imposing syntactic restrictions on Σ becomes necessary.

We next introduce information-theoretic query upper bounds, after which we move on to discuss the complexity of related syntactic restrictions and semantic modifications of the information inequality problem.

4.1 Query upper bounds

Fix a relation R over a variable set $\boldsymbol{X}$. Given vectors $\boldsymbol{U}, \boldsymbol{V}$ of variables from $\boldsymbol{X}$, and values $\boldsymbol{u} \in \mathrm{Dom}(\boldsymbol{U})$, the *$V$-degree* of $\boldsymbol{U} = \boldsymbol{u}$ in R, denoted $\deg_R(\boldsymbol{V} \mid \boldsymbol{U} = \boldsymbol{u})$, is the number of distinct values of $\boldsymbol{V}$ that occur in R together with the value $\boldsymbol{u}$ of $\boldsymbol{U}$. The *max-$\boldsymbol{V}$-degree* of $\boldsymbol{U}$, denoted $\deg_R(\boldsymbol{V} \mid \boldsymbol{U})$ is the maximum V-degree of $\boldsymbol{U} = \boldsymbol{u}$ over all $\boldsymbol{u}$. Expressions of the form $\deg_R(\sigma) \leq B$ (omitting the parentheses of σ), where σ is a conditional and $B \geq 1$, are usually called *degree constraints*. Note that $\deg_R(\boldsymbol{V} \mid \boldsymbol{U}) = 1$ if and only if R satisfies the functional dependency $\boldsymbol{U} \to \boldsymbol{V}$. A *$\Sigma$-inequality* is an information inequality $\phi_\Sigma(\boldsymbol{X}, \boldsymbol{w})$ of the form

$$\sum_{\sigma\in\Sigma} w_\sigma h(\sigma) \geq h(\boldsymbol{X}), \tag{9}$$

where $\boldsymbol{w} = (w_\sigma)_{\sigma\in\Sigma}$ is a sequence of non-negative reals.

Fix a *self-join-free full conjunctive query*, i.e., a quantifier-free first-order formula of the form

$$Q(\boldsymbol{X}) = R_1(\boldsymbol{X}_1) \wedge \cdots \wedge R_n(\boldsymbol{X}_n),$$

where $R_i(\boldsymbol{X}_i)$ are *relational atoms* over distinct *relation names* R_i, and variable sequences $\boldsymbol{X}_i$ such that $\boldsymbol{X}$ lists all the variables occurring in them. Note that this incurs a slight abuse of notation, because $R_i(\boldsymbol{X}_i)$ could also refer to a relation R_i over $\boldsymbol{X}_i$. We also blur the distinction between a set and a sequence of variables $\boldsymbol{X}_i$, and say that a set of conditionals Σ is *guarded* by Q if every $\sigma = (\boldsymbol{V} \mid \boldsymbol{U})$ from Σ is associated with a relation name R_i, called the *guard* of σ and denoted R_σ, such that $\boldsymbol{U}\boldsymbol{V} \subseteq \boldsymbol{X}_i$. A sequence of the form $\boldsymbol{B} = (B_\sigma)_{\sigma\in\Sigma}$, $B_\sigma \geq 1$, form the *degree values* associated with Σ. A database D containing relations R_σ, $\sigma \in \Sigma$, *satisfies* a conditionals-values pair $(\Sigma, \boldsymbol{B})$, written $D \models (\Sigma, \boldsymbol{B})$, if $\deg_{R_\sigma}(\sigma) \leq B_\sigma$ for all $\sigma \in \Sigma$.

For a set $S \subseteq \mathbb{R}^{2^n}$, and a set of conditionals Σ guarded by $Q(\boldsymbol{X})$ and associated with values $\boldsymbol{B}$, define the *bound* of Q w.r.t. $\Sigma, S, \boldsymbol{B}$ as

$$\mathrm{Bound}_S(Q, \Sigma, \boldsymbol{B}) \coloneqq \inf_{\substack{\boldsymbol{w} \geq 0 \\ \mathsf{S} \models \phi_\Sigma(\boldsymbol{X}, \boldsymbol{w})}} \prod_{\sigma \in \Sigma} B_\sigma^{w_\sigma}.$$

The bounds $\mathrm{Bound}_{\mathrm{Mod}_n}, \mathrm{Bound}_{\mathrm{N}_n}, \mathrm{Bound}_{\Gamma_n^*}, \mathrm{Bound}_{\Gamma_n}$ are often referred to as the *modular bound*, the *coverage bound*, the *entropic bound*, and the *polymatroid bound*. Writing $Q(D)$ for the output of Q on a database $D = \{R_1^D(\boldsymbol{X}_1), \ldots, R_n^D(\boldsymbol{X}_n)\}$, one can prove that the entropic bound is valid: $|Q(D)| \leq \mathrm{Bound}_{\Gamma_n^*}(Q, \Sigma, \boldsymbol{B})$ whenever $D \models (\Sigma, \boldsymbol{B})$. Since $\Gamma_n^* \subseteq \Gamma_n$, the entropic bound is less than or equal to the polymatroid bound. The entropic bound is asymptotically tight, but it is open whether or not the bound is computable. The polymatroid bound can be attained by solving a linear program of exponential size, but it is not tight. For a derivation of the entropic bound and a discussion on the asymptotic tightness (or lack thereof) of these bounds, we refer the reader to [26].

Fortunately, there are well-behaving syntactic restrictions for sets of conditionals Σ, some of which are presented next.

- For $\sigma = (\boldsymbol{V} \mid \boldsymbol{U})$, where $\boldsymbol{U} = \emptyset$, degree constraints of the form $\deg_R(\sigma) \leq B$ are called *cardinality constraints*. The AGM bound [2] can be viewed as the entropic bound over a specific set of cardinality constraints.
- Σ is called *acyclic* if the following directed graph is acyclic: the vertices are the variables in $\boldsymbol{X}$, and there is an edge from A to B if $A \in \boldsymbol{X}$ and $B \in \boldsymbol{Y} \setminus \boldsymbol{X}$, for some $(\boldsymbol{Y} \mid \boldsymbol{X}) \in \Sigma$.
- Σ is called *simple* if $|\boldsymbol{U}| \leq 1$ for each $(\boldsymbol{V} \mid \boldsymbol{U}) \in \Sigma$.

The entropic bound is polynomial-time computable in all of these cases. Let us call the Σ-inequality (9) *acyclic* (resp. *simple*) if the underlying set Σ is acyclic (resp. simple). The sets of conditionals underlying cardinality constraints are vacuously acyclic, and validity for acyclic Σ-inequalities coincides for modular functions, entropic functions, and polymatroids [23]. Consequently, the entropic bound becomes computable in polynomial time through a linear program describing the validity of Eq. (9) over basic modular functions. Validity for simple Σ-inequalities similarly coincides for entropic functions, polymatroids, and normal polymatroids. This does not immediately entail that the entropic bound for simple Σ is computable in polynomial time, because normal polymatroids are constructed with step functions, and there are exponentially many step functions in the number of variables. The entropic bound is nevertheless known to be polynomial-time computable in this case, as has been shown recently [11].

