Integrity Constraints Revisited: From Exact to Approximate Implication

Batya Kenig
University of Washington, Seattle, WA, USA
batyak@cs.washington.edu

Dan Suciu
University of Washington, Seattle, WA, USA
https://homes.cs.washington.edu/~suciu/
suciu@cs.washington.edu

Abstract

Integrity constraints such as functional dependencies (FD), and multi-valued dependencies (MVD) are fundamental in database schema design. Likewise, probabilistic conditional independences (CI) are crucial for reasoning about multivariate probability distributions. The implication problem studies whether a set of constraints (antecedents) implies another constraint (consequent), and has been investigated in both the database and the AI literature, under the assumption that all constraints hold *exactly*. However, many applications today consider constraints that hold only *approximately*. In this paper we define an approximate implication as a linear inequality between the degree of satisfaction of the antecedents and consequent, and we study the *relaxation problem*: when does an exact implication relax to an approximate implication? We use information theory to define the degree of satisfaction, and prove several results. First, we show that any implication from a set of data dependencies (MVDs+FDs) can be relaxed to a simple linear inequality with a factor at most quadratic in the number of variables; when the consequent is an FD, the factor can be reduced to 1. Second, we prove that there exists an implication between CIs that does not admit any relaxation; however, we prove that every implication between CIs relaxes "in the limit". Finally, we show that the implication problem for differential constraints in market basket analysis also admits a relaxation with a factor equal to 1. Our results recover, and sometimes extend, several previously known results about the implication problem: implication of MVDs can be checked by considering only 2-tuple relations, and the implication of differential constraints for frequent item sets can be checked by considering only databases containing a single transaction.

2012 ACM Subject Classification Theory of computation → Database theory; Theory of computation → Database constraints theory

Keywords and phrases Integrity constraints, The implication problem

Digital Object Identifier 10.4230/LIPIcs.ICDT.2020.18

Related Version A full version of the paper is available at https://arxiv.org/abs/1812.09987.

Funding This work was supported by NSF under grants III-1614738 and IIS 1907997.

1 Introduction

Traditionally, integrity constraints are assertions about a database that are stated by the database administrator and enforced by the system during updates. However, in several applications of Big Data, integrity constraints are discovered, or mined in a database instance, as opposed to being asserted by the administrator [13, 34, 7, 3, 20]. For example, data cleaning can be done by first learning conditional functional dependencies in some reference data, then using them to identify inconsistencies in the test data [16, 7]. Causal reasoning [35, 28, 31] and learning sum-of-product networks [29, 11, 26] repeatedly discover conditional independencies in the data. Constraints also arise in many other domains, for example in the *frequent itemset*

23rd International Conference on Database Theory (ICDT 2020).
Editors: Carsten Lutz and Jean Christoph Jung; Article No. 18; pp. 18:1–18:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Table 1 Summary of results: relaxation bounds for the implication $\Sigma \Rightarrow \tau$ for the sub-cones of Γ_n under various restrictions. (1) *General*; no restrictions to either Σ or τ (2) Σ is a set of saturated CIs and conditional entropies (i.e., MVDs+FDs in databases), and τ is a conditional entropy. (3) Σ is a set of saturated CIs and conditional entropies, τ is *any* CI (4) *Disjoint* integrity constraints. The terms in Σ are both saturated and *disjoint* (see definition 10 in Sec. 4), and τ is saturated.

Cone	Relaxation Bounds			
	General	MVDs+FDs $\Rightarrow$ FD	MVDs+FDs $\Rightarrow$ any	Disjoint MVDs+FDs $\Rightarrow$ MVD/FD
Γ_n	$(2^n)!$ (Thm. 21)	1 (Thm. 6)	$\frac{n^2}{4}$ (Thm. 6)	1 (Thm. 11)
Γ_n^*	∞ (Thm. 16)	1 (Thm. 6)	$\frac{n^2}{4}$ (Thm. 6)	1 (Thm. 11)
$\mathcal{P}_n$	1 (Thm. 23)	1 (Thm. 23)	1 (Thm. 23)	1 (Thm. 23)

problem (FIS) [22, 5], or as *measure based constraints* [32] in applications like Dempster-Shafer theory, possibilistic theory, and game theory (see discussion in [32]). In all these applications, quite often the constraints are learned from the data, and are not required to hold exactly, but it suffices if they hold only to a certain degree.

The classical *implication problem* asks whether a set of constraints, called the *antecedents*, logically imply another constraint called the *consequent*. In this setting, both antecedents and consequent are required to hold exactly, hence we refer to it as an *exact implication* (EI). The database literature has extensively studied the EI problem for integrity constraints and shown that the implication problem is decidable and axiomatizable for Functional Dependencies (FDs) and Multivalued Dependencies (MVDs) [23, 19, 1, 2], and undecidable for Embedded Multivalued Dependencies (EMVDs) [15]. The AI community has studied extensively the EI problem for Conditional Independencies (CI), which are assertions of the form $X \perp Y \mid Z$, stating that X is independent of Y conditioned on Z, and has shown that the implication problem is decidable and axiomatizable for saturated CIs [12] (where $XYZ =$ all variables), but not finitely axiomatizable in general [36]. In the FIS problem, a constraint like $X \rightarrow Y \vee Z \vee U$ means that every basket that contains X also contains at least one of Y, Z, U, and the implication problem here is also decidable and axiomatizable [33].