Eq. (8) can now be viewed as an Σ-inequality (9) (up to scaling) arising from Σ that does not belong to any of the aforementioned well-behaving classes. Since validity of inequalities of the form Eq. (8) is coNP-hard over step functions (Appendix A), this immediately gives us the following result.

▶ **Theorem 4.** *The information inequality problem over normal polymatroids w.r.t. Σ-inequalities is* coNP*-complete. This problem remains* coNP*-hard even if $|\boldsymbol{U}| \leq 2$ for all $(\boldsymbol{V} \mid \boldsymbol{U}) \in \Sigma$.*

Related to the previous result, computing the coverage bound over a set of conditionals is known to be NP-hard, and computing the polymatroid bound over an arbitrary set of conditionals can be efficiently reduced to computing the polymatroid bound over another set of conditionals $(\boldsymbol{V} \mid \boldsymbol{U})$ such that $|\boldsymbol{U}| \leq 2$ and $|\boldsymbol{V}| \leq 3$ [11].

We now turn to discuss tractable cases of the information inequality problem obtained either by syntactic restrictions or semantic modifications. The syntactic restrictions we consider correspond quite closely to the aforementioned acyclic/simple Σ-inequalities.

4.2 Modular functions

Since modular functions can be constructed as positive combinations of basic modular functions, the information inequality problem is trivially polynomial-time computable in this context.

▶ **Proposition 5.** *The information inequality problem over modular functions is in polynomial time.*

One example of a syntactic class with respect to which validity over entropic functions corresponds to validity over modular functions are the acyclic Σ-inequalities. Given an acyclic set Σ of conditionals $(\boldsymbol{V} \mid \boldsymbol{U})$, and a polymatroid $\boldsymbol{h}$ over $\boldsymbol{X}$, one can construct a modular function $\boldsymbol{f}$ such that (i) $f(\boldsymbol{X}) = h(\boldsymbol{X})$, and (ii) $f(\boldsymbol{V} \mid \boldsymbol{U}) \leq h(\boldsymbol{V} \mid \boldsymbol{U})$ for all $(\boldsymbol{V} \mid \boldsymbol{U}) \in \Sigma$ [23]. Consequently, validity of acyclic Σ-inequalities (9) coincides for polymatroids, entropic functions, normal polymatroids, and modular functions. Thus it is known that all the aforementioned bounds (modular, coverage, entropic, and polymatroid bounds) coincide and are polynomial-time computable if Σ is acyclic. With respect to the information inequality problem, we analogously obtain the following result.

▶ **Proposition 6.** *Let* $\mathrm{Mod}_n \subseteq K \subseteq \Gamma_n$. *The information inequality problem over* K *w.r.t. acyclic* Σ*-inequalities is in polynomial time.*

4.3 Monotone functions

Next we will show that, at the other extreme direction, the information inequality problem over monotone functions is also in polynomial time.

▶ **Theorem 7.** *The information inequality problem over monotone functions is in polynomial time.*

Analogously to the previous section, this semantic modification of the information inequality problem helps us identify syntactic classes with respect to which the general information inequality problem is tractable. Before we proceed into details, let us give a short sketch of the proof of this theorem. We associate an information inequality (3) with a *set representation* (S^+, S^-), where

$$S^+ := \{(\boldsymbol{X}_i, k) \mid k \in [|c_i|], c_i > 0\}, \text{ and}$$
$$S^- := \{(\boldsymbol{X}_i, k) \mid k \in [|c_i|], c_i < 0\}.$$

Then, we present a fixed-point algorithm (Alg. 1) to capture validity of information inequalities over monotone functions. This algorithm iteratively decomposes an input inequality into monotonicity axioms. For this, it maintains a bipartite directed graph G initialized as $G_S = (S^+ \cup S^-, E)$, where E is the set of edges from S^+ to S^- that correspond to possible monotonicity axioms (*forward edges*). The initial graph contains no edges from S^- to S^+ (*backward edges*). The number of these backward edges, which represent those monotonicity axioms that are currently selected for the decomposition, is increased in each iteration. Although the algorithm as such does not run in polynomial time (it runs in *pseudo-polynomial time*, i.e., in polynomial time in the length of the input and the numeric values of the coefficients), it does guide us toward a characterization of valid inequalities as positive combinations of monotonicity axioms and *non-negativity axioms* $h(\boldsymbol{X}) \geq 0$, which are derivable as combinations of the first two polymatroidal axioms. These combinations are polynomial in the input length, and consequently can be found through a linear program of polynomial size, which entails the desired result.

Algorithm 1 Decomposition algorithm for inequalities.

Input: Set representation $S = (S^+, S^-)$ of ϕ

Output: true iff ϕ is $\mathcal{M}_n$-valid

$G \leftarrow G_S, S_0 \leftarrow S^+, S_1 \leftarrow S^-$

while G contains a path $u_0, \dots, u_m$ from S_0 to S_1 **do**

  remove u_0 from S_0 and u_m from S_1

  remove from G (backward) edges $(u_1, u_2), (u_3, u_4), \dots, (u_{m-2}, u_{m-1})$

  add to G (backward) edges $(u_1, u_0), (u_3, u_2), \dots, (u_m, u_{m-1})$

return true if S_1 is empty, otherwise false

The following example demonstrates the use of Alg. 1.

▶ **Example 8.** Consider an information inequality of the form

$$\boldsymbol{XY} + \boldsymbol{YZ} + 2\boldsymbol{XZ} + \boldsymbol{X} \geq \boldsymbol{Y} + 3\boldsymbol{Z}. \tag{10}$$

The set representation is (S^+, S^-), where

$$S^+ = \{(\boldsymbol{XY}, 1), (\boldsymbol{YZ}, 1), (\boldsymbol{XZ}, 1), (\boldsymbol{XZ}, 2), (\boldsymbol{X}, 1)\}, \text{ and}$$
$$S^- = \{(\boldsymbol{Y}, 1), (\boldsymbol{Z}, 1), (\boldsymbol{Z}, 2), (\boldsymbol{Z}, 3)\}.$$

Clearly, the inequality (10) is valid over monotone functions. Alg. 1 also returns true after four iterations. The leftmost graph in Fig. 2 illustrates the starting point for the last iteration in one possible implementation. The edges from right to left (backward edges) represent monotonicity axioms that have been selected in the previous iteration. The edges from left to right (forward edges), some of which are visible in the middle graph of Fig. 2, represent possible monotonicity axioms. Since there is a path from S_0 to S_1, we can increase the number of selected monotonicity axioms by deleting the backward edges in the path, and changing the direction of the forward edges in the path. The rightmost graph illustrates the result of this modification. Since S_1 becomes empty, the algorithm terminates returning true. The final state of the algorithm represents an integral decomposition of Eq. (10) into monotonicity and non-negativity axioms.