The Relaxation Problem. In this paper we consider a new problem, called the *relaxation problem*: if an exact implication holds, does an approximate implication hold too? For example, suppose we prove that a given set of FDs implies another FD, but the input data satisfies the antecedent FDs only to some degree: to what degree does the consequent FD hold on the database? An *approximate implication* (AI) is an inequality that (numerically) bounds the consequent by a linear combination of the antecedents. The relaxation problem asks whether we can convert an EI into an AI. When relaxation holds, then any inference system for proving exact implication, e.g. using a set of axioms or some algorithm, can be used to infer an approximate implication.

In order to study the relaxation problem we need to measure the degree of satisfaction of a constraint. In this paper we use Information Theory. This is the natural semantics for modeling CIs of multivariate distributions, because $X \perp Y \mid Z$ iff $I(X;Y|Z) = 0$ where I is the *conditional mutual information*. FDs and MVDs are special cases of CIs [21, 8, 38] (reviewed in Sec. 2.1), and thus they are naturally modeled using the information theoretic measure $I(X;Y|Z)$ or $H(Y|X)$; in contrast, EMVDs do not appear to have a natural interpretation using information theory, and we will not discuss them here. Several papers have argued that information theory is a suitable tool to express integrity constraints [21, 8, 38, 24, 13].

An exact implication (EI) becomes an assertion of the form $(\sigma_1 = 0 \wedge \sigma_2 = 0 \wedge \ldots) \Rightarrow (\tau = 0)$, while an approximate implication (AI) is a linear inequality $\tau \leq \lambda \cdot \left(\sum \sigma_i\right)$, where $\lambda \geq 0$, and $\tau, \sigma_1, \sigma_2, \ldots$ are information theoretic measures. We say that a class of constraints can be *relaxed* if EI implies AI; we also say that it λ-relaxes, when we want to specify the factor λ in the AI. We notice an AI always implies EI.

Results. We make several contributions, summarized in Table 1. We start by showing in Sec. 4 that MVDs+FDs admit an $n^2/4$-relaxation, where n is the number of variables. When the consequent is an FD, we show that implication admits a 1-relaxation. Thus, whenever an exact implication holds between MVD+FDs, a simple linear inequality also holds between their associated information theoretic terms. In fact, we prove a stronger result that holds for CIs in general, which implies the result for MVDs+FDs. In addition, under some mild syntactic restrictions to the antecedents, we strengthen the result from a $n^2/4$-relaxation to a 1-relaxation, even when the consequent is an MVD; we leave open the question whether 1-relaxation exists in general.

So far, we have restricted ourselves to saturated or conditional CIs (which correspond to MVDs or FDs). In Sec. 5 we remove any restrictions, and prove a negative result: there exists an EI that does not relax (Eq. (9), based on an example in [17]). Nevertheless, we show that every EI can be relaxed to an AI plus an error term, which can be made arbitrarily small, at the cost of increasing the factor λ. This result implies that *every* EI can be proven from some inequality, corresponding to the AI associated to the EI, plus an error term. In fact, the EI in Eq. (9) follows from an inequality by Matúš [25], which is precisely the associated AI plus an error term; our result shows that every EI can be proven in this style.

Next, we consider two restrictions, which are commonly used in model theory. First, in Sec. 6 we restrict the class of *axioms* used to prove implications, to Shannon's inequalities (monotonicity and submodularity, reviewed in Sec. 2.2). In general, Shannon's inequalities are sound but incomplete for proving exact and approximate implications that hold for all probability distributions [41, 42], but they are complete for deriving inequalities that hold for all polymatroids [40]. We show that they are also complete for saturated+conditional constraints (as we show in Sec 4), and for measure-based constraints [32] (Sec. 7). We prove that every exact implication that holds for all polymatroids relaxes to an approximate implication, and prove an upper bound $\lambda \leq (2^n)!$, and a lower bound $\lambda \geq 3$; the exact bound remains open. Second, in Sec. 7 we restrict the class of *models* used to check an implication, to probability distributions with exactly 2 outcomes (tuples), each with probability 1/2; we justify this shortly. We prove that, under this restriction, the implication problem has a 1-relaxation. Restricting the models leads to a complete but unsound method for checking general implication, however this method is sound for saturated+conditional (as we show in Sec 4) and is also sound for checking FIS constraints (as we show in Sec. 7).

Two Consequences. While our paper is focused on relaxation, our results have two consequences for the exact implication problem. The first is a *2-tuple model property*: an exact implication, where the antecedents are saturated or conditional CIs, can be verified on uniform probability distributions with 2 tuples. A similar result is known for MVD+FDs [30]. Geiger and Pearl [12], building on an earlier result by Fagin [10], prove that every set of CIs has an *Armstrong model*: a discrete probability distribution that satisfies only the CIs and their consequences, and no other CI. The Armstrong model is also called a *global witness*, and, in general, can be arbitrarily large. Our result concerns a *local witness*: for any EI, if it fails on some probability distribution, then it fails on a 2-tuple uniform distribution.

The second consequence concerns the equivalence between the implication problem of saturated+conditional CIs with that of MVD+FDs. It is easy to check that the former implies the latter (Sec. 2). Wong et al. [38] prove the other direction, relying on the sound and complete axiomatization of MVDs [2]. Our 2-tuple model property implies the other direction immediately.

2 Notation and Preliminaries

We denote by $[n] = \{1, 2, \dots, n\}$. If $\Omega = \{X_1, \dots, X_n\}$ denotes a set of variables and $U, V \subseteq \Omega$, then we abbreviate the union $U \cup V$ with UV.

2.1 Integrity Constraints and Conditional Independence

A *relation instance* R over signature $\Omega = \{X_1, \dots, X_n\}$ is a finite set of tuples with attributes Ω. Let $X, Y, Z \subseteq \Omega$. We say that the instance R satisfies the *functional dependency* (FD) $X \to Y$, and write $R \models X \to Y$, if forall $t_1, t_2 \in R$, $t_1[X] = t_2[X]$ implies $t_1[Y] = t_2[Y]$. We say that R satisfies the *embedded multivalued dependency* (EMVD) $X \twoheadrightarrow Y \mid Z$, and write $R \models X \twoheadrightarrow Y \mid Z$, if for all $t_1, t_2 \in R$, $t_1[X] = t_2[X]$ implies $\exists t_3 \in R$ such that $t_1[XY] = t_3[XY]$ and $t_2[XZ] = t_3[XZ]$. One can check that $X \twoheadrightarrow Y \mid Y$ iff $X \to Y$. When $XYZ = \Omega$, then we call $X \twoheadrightarrow Y \mid Z$ a *multivalued dependency*, MVD; notice that X, Y, Z are not necessarily disjoint [2].

A set of constraints Σ *implies* a constraint τ, in notation $\Sigma \Rightarrow \tau$, if for every instance R, if $R \models \Sigma$ then $R \models \tau$. The implication problem has been extensively studied in the literature; Beeri et al. [2] gave a complete axiomatization of FDs and MVDs, while Herrman [15] showed that the implication problem for EMVDs is undecidable.

Recall that two discrete random variables X, Y are called *independent* if $p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$ for all outcomes x, y. Fix $\Omega = \{X_1, \dots, X_n\}$ a set of n jointly distributed discrete random variables with finite domains $\mathcal{D}_1, \dots, \mathcal{D}_n$, respectively; let p be the probability mass. For $\alpha \subseteq [n]$, denote by X_α the joint random variable $(X_i : i \in \alpha)$ with domain $\mathcal{D}_\alpha \stackrel{\text{def}}{=} \prod_{i \in \alpha} D_i$. We write $p \models X_\beta \perp X_\gamma | X_\alpha$ when X_β, X_γ are conditionally independent given X_α; in the special case $\beta = \gamma$, then $p \models X_\beta \perp X_\beta | X_\alpha$ iff X_α functionally determines[1] X_β, and we write $p \models X_\alpha \to X_\beta$.

An assertion $Y \perp Z | X$ is called a *Conditional Independence* statement, or a CI; this includes $X \to Y$ as a special case. When $XYZ = \Omega$ we call it *saturated*, and when $Z = \emptyset$ we call it *marginal*. A set of CIs Σ *implies* a CI τ, in notation $\Sigma \Rightarrow \tau$, if every probability distribution that satisfies Σ also satisfies τ. This implication problem has also been extensively studied: Pearl and Paz [27] gave a sound but incomplete set of *graphoid axioms*, Studeny [36] proved that no finite axiomatization exists, while Geiger and Pearl [12] gave a complete axiomatization for saturated, and marginal CIs.

Lee [21] observed the following connection between database constraints and CIs. The *empirical distribution* of a relation R is the uniform distribution over its tuples, in other words, $\forall t \in R$, $p(t) = 1/|R|$. Then:

▶ **Lemma 1.** ([21]) *Forall $X, Y, Z \subset \Omega$ such that $XYZ = \Omega$.*

$$R \models X \to Y \quad \Leftrightarrow \quad p \models X \to Y \qquad \textit{and} \qquad R \models X \twoheadrightarrow Y | Z \quad \Leftrightarrow \quad p \models (Y \perp Z | X) \tag{1}$$

[1] This means: $\forall u \in D_\alpha$, if $p(X_\alpha = u) \neq 0$ then $\exists v \in D_\beta$ s.t. $p(X_\beta = v | X_\alpha = u) = 1$, and v is unique.

The lemma no longer holds for EMVDs, and for that reason we no longer consider EMVDs in this paper. The lemma immediately implies that if Σ, τ are saturated and/or conditional CIs and the implication $\Sigma \Rightarrow \tau$ holds for all probability distributions, then the corresponding implication holds in databases, where the CIs are interpreted as MVDs or FDs respectively. Wong [38] gave a non-trivial proof for the other direction; we will give a much shorter proof in Corollary 8.