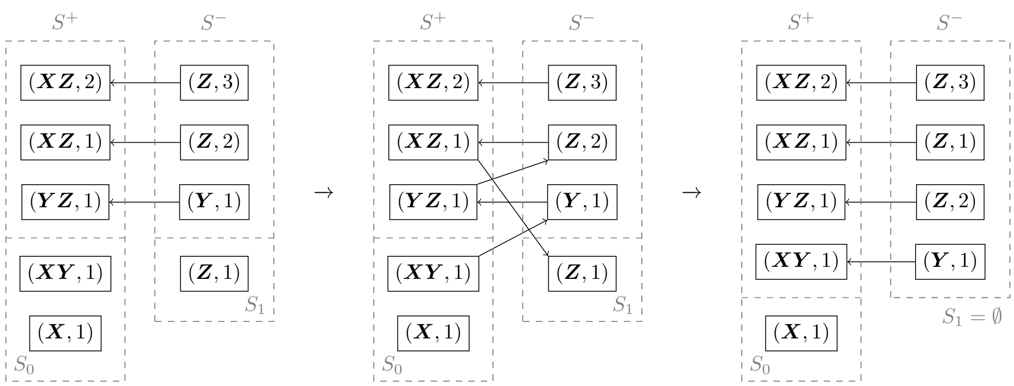

Figure 2 Last iteration of Alg. 1 for Eq. (10).

An inequality ϕ of the form (3) can be identified with its *coefficient function* $\boldsymbol{c}_\phi$, where $c_\phi(\boldsymbol{X}) = c$ if the term $ch(\boldsymbol{X})$ appears in (3), and otherwise $c_\phi(\boldsymbol{X}) = 0$. We then say that ϕ is a (positive) combination of inequalities $\phi_1, \ldots, \phi_n$ if $\boldsymbol{c}_\phi$ is a (positive) combination of $\boldsymbol{c}_{\phi_1}, \ldots, \boldsymbol{c}_{\phi_n}$. We furthermore say that a combination of functions $c_1\boldsymbol{h}_1 + \cdots + c_n\boldsymbol{h}_n$ is *separable* if there exist no i, j and $\boldsymbol{Y}$ such that $h_i(\boldsymbol{Y}) < 0$ and $h_j(\boldsymbol{Y}) > 0$, while $c_i \neq 0 \neq c_j$. This definition is extended to combinations of inequalities in the natural way. For example, any positive combination of $h(A) + h(B) \geq 0$ and $h(B) - h(AB) \geq 0$ is separable, but no positive combination of inequalities $h(A) - h(B) \geq 0$ and $h(A) + h(B) \geq 0$ is separable. In particular, if ϕ is a positive and separable combination of monotonicity and non-negativity axioms, then $h(\boldsymbol{X})$ cannot appear in the left-hand side of ψ_0 and in the right-hand side of ψ_1, for any two axioms ψ_0 and ψ_1 appearing in the combination.

We also say that a set function $\boldsymbol{h}$ is Boolean-valued if it maps every $\boldsymbol{X}$ to either 0 or 1. The proof the following lemma is located in Appendix B.

▶ **Lemma 9.** *Let ϕ be an information inequality of the form*

$$c_1 h(\boldsymbol{X}_1) + \cdots + c_k h(\boldsymbol{X}_k) \geq 0 \quad (c_i \in \mathbb{R}). \tag{11}$$

The following are equivalent:
1. *ϕ is valid over monotone functions.*
2. *ϕ is valid over monotone, Boolean-valued functions.*
3. *ϕ is a positive and separable combination of monotonicity and non-negativity axioms.*

Since linear programming is in polynomial time, we can now establish Theorem 7 as a consequence of the following lemma. This lemma will also be applied in the next section that focuses on simple Σ-inequalities.

▶ **Lemma 10.** *For each information inequality ϕ of the form*

$$c_1 h(\boldsymbol{X}_1) + \cdots + c_k h(\boldsymbol{X}_k) \geq d_1 h(\boldsymbol{Y}_1) + \cdots + d_l h(\boldsymbol{Y}_l) \quad (c_i, d_i > 0), \tag{12}$$

there exists a matrix M such that the inequality $M\boldsymbol{x} \geq \boldsymbol{cd}$ has a solution $\boldsymbol{x} \geq 0$ if and only if ϕ is valid over monotone functions. In particular, M can be constructed in polynomial time from ϕ (with rational coefficients).

Proof. Consider first a set $\boldsymbol{Y}_i$ from the right-hand side of the inequality. Let $\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_m}$ list all those sets $\boldsymbol{X}_j$ from the left-hand side that contain $\boldsymbol{Y}_i$ as a subset. We need to describe how the term $d_i h(\boldsymbol{Y}_i)$ is distributed to monotonicity axioms. For this, define

$$x^i_{i_1} + \cdots + x^i_{i_m} \geq d_i, \tag{13}$$

where x^i_j is a variable denoting the coefficient of the monotonicity axiom $h(\boldsymbol{X}_j) \geq h(\boldsymbol{Y}_i)$. We also need to ensure that this variable does not grow exceedingly large. Consider a set $\boldsymbol{X}_j$ from the left-hand side of the inequality, and let $\boldsymbol{Y}_{j_1}, \ldots, \boldsymbol{Y}_{j_n}$ list all those sets from the right-hand side that are contained in $\boldsymbol{X}_j$.

$$x^{j_1}_j + \cdots + x^{j_n}_j \leq c_j. \tag{14}$$

Combining Eqs. (13) and (14) we obtain an inequality $M\boldsymbol{x} \geq \boldsymbol{cd}$, where M is a $((k+l) \times kl)$-matrix with entries of from $-1, 0, 1$, $\boldsymbol{x}$ is a vector of length kl, and $\boldsymbol{cd}$ (i.e., $\boldsymbol{c}$ and $\boldsymbol{d}$ concatenated) is a vector of length $k + l$. Obviously M can be constructed in polynomial time given ϕ. Moreover, $M\boldsymbol{x} \geq \boldsymbol{cd}$ has a solution $\boldsymbol{x} \geq \boldsymbol{0}$ if and only if ϕ is a positive and separable combination of monotonicity and non-negativity axioms. The statements of the theorem then follow by Lemma 9. ◀