2.2 Background on Information Theory

We adopt required notation from the literature on information theory [40, 6]. For $n > 0$, we identify vectors in $\mathbb{R}^{2^n}$ with functions $2^{[n]} \to \mathbb{R}$.

Polymatroids. A function[2] $h \in \mathbb{R}^{2^n}$ is called a *polymatroid* if $h(\emptyset) = 0$ and satisfies the following inequalities, called *Shannon inequalities*:

1. Monotonicity: $h(A) \leq h(B)$ for $A \subseteq B$
2. Submodularity: $h(A \cup B) + h(A \cap B) \leq h(A) + h(B)$ for all $A, B \subseteq [n]$

The set of polymatroids is denoted $\Gamma_n \subseteq \mathbb{R}^{2^n}$, and forms a polyhedral cone (reviewed in Sec. 5). For any polymatroid h and subsets $A, B, C \subseteq [n]$, we define[3]

$$h(B|A) \stackrel{\text{def}}{=} h(AB) - h(A) \tag{2}$$

$$I_h(B;C|A) \stackrel{\text{def}}{=} h(AB) + h(AC) - h(ABC) - h(A) \tag{3}$$

Then, $\forall h \in \Gamma_n$, $I_h(B;C|A) \geq 0$ and $h(B|A) \geq 0$. The *chain rule* is the identity:

$$I_h(B;CD|A) = I_h(B;C|A) + I_h(B;D|AC) \tag{4}$$

We call $I_h(B;C|A)$ *saturated* if $ABC = [n]$, and *elemental* if $|B| = |C| = 1$; $h(B|A)$ is a special case of I_h, because $h(B|A) = I_h(B;B|A)$.

Entropic Functions. If X is a random variable with a finite domain $\mathcal{D}$ and probability mass p, then $H(X)$ denotes its entropy

$$H(X) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{D}} p(x) \log \frac{1}{p(x)} \tag{5}$$

For a set of jointly distributed random variables $\Omega = \{X_1, \ldots, X_n\}$ we define the function $h : 2^{[n]} \to \mathbb{R}$ as $h(\alpha) \stackrel{\text{def}}{=} H(X_\alpha)$; h is called an *entropic function*, or, with some abuse, an *entropy*. The set of entropic functions is denoted Γ_n^*. The quantities $h(B|A)$ and $I_h(B;C|A)$ are called the *conditional entropy* and *conditional mutual information* respectively. The conditional independence $p \models B \perp C \mid A$ holds iff $I_h(B;C|A) = 0$, and similarly $p \models A \to B$ iff $h(B|A) = 0$, thus, entropy provides us with an alternative characterization of CIs.

2-Tuple Relations and Step functions. 2-tuple relations play a key role for the implication problem of MVDs+FDs: if an implication fails, then there exists a witness consisting of only two tuples [30]. We define a *step function* as the entropy of the empirical distribution of a

[2] Most authors consider rather the space $\mathbb{R}^{2^n-1}$, by dropping $h(\emptyset)$ because it is always 0.
[3] Recall that AB denotes $A \cup B$.

X_1	X_2	U_1	U_2	Pr
0	0	0	0	1/2
1	1	0	0	1/2

(a)

X	Y	Z	Pr
0	0	0	1/4
0	1	1	1/4
1	0	1	1/4
1	1	0	1/4

(b)

A	B	C	D	Pr
0	0	0	0	$1/2-\varepsilon$
0	1	0	1	$1/2-\varepsilon$
1	0	1	0	ε
1	1	0	0	ε

(c)

Figure 1 Two relations and their empirical distribution (a),(b); a distribution from [17] (c).

2-tuple relation; $R = \{t_1, t_2\}$, $t_1 \neq t_2$, and $p(t_1) = p(t_2) = 1/2$. We denote the step function by h_U, where $U \subsetneq \Omega$ is the set of attributes where t_1, t_2 agree. One can check:

$$h_U(W) = \begin{cases} 0 & \text{if } W \subseteq U \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

If we set $U = \Omega$ in (6) then $h_\Omega \equiv 0$. Unless otherwise stated, in this paper we do not consider h_Ω to be a step function. Thus, there are $2^n - 1$ step functions and their set is denote $\mathcal{S}_n$. We will use the following fact extensively in this paper: $I_{h_U}(Y; Z|X) = 1$ if $X \subseteq U$ and $Y, Z \not\subseteq U$, and $I_{h_U}(Y; Z|X) = 0$ otherwise.

▶ **Example 2.** Consider the relational instance in Fig. 1 (a). It's entropy is the step function $h_{U_1U_2}(W)$, which is 0 for $W \subseteq U_1U_2$ and 1 otherwise. $R \models X_1 \rightarrow X_2$ because $h(X_2|X_1) = h(X_1X_2) - h(X_1) = 1 - 1 = 0$, and $R \not\models U_1 \rightarrow X_1$ because $h(X_1|U_1) = h(X_1U_1) - h(U_1) = 1 - 0 \neq 0$.

The relational instance $R = \{(x, y, z) \mid x + y + z \mod 2 = 0\}$ in Fig. 1 (b) is called the *parity function*. It's entropy is $h(X) = h(Y) = h(Z) = 1$, $h(XY) = h(XZ) = h(YZ) = h(XYZ) = 2$. We have that $R \models Y \perp Z$ because $I_h(Y; Z) = h(Y) + h(Z) - h(YZ) = 1 + 1 - 2 = 0$, but $R \not\models Y \perp Z|X$ because $I_h(Y; Z|X) = 1$ [4].

2.3 Discussion

This paper studies exact and approximate implications, expressed as equalities or inequalities of entropic functions h. For example, the augmentation axiom for MVDs [2] $A \twoheadrightarrow B|CD \Rightarrow AC \twoheadrightarrow B|D$ is expressed as $I_h(B; CD|A) = 0 \Rightarrow I_h(B; D|AC) = 0$, which holds by the chain rule (4). Thus, our golden standard is to prove that (in)equalities hold forall entropic functions, Γ_n^*. It is known that Γ_n^* is not topologically closed [40]; its topological closure, $\mathbf{cl}(\Gamma_n^*)$, is called the set of *almost entropic functions*. If an *inequality* holds for all entropic functions $h \in \Gamma_n^*$, then, by continuity, it also holds for all almost entropic functions $h \in \mathbf{cl}(\Gamma_n^*)$. However, this observation does not extend to *implications* of (in)equalities; Kaced and Romashchenko [17] gave an example of an exact implication that holds only for entropic functions but fails for almost entropic functions. Thus, when discussing an EI, it matters whether we assume that it holds for Γ_n^* or for $\mathbf{cl}(\Gamma_n^*)$. The only result in this paper where this distinction matters are the two main theorems in Sec. 5: the negative result Theorem 16 holds for both Γ_n^* and for $\mathbf{cl}(\Gamma_n^*)$, while the positive result Theorem 17 holds only for $\mathbf{cl}(\Gamma_n^*)$. The results in Sec. 4 apply to *any* set of polymatroids K that contains all step functions, i.e. $\mathcal{S}_n \subseteq K \subseteq \Gamma_n$, thus they apply to both Γ_n^* and $\mathbf{cl}(\Gamma_n^*)$, while those in Sec 6 and Sec. 7 are stated only for Γ_n and only for (the conic closure of) $\mathcal{S}_n$ respectively.

[4] $h(XY) + h(XZ) - h(X) - h(XYZ) = 2 + 2 - 1 - 2 = 1$

3 Definition of the Relaxation Problem

We now formally define the relaxation problem. We fix a set of variables $\Omega = \{X_1, \dots, X_n\}$, and consider formulas of the form $\sigma = (Y; Z|X)$, where $X, Y, Z \subseteq \Omega$, which we call a *conditional independence*, CI; when $Y = Z$ then we write it as $X \to Y$ and call it a *conditional*. An *implication* is a formula $\Sigma \Rightarrow \tau$, where Σ is a set of CIs called *antecedents* and τ is a CI called *consequent*. For a CI $\sigma = (B; C|A)$, we define $h(\sigma) \stackrel{\text{def}}{=} I_h(B; C|A)$, for a set of CIs Σ, we define $h(\Sigma) \stackrel{\text{def}}{=} \sum_{\sigma \in \Sigma} h(\sigma)$. Fix a set K s.t. $\mathcal{S}_n \subseteq K \subseteq \Gamma_n$.

▶ **Definition 3.** *The* exact implication *(EI) $\Sigma \Rightarrow \tau$ holds in K, denoted $K \models_{EI} (\Sigma \Rightarrow \tau)$ if, forall $h \in K$, $h(\Sigma) = 0$ implies $h(\tau) = 0$. The* λ-approximate implication *(λ-AI) holds in K, in notation $K \models \lambda \cdot h(\Sigma) \geq h(\tau)$, if $\forall h \in K$, $\lambda \cdot h(\Sigma) \geq h(\tau)$. The* approximate implication *holds, in notation $K \models_{AI} (\Sigma \Rightarrow \tau)$, if there exist a $\lambda \geq 0$ such that the λ-AI holds.*

We will sometimes consider an equivalent definition for AI, as $\sum_{\sigma \in \Sigma} \lambda_\sigma h(\sigma) \geq h(\tau)$, where $\lambda_\sigma \geq 0$ are coefficients, one for each $\sigma \in \Sigma$; these two definitions are equivalent, by taking $\lambda = \max_\sigma \lambda_\sigma$. Notice that both EI and AI are preserved under subsets of K in the sense that $K_1 \subseteq K_2$ and $K_2 \models_x (\Sigma \Rightarrow \tau)$ implies $K_1 \models_x (\Sigma \Rightarrow \tau)$, for $x \in \{EI, AI\}$.