4.4 Simple Σ-inequalities

Let us first recall the reason why normal and general polymatroids are known to agree on the validity of simple Σ-inequalities. On the one hand, any such Σ-inequality over a variable set $\boldsymbol{X}$ can be presented in the form

$$c_1 h(\boldsymbol{X}_1) + \cdots + c_n h(\boldsymbol{X}_n) \geq d_1 h(\boldsymbol{Y}_1) + \cdots + d_m h(\boldsymbol{Y}_m) \quad (c_i, d_i > 0), \tag{15}$$

where each $\boldsymbol{Y}_i$ is either the full set $\boldsymbol{X}$ or some singleton set $\{X\}$. On the other hand, every polymatroid $\boldsymbol{h}$ over $\boldsymbol{X}$ can be associated with a normal polymatroid $\boldsymbol{f}$ over $\boldsymbol{X}$ such that $f(\boldsymbol{Y}) \leq h(\boldsymbol{Y})$ for all $\boldsymbol{Y} \subseteq \boldsymbol{X}$, $f(\boldsymbol{X}) = h(\boldsymbol{X})$, and $f(X) = h(X)$ for all $X \in \boldsymbol{X}$ [26]. Hence, if $\boldsymbol{h}$ is a counterexample for Eq. (15), then $\boldsymbol{f}$ must also be a counterexample. Since normal polymatroids are positive combinations step functions, it follows that at least one step function is also a counterexample. This brings us to the following result.

▶ **Theorem 11** ([26]). *Let $\phi(\boldsymbol{X})$ be an information inequality of the form Eq. (15), where each $\boldsymbol{Y}_i$ is either the full set $\boldsymbol{X}$ or a singleton set. Then, ϕ is valid over step functions if and only if it is valid over polymatroids.*

Since $\mathrm{S}_n \subseteq \mathrm{N}_n \subseteq \Gamma^*_n \subseteq \Gamma_n$, it follows that validity for simple Σ-inequalities coincides for step functions, normal polymatroids, entropic functions, and polymatroids. If we remove terms of the form $h(\boldsymbol{X})$ from the right-hand side of Eq. (15), the previous result extends to monotone functions.

▶ **Theorem 12.** *Let $\phi(\boldsymbol{X})$ be an information inequality of the form Eq. (15), where each $\boldsymbol{Y}_i$ is a singleton set. Then, ϕ is valid over step functions if and only if it is valid over monotone functions.*

Proof. Since step functions are monotone, we only need to consider the "only-if" direction. To show the contraposition, assume that ϕ is not valid over monotone functions. By Lemma 9, we find a monotone, Boolean-valued function h such that Eq. (15) becomes false. Consider the step function $s^{\boldsymbol{U}}$, where $\boldsymbol{U}$ consists of all those variables A_i that are mapped to 1 by h. Clearly, h and $s^{\boldsymbol{U}}$ agree on the right-hand side of Eq. (15). Furthermore, for any set $\boldsymbol{Z}$, we have $s^{\boldsymbol{U}}(\boldsymbol{Z}) \leq h(\boldsymbol{Z})$ by monotonicity of h. Consequently, Eq. (15) is also false for $s^{\boldsymbol{U}}$, meaning that ϕ is not valid over step functions. ◀

It follows that validity for information inequalities of the form (15), where $\boldsymbol{Y}_i$ are singletons, is decidable in polynomial time with respect to any K such that $\mathrm{S}_n \subseteq K \subseteq \mathrm{Mon}_n$, including $K = \Gamma^*_n$. Note that simple Σ-inequalities are not of this form; rewritten in the form (15) one of the sets $\boldsymbol{Y}_i$ is the full variable set. However, as we will see next, it is possible to remove such terms in a single step.

Continuing our analysis of $\phi(\boldsymbol{X})$ of the form (15), fix a variable A from $\boldsymbol{X}$. Define sums

$$c_A = \sum_{\substack{i \in [n] \\ A \in \boldsymbol{X}_i}} c_i \quad \text{and} \quad d_A = \sum_{\substack{i \in [n] \\ A \in \boldsymbol{Y}_i}} d_i,$$

and define the *A-reduction* of ϕ as the inequality $\phi^A(\boldsymbol{X} \setminus \{A\})$ given as

$$(c_A - d_A) h(\boldsymbol{X} \setminus \{A\}) + \sum_{\substack{i \in [n] \\ A \notin \boldsymbol{X}_i}} c_i h(\boldsymbol{X}_i) \geq \sum_{\substack{i \in [n] \\ A \notin \boldsymbol{Y}_i}} d_i h(\boldsymbol{Y}_i), \tag{16}$$

▶ **Lemma 13.** *An information inequality $\phi(\boldsymbol{X})$ of the form* (15) *(having no restrictions on sets $\boldsymbol{Y}_i$) is valid over step functions if and only if for all $A \in \boldsymbol{X}$, the A-reduction ϕ^A of ϕ is valid over step functions.*

1. $c_A \geq d_A$*, and*
2. *the A-reduction $\phi^A(\boldsymbol{X} \setminus \{A\})$ is valid over step functions.*

Proof. Note that $s^{\{A\}} \models \phi$ if and only if $c_A \geq d_A$. Also, if $\emptyset \neq \boldsymbol{Y} \subseteq \boldsymbol{X} \setminus \{A\}$, we have $s^{\boldsymbol{Y} \cup \{A\}} \models \phi$ if and only if $s^{\boldsymbol{Y}} \models \phi^A$, where $s^{\boldsymbol{Y} \cup \{A\}}$ and $s^{\boldsymbol{Y}}$ refer specifically to the set functions over $\boldsymbol{X}$ and $\boldsymbol{X} \setminus \{A\}$, respectively. The statement of the lemma follows. ◀

In particular, if each $\boldsymbol{Y}_i$ in the inequality (15) is either a singleton or the full set $\boldsymbol{X}$, then checking validity of this inequality reduces to checking validity of a linear number of inequalities in which the sets appearing in the right-hand side are all singletons. Theorems 7, 11, and 12, and Lemma 13 thus entail that the validity of such inequalities (15) over normal polymatroids, entropic functions, and polymatroids can be determined in polynomial time. This leads us to the following corollary.

▶ **Corollary 14.** *The information inequality problem w.r.t. simple Σ-inequalities is in polynomial time over normal polymatroids, entropic functions, and polymatroids.*

One may recall from Theorem 4 that, at least in the context of step functions, the requirement of Σ being simple is necessary.

We conclude this section by offering an alternative proof for the fact that the entropic bound for simple sets of conditionals Σ is polynomial-time computable. In order to formulate this statement precisely, we need the concept of the logarithmic bound. Similarly to the degree values, the *log-degree values* associated with Σ are defined as a sequence $\boldsymbol{b} = (b_\sigma)_{\sigma \in \Sigma}$, where $b_\sigma \geq 0$. A function $\boldsymbol{h}$ *satisfies* $(\Sigma, \boldsymbol{b})$, denoted $\boldsymbol{h} \models (\Sigma, \boldsymbol{b})$, if $h(\sigma) \leq b_\sigma$ for all $\sigma \in \Sigma$. For a set $S \subseteq \mathbb{R}^{2^n}$, and a set of conditionals Σ guarded by a query $Q(\boldsymbol{X})$ and associated with values $\boldsymbol{b}$, define the *log-bound* of Q w.r.t. S as

$$\text{Log-Bound}_S(Q, \Sigma, \boldsymbol{b}) := \inf_{\substack{\boldsymbol{w} \geq 0 \\ S \models \phi_\Sigma(\boldsymbol{X}, \boldsymbol{w})}} \sum_{\sigma \in \Sigma} w_\sigma b^\sigma,$$

where $\phi_\Sigma(\boldsymbol{X}, \boldsymbol{w})$ is the Σ-inequality (9). It is known that the *entropic log-bound* $\text{Log-Bound}_{\Gamma_n^*}$ is computable in polynomial time [11]. In the following, we present an alternative proof for this fact via monotone functions.

▶ **Theorem 15.** *Let Σ be a set of conditionals that is guarded by a query $Q(\boldsymbol{X})$ and associated with values $\boldsymbol{b}$. If Σ is simple, the entropic log-bound* $\text{Log-Bound}_{\Gamma_n^*}(Q, \Sigma, \boldsymbol{b})$ *is computable in polynomial time in the size of the input $(Q, \Sigma, \boldsymbol{b})$.*

Proof. We construct a linear program that is polynomial in the size of the input and such that its optimal value is attained at the entropic log-bound. Theorem 11 entails

$$\Gamma_n^* \models \phi_\Sigma(\boldsymbol{X}, \boldsymbol{w}) \iff \mathrm{S}_n \models \phi_\Sigma(\boldsymbol{X}, \boldsymbol{w}), \tag{17}$$

where ϕ_Σ is the Σ-inequality (9). Lemma 13 implies that

$$\mathrm{S}_n \models \phi_\Sigma(\boldsymbol{X}, \boldsymbol{w}) \iff \forall A \in \boldsymbol{X} : c_A \geq d_A \text{ and } \mathrm{S}_{n-1} \models \phi_\Sigma^A, \tag{18}$$

where c_A, d_A are the sums of coefficients w_σ computed from ϕ_Σ for a variable A. Since ϕ_Σ^A contain only singletons on their right-hand sides, Lemma 12 yields

$$\mathrm{S}_{n-1} \models \phi_\Sigma^A \iff \mathrm{Mon}_{n-1} \models \phi_\Sigma^A.$$

By Theorem 7 we can construct in polynomial time matrices M_A such that

$$\mathrm{Mon}_{n-1} \models \phi^A_\Sigma \iff M_A \boldsymbol{x}_A \geq \boldsymbol{w}_A \text{ for some } \boldsymbol{x}_A \geq 0,$$

where $\boldsymbol{w}_A$ is a list (with possible repetitions) of coefficients w_σ that appear in ϕ^A_Σ. Note that we should now treat w_σ as variables, since we are interested in optimizing their values. Thus we rewrite $M_A \boldsymbol{x}_A \geq \boldsymbol{w}_A \wedge c_A \geq d_A$ as $M'_A \boldsymbol{x}_A \boldsymbol{w}_A \geq 0$, where M'_A is obtained from $(M_A \mid -I_{|\boldsymbol{w}_A|})$ by adding one extra row to describe the inequality $c_A \geq d_A$. Then we construct a single matrix M^* such that $M^* \boldsymbol{x}\boldsymbol{w} = (\boldsymbol{x}_A \boldsymbol{w}_A)_{A \in \boldsymbol{X}}$, where $\boldsymbol{w} = (w_\sigma)_{\sigma \in \Sigma}$, and $\boldsymbol{x}$ is the concatenation of all $\boldsymbol{x}_A$. Finally, composing M'_A diagonally into a single matrix $M_{\boldsymbol{X}}$, and writing $M_\sigma = M_{\boldsymbol{X}} M^*$, we obtain

$$\forall A \in \boldsymbol{X} : c_A \geq d_A \text{ and } \mathrm{S}_{n-1} \models \phi^A_\Sigma \iff M_\Sigma \boldsymbol{x}\boldsymbol{w} \geq 0. \tag{19}$$

By Eqs. (17), (18), and (19) we obtain

$$\text{Log-Bound}_{\Gamma^*_n}(Q, \Sigma, \boldsymbol{b}) = \inf_{\substack{\boldsymbol{w} \geq 0 \\ \Gamma^*_n \models \phi_\Sigma(\boldsymbol{X}, \boldsymbol{w})}} \sum_{\sigma \in \Sigma} w_\sigma b^\sigma = \min_{\substack{\boldsymbol{x}\boldsymbol{w} \geq 0 \\ M_\Sigma \boldsymbol{x}\boldsymbol{w} \geq 0}} \sum_{\sigma \in \Sigma} w_\sigma b^\sigma.$$

Since M_Σ can be constructed in polynomial time in the size of $(Q, \Sigma, \boldsymbol{b})$, we can compute in polynomial time the entropic log-bound $\text{Log-Bound}_{\Gamma^*_n}(Q, \Sigma, \boldsymbol{b})$ as the optimal value of the linear program

$$\begin{aligned} \text{minimize} \quad & \sum_{\sigma \in \Sigma} w_\sigma b^\sigma \\ \text{subject to} \quad & M_\Sigma \boldsymbol{x}\boldsymbol{w} \geq \boldsymbol{0} \\ & \boldsymbol{x}\boldsymbol{w} \geq \boldsymbol{0}. \end{aligned}$$ ◀

5 Conclusion

The present paper marks the first attempt to demarcate the tractability boundary for different variants of the information inequality problem, introduced in [14]. We established that this problem is coNP-complete over normal polymatroids, and in polynomial time over monotone functions. Restricted to Σ-inequalities where $|\boldsymbol{U}| \leq 2$ for all $(\boldsymbol{V} \mid \boldsymbol{U}) \in \Sigma$, we proved that the information inequality problem remains coNP-hard over normal polymatroids. The same problem was shown to be in polynomial time over normal polymatroids, entropic functions, and polymatroids if $|\boldsymbol{U}| \leq 1$. If every set in the right-hand side of Eq. (15) is a singleton or the full variable set, we proved that the information inequality problem is in polynomial time over any K that falls inbetween normal polymatroids and monotone functions. Using this result, we constructed an alternative proof for the polynomial-time computability of the entropic bound in the case where the set of conditionals Σ is simple.

Based on these findings we may delineate a preliminary complexity classification of information inequalities over different set functions and syntactic classes. Consider an information inequality ϕ over n variables, presented in the form (15). If A and B are subsets of $[n]$, we say that ϕ is of *type* (A, B) if $|\boldsymbol{X}_i| \in A$ for each $\boldsymbol{X}_i$ appearing in the left-hand side in (15), and $|\boldsymbol{Y}_j| \in A$ for each $\boldsymbol{Y}_j$ appearing in the right-hand side in (15). For instance, the inequality (8) is of type $(\{1, n\}, \{2\})$, and all simple Σ-inequalities are of type $([n], \{1, n\})$. Using this convention, Tab. 3 summarizes the results of this paper.