AI always implies EI. Indeed, $h(\tau) \leq \lambda \cdot h(\Sigma)$ and $h(\Sigma) = 0$, implies $h(\tau) \leq 0$, which further implies $h(\tau) = 0$, because $h(\tau) \geq 0$ for every CI τ, and every polymatroid h. In this paper we study the reverse.

▶ **Definition 4.** *Let $\mathcal{I}$ be a syntactically-defined class of implication statements $(\Sigma \Rightarrow \tau)$, and let $K \subseteq \Gamma_n$. We say that $\mathcal{I}$* admits a relaxation *in K if, every implication statement $(\Sigma \Rightarrow \tau)$ in $\mathcal{I}$ that holds exactly, also holds approximately: $K \models_{EI} \Sigma \Rightarrow \tau$ implies $K \models_{AI} \Sigma \Rightarrow \tau$. We say that $\mathcal{I}$ admits a* λ-relaxation *if every EI admits a λ-AI.*

▶ **Example 5.** Let Σ=$\{(A; B|\emptyset), (A; C|B)\}$ and τ=$(A; C|\emptyset)$. Since $I_h(A; C|\emptyset) \leq I_h(A; B|\emptyset) + I_h(A; C|B)$ by the chain rule (4), then the exact implication $\Gamma_n \models_{EI} \Sigma \Rightarrow \tau$ admits a 1-AI.

4 Relaxation for FDs and MVDs: Always Possible

In this section we consider the implication problem where the antecedents are either saturated CIs, or conditionals. This is a case of special interest in databases, because the constraints correspond to MVDs, or FDs. Recall that a CI $(B; C|A)$ is *saturated* if $ABC = \Omega$ (i.e., the set of all attributes). Our main result in this section is:

▶ **Theorem 6.** *Assume that each formula in Σ is either saturated, or a conditional, and let τ be an arbitrary CI. Assume $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$. Then:*
1. *$\Gamma_n \models \frac{n^2}{4} h(\Sigma) \geq h(\tau)$.*
2. *If τ is a conditional, $Z \to X$, then $\Gamma_n \models h(\Sigma) \geq h(\tau)$.*

Before we prove the theorem, we list two important consequences.

▶ **Corollary 7.** *Let Σ consist of saturated CIs and/or conditionals, and let τ be any CI. Then $\mathcal{S}_n \models \Sigma \Rightarrow_{EI} \tau$ implies $\Gamma_n \models \Sigma \Rightarrow_{EI} \tau$*

Proof. If $\mathcal{S}_n \models \Sigma \Rightarrow_{EI} \tau$ then $\forall h \in \Gamma_n$, $h(\tau) \leq \frac{n^2}{4} h(\Sigma)$, thus $h(\Sigma) = 0$ implies $h(\tau) = 0$. ◀

The corollary has an immediate application to the inference problem in graphical models [12]. There, the problem is to check if every probability distribution that satisfies all CIs in Σ also satisfies the CI τ; we have seen that this is equivalent to $\Gamma_n^* \models_{EI} \Sigma \Rightarrow \tau$. The corollary states that it is enough that this implication holds on all of the uniform 2-tuple distributions, i.e. $\mathcal{S}_n \models \Sigma \Rightarrow_{EI} \tau$, because this implies the (even stronger!) statement $\Gamma_n \models \Sigma \Rightarrow_{EI} \tau$. Decidability was already known: Geiger and Pearl [12] proved that the set of graphoid axioms is sound and complete for the case when both Σ and τ are saturated, while Gyssens et al. [14] improve this by dropping any restrictions on τ.

The second consequence is the following:

▶ **Corollary 8.** *Let Σ, τ consist of saturated CIs and/or conditionals. Then the following two statements are equivalent:*

1. *The implication $\Sigma \Rightarrow \tau$ holds, where we interpret Σ, τ as MVDs and/or FDs.*
2. $\Gamma_n \models_{EI} \Sigma \Rightarrow \tau$.

Proof. We have shown right after Lemma 1 that (2) implies (1). For the opposite direction, by Th. 6, we need only check $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$, which holds because on every uniform probability distribution a saturated CI holds iff the corresponding MVD holds, and similarly for conditionals and FDs. Since the 2-tuple relation satisfies the implication for MVDs+FDs, it also satisfies the implication for CIs, proving the claim. ◀

Wong et al. [38] have proven that the implication for MVDs is equivalent to that of the corresponding saturated CIs (called there BMVD); they did not consider FDs. For the proof in the hard direction, they use the sound and complete axiomatization of MVDs in [2]. In contrast, our proof is independent of any axiomatic system, and is also much shorter. Finally, we notice that the corollary also implies that, in order to check an implication between MVDs and/or FDs, it suffices to check it on all 2-tuple databases: indeed, this is equivalent to checking $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$, because this implies Item (2), which in turn implies item (1). This rather surprising fact was first proven in [30].