Specifically, we showed that results on step functions and monotone functions lead to a polynomial-time algorithm for the entropic bound over simple degree constraints. To find more results of this kind, it may be useful to extend investigations to also other classes

Table 3 Complexity of the information inequality problem for different syntactic types and set functions.

types/set functions	Mod_n	N_n	Γ_n^*	Γ_n	Mon_n
$([n],[n]),(\{1,n\},\{2\})$	$\in$ P[23]	coNP-complete	$\in \Pi_1^0$ [14]	$\in$ EXP [27]	$\in$ P
$([n],\{1,n\})$	$\in$ P[23]	$\in$ P	$\in$ P	$\in$ P	$\in$ P

of set functions. For instance, as illustrated in Examples 1 and 3, as k grows uniform distributions over (bags of) k tuples yield increasingly accurate answers to questions about entropic constraints, compared to step functions derived from two tuples. One way to identify more decidable classes of information inequalities would be to find syntactic restrictions for which validity is captured by uniform distributions over k tuples, for some fixed k.

References

1 Marcelo Arenas and Leonid Libkin. An information-theoretic approach to normal forms for relational and XML data. *J. ACM*, 52(2):246–283, 2005. `doi:10.1145/1059513.1059519`.

2 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`.

3 Catriel Beeri, Ronald Fagin, and John H. Howard. A complete axiomatization for functional and multivalued dependencies in database relations. In *SIGMOD Conference*, pages 47–61. ACM, 1977. `doi:10.1145/509404.509414`.

4 Randall Dougherty, Chris Freiling, and Kenneth Zeger. Non-shannon information inequalities in four random variables, 2011. `doi:10.48550/arXiv.1104.3602`.

5 E. Mark Gold. Complexity of automaton identification from given data. *Inf. Control.*, 37(3):302–320, 1978. `doi:10.1016/S0019-9958(78)90562-4`.

6 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`.

7 Martin Grohe and Dániel Marx. Constraint solving via fractional edge covers. *ACM Trans. Algorithms*, 11(1):4:1–4:20, 2014. `doi:10.1145/2636918`.

8 Emirhan Gürpinar and Andrei E. Romashchenko. How to use undiscovered information inequalities: Direct applications of the copy lemma. In *ISIT*, pages 1377–1381. IEEE, 2019. `doi:10.1109/ISIT.2019.8849309`.

9 Miika Hannula. Information inequality problem over set functions. *CoRR*, abs/2309.11818, 2023. `doi:10.48550/arXiv.2309.11818`.

10 Christian Herrmann. On the undecidability of implications between embedded multivalued database dependencies. *Information and Computation*, 122(2):221–235, 1995. `doi:10.1006/inco.1995.1148`.

11 Sungjin Im, Benjamin Moseley, Hung Q. Ngo, Kirk Pruhs, and Alireza Samadian. Optimizing polymatroid functions. *CoRR*, abs/2211.08381, 2022. `doi:10.48550/arXiv.2211.08381`.

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

13 Batya Kenig and Dan Suciu. Integrity constraints revisited: From exact to approximate implication. *Log. Methods Comput. Sci.*, 18(1), 2022. `doi:10.46298/lmcs-18(1:5)2022`.

14 Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. Decision problems in information theory. In *ICALP*, volume 168 of *LIPIcs*, pages 106:1–106:20, 2020. `doi:10.4230/LIPIcs.ICALP.2020.106`.

15 Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suciu. Bag query containment and information theory. *ACM Trans. Database Syst.*, 46(3):12:1–12:39, 2021. `doi:10.1145/3472391`.

16 Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu. Computing join queries with functional dependencies. In *PODS*, pages 327–342. ACM, 2016. doi:10.1145/2902251.2902289.
17 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 *PODS*, pages 429–444. ACM, 2017. doi:10.1145/3034786.3056105.
18 Lukas Kühne and Geva Yashfe. On entropic and almost multilinear representability of matroids. *CoRR*, abs/2206.03465, 2022. doi:10.48550/arXiv.2206.03465.
19 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.
20 Tony T. Lee. An information-theoretic analysis of relational databases - part II: information structures of database schemas. *IEEE Trans. Software Eng.*, 13(10):1061–1072, 1987. doi:10.1109/TSE.1987.232848.
21 Cheuk Ting Li. Undecidability of network coding, conditional information inequalities, and conditional independence implication. *IEEE Trans. Inf. Theory*, 69(6):3493–3510, 2023. doi:10.1109/TIT.2023.3247570.
22 Wing Ning Li. Two-segmented channel routing is strong np-complete. *Discret. Appl. Math.*, 78(1-3):291–298, 1997. doi:10.1016/S0166-218X(97)00020-6.
23 Hung Q. Ngo. Worst-case optimal join algorithms: Techniques, results, and open problems. In *PODS*, pages 111–124. ACM, 2018. doi:10.1145/3196959.3196990.
24 Hung Q. Ngo, Ely Porat, Christopher Ré, and Atri Rudra. Worst-case optimal join algorithms. *J. ACM*, 65(3):16:1–16:40, 2018. doi:10.1145/3180143.
25 Nicholas Pippenger. What are the laws of information theory. In *Special Problems on Communication and Computation Conference*, pages 3–5, 1986.
26 Dan Suciu. Applications of information inequalities to database theory problems. In *LICS*, pages 1–30, 2023. doi:10.1109/LICS56636.2023.10175769.
27 Raymond W. Yeung. *Information Theory and Network Coding*. Springer Publishing Company, Incorporated, 1 edition, 2008.
28 Z. Zhang and R.W. Yeung. A non-shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory*, 43(6):1982–1986, 1997. doi:10.1109/18.641561.

A Alternative coNP-hardness proof

Recall that validity coincides for step functions and normal polymatroids, and thus it suffices to consider validity in the former sense. We reduce from three-colorability. Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a graph consisting of a vertex set $\boldsymbol{V}$ and a set of undirected edges $\boldsymbol{E}$. For each node $A \in \boldsymbol{V}$, we introduce variables A_r, A_g, A_b representing possible colors of A. Assume that the graph contains n vertices. We define

$$\sum_{\substack{c\in\{r,g,b\}\\ A\in\boldsymbol{V}}} h(A_c)+\sum_{\substack{c,d\in\{r,g,b\}\\ c\neq d\\ A\in\boldsymbol{V}}}(2n+1)h(\boldsymbol{V}\mid A_cA_d)+\sum_{\substack{c\in\{r,g,b\}\\ \{A,B\}\in\boldsymbol{E}}}(2n+1)h(\boldsymbol{V}\mid A_cB_c)\geq(2n+1)h(\boldsymbol{V}). \tag{20}$$

We claim that G is three-colorable if and only if Eq. (20) is not valid over S_n.

Assume first Eq. (20) is not valid, and let $s_{\boldsymbol{U}}$, $\boldsymbol{U} \subseteq \boldsymbol{V}$, be a step function such that Eq. (20) is false for $h = s_{\boldsymbol{U}}$. We claim that the function that maps each vertex A to a color c if $A_c \in \boldsymbol{U}$ is well-defined and constitutes a coloring of the graph. Since the entropy and the conditional entropy are non-negative for all step functions, we have $s_{\boldsymbol{U}}(\boldsymbol{V}) = 1$, and thus the right-hand side of Eq. (20) is $2n+1$. Consequently, the left-hand side is at most $2n$. From the first summation term, we obtain that $\boldsymbol{U}$ must contain at least n elements. Moreover, each term of the form $h(\boldsymbol{V} \mid A_cA_d)$ or $h(\boldsymbol{V} \mid A_cB_c)$ must be zero. In particular, we have $A_cA_d \not\subseteq \boldsymbol{U}$ and $A_cB_c \not\subseteq \boldsymbol{U}$, which entails that each vertex is assigned exactly one color, and no two vertices connected by an edge are assigned the same color. We conclude that the function defined by the step function is well defined and constitutes a graph coloring.

Assume then Eq. (20) is valid. For each coloring of the vertices we may define a subset $\boldsymbol{U} \subseteq \boldsymbol{V}$ such that $A_c \in \boldsymbol{U}$ if and only if vertex A is assigned color c. Then, the first summation term in the left-hand side of Eq. (20) is n, and the second summation term is zero. By hypothesis, some term of the form $h(\boldsymbol{V} \mid A_cB_c)$ must be non-zero, which means that there exists an edge whose endpoints are assigned the same color. This concludes the proof of the claim.

Since the reduction is in polynomial time, and each coefficient is bounded by a polynomial in the input size, strong coNP-completeness again follows. ◀

B Completeness of fixed-point algorithm

▶ **Lemma 9.** *Let ϕ be an information inequality of the form*

$$c_1h(\boldsymbol{X}_1) + \cdots + c_kh(\boldsymbol{X}_k) \geq 0 \quad (c_i \in \mathbb{R}). \tag{11}$$

The following are equivalent:

1. *ϕ is valid over monotone functions.*
2. *ϕ is valid over monotone, Boolean-valued functions.*
3. *ϕ is a positive and separable combination of monotonicity and non-negativity axioms.*

Proof. The implications (3) ⇒ (1) and (1) ⇒ (2) are immediate. We prove that (2) ⇒ (3). Clearly, if this implication holds w.r.t. $c_i \in \mathbb{Z}$, then it holds w.r.t. $c_i \in \mathbb{Q}$. We first prove the following claim.

▷ Claim 16. If the implication (2) ⇒ (3) holds w.r.t. $c_i \in \mathbb{Q}$, then it holds w.r.t. $c_i \in \mathbb{R}$.

Proof. To prove this, assume ϕ is valid over $\mathrm{Mon}_n^{0,1}$. Let (ϕ_n) be a sequence of information inequalities

$$c_1^nh(\boldsymbol{X}_1) + \cdots + c_k^nh(\boldsymbol{X}_k) \geq 0 \quad (c_i^n \in \mathbb{Q}, n \geq 1), \tag{21}$$

where $\lim_{n\to\infty} c_i^n = c_i$ and $c_i^n \geq c_i^{n+1}$. We may assume that c_i^1 is negative if c_i is negative. That is, (c_i^n) is a sequence of positive (resp. negative) values if c_i is positive (resp. negative). Clearly, if ϕ is valid over Boolean-valued, monotone functions, then so are ϕ_n. By hypothesis, ϕ_n decompose into positive and separable combinations of monotonicity and non-negativity axioms. Writing $\boldsymbol{c}_\phi$ for the coefficient function arising from ϕ, we may write

$$\boldsymbol{c}_{\phi_n} = d_1^n\boldsymbol{c}_{\psi_1} + \cdots + d_m^n\boldsymbol{c}_{\psi_m} \quad (d_i^n \geq 0), \tag{22}$$

where ψ_l list all possible monotonicity and non-negativity axioms respectively of the form $h(\boldsymbol{X}_i) \geq 0$ and $h(\boldsymbol{X}_i) - h(\boldsymbol{X}_j) \geq 0$, where $i, j \in [k]$ and $\boldsymbol{X}_j \subseteq \boldsymbol{X}_i$, excluding those ψ_l for which the coefficient d_l^n is always zero. That is, the combinations (22) are separable and have fixed length over all $n \geq 1$; recall that separability was defined with respect to terms having a non-zero coefficient. Fix attention to an arbitrary ψ_l being either of the form $h(\boldsymbol{X}_i) \geq 0$ or $h(\boldsymbol{X}_i) - h(\boldsymbol{X}_j) \geq 0$. In this case, the coefficient function $\boldsymbol{c}_{\psi_l}$ maps $\boldsymbol{X}_i$ to 1, that is, $c_{\psi_l}(\boldsymbol{X}_i) = 1$. We claim that the coefficient c_i of $h(\boldsymbol{X}_i)$ in Eq. (11) is positive. For this, consider some $p \geq 1$ such that the coefficient d_l^p of $\boldsymbol{c}_{\psi_l}$ is strictly positive. Assume toward contradiction that c_i is not positive, meaning that it is negative. Then by construction, c_i^p is negative (i.e., $c_{\phi_p}(\boldsymbol{X}_i) < 0$), whence $c_{\psi_{l'}}(\boldsymbol{X}_i) < 0$ for some $l' \neq l$ associated with a strictly positive coefficient $d_{l'}^p$. Since $c_{\psi_l}(\boldsymbol{X}_i) > 0$, this contradicts separability of (22), proving our claim. The claim entails by construction that c_i^n are positive for all $n \geq 1$. Hence we obtain $d_l^n \leq c_i^n \leq c_i^1$ by separability of (22).