We now turn to the proof of Theorem 6. Before proceeding, we note that we can assume w.l.o.g. that Σ consists only of saturated CIs. Indeed, if Σ contains a non-saturated term, then by assumption it is a conditional, $X \to Y$, and we will replace it with two saturated terms: $(Y; Z|X)$ and $XZ \to Y$, where $Z = \Omega \setminus XY$. Denoting Σ' the new set of formulas, we have $h(\Sigma) = h(\Sigma')$, because $h(Y|X) = I_h(Y; Z|X) + h(Y|XZ)$. Thus, we will assume w.l.o.g. that all formulas in Σ are saturated.

Theorem 6 follows from the next result, which is also of independent interest. We say that a CI $(X; Y|Z)$ is *elemental* if $|X| = |Y| = 1$. We say that σ *covers* τ if all variables in τ are contained in σ; for example $\sigma = (abc; d|e)$ covers $\tau = (cd; be)$. Then:

▶ **Theorem 9.** *Let τ be an elemental CI, and suppose each formula in Σ covers τ. Then $\mathcal{S}_n \models_{EI} (\Sigma \Rightarrow \tau)$ implies $\Gamma_n \models h(\tau) \leq h(\Sigma)$.*

Notice that this result immediately implies Item (1) of Theorem 6, because every $\tau = (Y; Z|X)$ can be written as a sum of $|Y| \cdot |Z| \leq n^2/4$ elemental terms (by the chain rule). In what follows we prove Theorem 9, then use it to prove item (2) of Theorem 6.

Finally, we consider whether (1) of Theorem 6 can be strengthened to a 1-relaxation; we give in Th. 11 below a sufficient condition, whose proof uses the notion of I-measure [40] and is included in the full paper [18], and leave open the question whether 1-relaxation holds in general for implications where the antecedents are saturated CIs and conditionals.

▶ **Definition 10.** *We say that two CIs $(X; Y|Z)$ and $(A; B|C)$ are* disjoint *if at least one of the following four conditions holds: (1) $X \subseteq C$, (2) $Y \subseteq C$, (3) $A \subseteq Z$, or (4) $B \subseteq Z$.*

If $\tau = (X; Y|Z)$ and $\sigma = (A; B|C)$ are disjoint, then for any step function h_W, it cannot be the case that both $h_W(\tau) \neq 0$ and $h_W(\sigma) \neq 0$. Indeed, if such W exists, then $Z, C \subseteq W$ and, assuming (1) $X \subseteq C$ (the other three cases are similar), we have $ZX \subseteq W$ thus $h_W(\tau) = 0$.

▶ **Theorem 11.** *Let Σ be a set of saturated, pairwise disjoint CI terms (Def. 10), and τ be a saturated mutual information. Then, $\mathcal{S}_n \models_{EI} (\Sigma \Rightarrow \tau)$ implies $\Gamma_n \models h(\tau) \leq h(\Sigma)$.*

4.1 Proof of Theorem 9

The following holds by the chain rule (proof in the appendix), and will be used later on.

▶ **Lemma 12.** *Let $\sigma = (A; B|C)$ and $\tau = (X; Y|Z)$ be CIs such that $X \subseteq A$, $Y \subseteq B$, $C \subseteq Z$ and $Z \subseteq ABC$. Then, $\Gamma_n \models h(\tau) \leq h(\sigma)$.*

We now prove theorem 9. We use lower case for single variables, thus $\tau = (x; y|Z)$ because it is elemental. We may assume w.l.o.g. $x, y \notin Z$ (otherwise $I_h(x; y|Z) = 0$ and the lemma holds trivially). The *deficit* of an elemental CI $\tau = (x; y|Z)$ is the quantity $|\Omega - Z|$. We prove by induction on the deficit of τ that $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$ implies $\Gamma_n \models h(\tau) \leq h(\Sigma)$.

Assume $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$, and consider the step function at Z. Since $h_Z(\tau) = 1$, there exists $\sigma \in \Sigma$, $\sigma = (A; B|C)$, such that $h_Z(\sigma) = 1$; this means that $C \subseteq Z$, and $A, B \not\subseteq Z$. In particular $x, y \notin C$, therefore $x, y \in AB$, because σ covers τ. If $x \in A$ and $y \in B$ (or vice versa), then $\Gamma_n \models h(\tau) \leq h(\sigma)$ by Lemma 12, proving the theorem. Therefore we assume w.l.o.g. that $x, y \in A$ and none is in B. Furthermore, since $B \not\subseteq Z$, there exists $u \in B - Z$.

Base case: τ is saturated. Then $u \notin xyZ$, contradicting the assumption that τ is saturated; in other words, in the base case, it is the case that $x \in A$ and $y \in B$.