We conclude that $(\boldsymbol{d}_n) = (d_1^n, \dots, d_m^n)$ is an infinite and bounded sequence of $\mathbb{R}^m$. The Bolzano-Weierstrass theorem entails that $(\boldsymbol{d}_n)$ has a subsequence $(\boldsymbol{d}_{n_p})$ that converges to some $\boldsymbol{d} = (d_1, \dots, d_m)$. Obviously the vector $\boldsymbol{d}$ is non-negative. By continuity,

$$\boldsymbol{c}_\phi = \lim_{p\to\infty} \boldsymbol{c}_{\phi_{n_p}} = \lim_{p\to\infty} d_1^{n_p}\boldsymbol{c}_{\psi_1} + \dots + d_m^{n_p}\boldsymbol{c}_{\psi_m} = d_1\boldsymbol{c}_{\psi_1} + \dots + d_m\boldsymbol{c}_{\psi_m}.$$

The obtained combination is separable, because otherwise some combination (22) is not separable for large enough n_p, which leads to a contradiction. We conclude that ϕ is a positive and separable combination of monotonicity and non-negativity axioms, which shows that (2) $\Rightarrow$ (3) w.r.t. $c_i \in \mathbb{R}$. ◁

It remains to prove that (2) $\Rightarrow$ (3) w.r.t. $c_i \in \mathbb{Z}$. Let $S = (S^+, S^-)$ be the set representation of ϕ. We associate S with a directed graph G_S, where

- the set of nodes are the elements of S^+ and S^-, and
- there is a directed edge from $(\boldsymbol{X}, i)$ to $(\boldsymbol{Y}, j)$ if $(\boldsymbol{X}, i) \in S^+$, $(\boldsymbol{Y}, j) \in S^-$, and $\boldsymbol{Y} \subseteq \boldsymbol{X}$.

Consider Alg. 1 which maintains a bipartite directed graph G that is initially set up as G_S. The monotonicity axioms isolated at the current step are represented as directed edges going from S^- to S^+ *backward edges*; in the beginning no such edges have been introduced yet. The edges that proceed from S^+ to S^- (*forward edges*) are kept fixed.

We say that a node u is *connected* to a node v in a directed graph if $u = v$, or there is a sequence of nodes (a *path* from u to v) in which the first node is u, the last node is v, and each node is connected to the following node by a directed edge. A set $\boldsymbol{U}$ is *connected* to another set $\boldsymbol{V}$ is some node in $\boldsymbol{U}$ is connected to some node in $\boldsymbol{V}$.

Consider the following claim.

▷ Claim 17. If Alg. 1 returns true on ϕ, then ϕ is a positive and separable combination of the monotonicity and non-negativity axioms.

Proof. Consider the graph G and the sets S_0 and S_1 after termination of the algorithm. Note that if G contains a backward edge (u, v), then the reverse edge (v, u) forms a forward edge of G_S and consequently corresponds to a monotonicity axiom. The backward edges also form a bijection from $S^- \setminus S_1$ to $S^+ \setminus S_0$. Since S_1 is empty by assumption, and S_0 can be viewed as representing non-negativity axioms, it can now be observed that ϕ decomposes into a positive and separable combination of monotonicity and non-negativity axioms. This proves the claim. ◁

We now prove the contraposition of (2) $\Rightarrow$ (3) w.r.t. $c_i \in \mathbb{Z}$. Suppose ϕ is not a positive and separable combination of the monotonicity axioms. The previous claim entails that the algorithm returns false. Consider again the graph G and the sets S_0, S_1 after termination of the algorithm. Note that S_1 is now non-empty. Let $\boldsymbol{V}$ denote the set of variables appearing in ϕ. Let $\mathcal{Y}$ be the (non-empty) collection of sets $\boldsymbol{Y} \subseteq \boldsymbol{V}$ such that for some j, $(\boldsymbol{Y}, j)$ belongs to S^- and is connected to S_1. Consider also its *upper closure* $\mathcal{Y}^\uparrow := \{\boldsymbol{Z} \subseteq \boldsymbol{V} \mid \exists \boldsymbol{Y} \in \mathcal{Y} : \boldsymbol{Y} \subseteq \boldsymbol{Z}\}$. Define a mapping h such that $h(\boldsymbol{Z}) = 1$ if $\boldsymbol{Z} \in \mathcal{Y}^\uparrow$, and otherwise $h(\boldsymbol{Z}) = 0$. Clearly, h is a Boolean, monotone function. We show that h does not satisfy ϕ.

Consider a pair $(\boldsymbol{X}, i) \in S^+$ such that $h(\boldsymbol{X}) = 1$. Then, $\boldsymbol{X}$ contains a set $\boldsymbol{Y}$ from $\mathcal{Y}$. Let j be such that $(\boldsymbol{Y}, j)$ belongs to S^- and is connected to S_1. Since there is an edge from $(\boldsymbol{X}, i)$ to $(\boldsymbol{Y}, j)$, it follows that $(\boldsymbol{X}, i)$ is connected to S_1. Now, if $(\boldsymbol{X}, i)$ belonged to S_0, the algorithm could not have terminated yet. Hence $(\boldsymbol{X}, i)$ must belong to $S^+ \setminus S_0$. Recall that the backward edges form a bijection from $S^- \setminus S_1$ to $S^+ \setminus S_0$. In particular, $(\boldsymbol{X}, i)$ is the target of a unique backward edge with a source node $(\boldsymbol{Z}, k)$. Since $(\boldsymbol{X}, i)$ is connected to S_1, it follows that $(\boldsymbol{Z}, k)$ is also connected to S_1. This entails that $h(\boldsymbol{Z}) = 1$. In particular, this shows that any $(\boldsymbol{X}, i) \in S^+$ such that $h(\boldsymbol{X}) = 1$ is paired by a backward edge with a unique $(\boldsymbol{Z}, k) \in S^- \setminus S_1$ such that $h(\boldsymbol{Z}) = 1$. In addition, because S_1 is non-empty, there exists an element $(\boldsymbol{U}, l) \in S^- \cap S_1$ such that $h(\boldsymbol{U}) = 1$. In particular, $(\boldsymbol{U}, l)$ is not the source node of any backward edge. These observations entail that h does not satisfy ϕ. This proves the contraposition of (2) $\Rightarrow$ (3) w.r.t. $c_i \in \mathbb{Z}$.

This concludes the proof of the direction (2) $\Rightarrow$ (3). ◀

The following example demonstrates that Alg. 1 correctly returns false on the submodularity axiom, as this axiom is not a consequence of monotonicity and non-negativity.

▶ **Example 18.** The submodularity axiom $XY + XZ - X - XYZ \geq 0$ is not valid over monotone functions. This can be also seen by referring to Alg. 1. The set representation is (S^+, S^-) where $S^+ = \{(XY, 1), (XZ, 1)\}$ and $S^- = \{(X, 1), (XYZ, 1)\}$. Suppose at the first step the algorithm introduces a backward edge from $(X, 1)$ to $(XY, 1)$; the only other option is the symmetric scenario where it introduces an edge from $(X, 1)$ to $(XZ, 1)$. After the first step we have $S_0 = \{(XZ, 1)\}$ and $S_1 = \{(XYZ, 1)\}$. Then, no path exists from S_0 to S_1, since no forward edge points to $(XYZ, 1)$. The algorithm therefore terminates returning false. Accordingly, the function that maps XYZ to 1 and all other sets to 0 is monotone, Boolean-valued, and does not satisfy the aforementioned submodularity axiom.