Step: Let $Z_A = Z \cap A$, and $Z_B = Z \cap B$. Since $C \subseteq Z$, and $\sigma = (A; B|C)$ covers τ, then $Z = Z_A Z_B C$. We also write $A = xyA'Z_A$ (since $x, y \in A$) and $B = uB'Z_B$. So, we have that $\sigma = (A; B|C) = (xyA'Z_A; uB'Z_B|C)$, and we use the chain rule to define σ_1, σ_2:

$$h(\sigma) = I_h(xyA'Z_A; uB'Z_B|C) = I_h(\underbrace{xyA'Z_A; uZ_B|C}_{\stackrel{\text{def}}{=}\sigma_1}) + I_h(\underbrace{xyA'Z_A; B'|uCZ_B}_{\stackrel{\text{def}}{=}\sigma_2})$$

We also partition Σ s.t. $h(\Sigma) = h(\sigma_1) + h(\Sigma_2)$, where $\Sigma_2 \stackrel{\text{def}}{=} (\Sigma \setminus \{\sigma\}) \cup \{\sigma_2\}$.
Next, define $\tau' \stackrel{\text{def}}{=} (x; uy|Z)$ and use the chain rule to define τ_1, τ_2:

$$h(\underbrace{x; y|Z}_{\tau}) \leq I_h(\underbrace{x; uy|Z}_{\tau'}) = I_h(\underbrace{x; u|Z}_{\stackrel{\text{def}}{=}\tau_1}) + I_h(\underbrace{x; y|uZ}_{\stackrel{\text{def}}{=}\tau_2}) \tag{7}$$

By Lemma 12, $\Gamma_n \models h(\sigma_1) \geq h(\tau_1)$. We will prove: $\mathcal{S}_n \models_{EI} \Sigma_2 \Rightarrow \tau_2$. This implies the theorem, because Σ_2 is saturated, and by the induction hypothesis $\Gamma_n \models h(\Sigma_2) \geq h(\tau_2)$ (since the deficit of τ_2 is one less than that of τ), and the theorem follows from $h(\Sigma) = h(\sigma_1) + h(\Sigma_2) \geq h(\tau_1) + h(\tau_2) = h(\tau') \geq h(\tau)$. It remains to prove $\mathcal{S}_n \models_{EI} \Sigma_2 \Rightarrow \tau_2$, and we start with a weaker claim:

▷ Claim 13. $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau_2$.

Proof. By Lemma 12 we have that $h(\sigma) = I_h(xyA'Z_A; uB'Z_B|C) \geq I_h(xy; u|Z) = I_h(y; u|Z) + I_h(x; u|yZ)$. Therefore, $\Sigma \Rightarrow (x; u|yZ)$. Since $\Sigma \Rightarrow (x; y|Z)$, then by the chain rule we have that $\Sigma \Rightarrow (x; uy|Z) = \tau'$, and the claim follows from (7). ◁

Finally, we prove $\mathcal{S}_n \models_{EI} \Sigma_2 \Rightarrow \tau_2$. Assume otherwise, and let h_W be a step function such that $h_W(\tau_2) = I_{h_W}(x; y|uZ) = 1$, and $h_W(\Sigma_2) = 0$. This means that $uZ \subseteq W$. Therefore $uZ_B \subseteq W$, implying $I_{h_W}(xyA'Z_A; uZ_B|C) = h_W(\sigma_1) = 0$ (because $uZ_BC \subseteq uZ$). Therefore, $h_W(\Sigma) = h_W(\sigma_1) + h_W(\Sigma_2) = 0$, contradicting the fact that $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau_2$.

4.2 Proof of Theorem 6 Item 2

▶ **Lemma 14.** *Suppose $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$, where $\tau = (X;Y|Z)$. Let $\sigma \in \Sigma$ such that τ, σ are disjoint (Def. 10). Then: $\mathcal{S}_n \models_{EI} \big(\Sigma \backslash \{\sigma\}\big) \Rightarrow \tau$.*

Proof. Let $\Sigma' \stackrel{\text{def}}{=} \Sigma \setminus \{\sigma\}$. Assume by contradiction that there exists a step function h_W such that $h_W(\Sigma') = 0$ and $h_W(\tau) = 1$. Since σ, τ are disjoint, $h_W(\sigma) = 0$. Then $h_W(\Sigma) = 0$, contradicting the assumption $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$. ◀

▶ **Lemma 15.** *Let Σ be a set of saturated CIs s.t. $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$. Suppose $\tau = (Z \rightarrow uX)$ (which, recall, is a shorthand for $(uX; uX|Z)$), and define $\tau_1 = (Z \rightarrow u)$, $\tau_2 = (uZ \rightarrow X)$; thus, $h(\tau) = h(\tau_1) + h(\tau_2)$. Then there exists Σ_1 and Σ_2 such that: (1) $h(\Sigma) = h(\Sigma_1) + h(\Sigma_2)$; we say that Σ_1, Σ_2 form a parition of Σ. (2) Σ_1 covers τ_1 and $\mathcal{S}_n \models_{EI} \Sigma_1 \Rightarrow \tau_1$. (3) Σ_2 is saturated and $\mathcal{S}_n \models \Sigma_2 \Rightarrow \tau_2$.*

Proof. We partition Σ into Σ_1 and Σ_2 as follows. For every $\sigma = (A; B|C) \in \Sigma$, if $u \in C$ then we place σ in Σ_2. Otherwise, assume w.l.o.g that $u \in A$, and we write $A = uA_ZA_XA'$ where $A_Z = A \cap Z$, $A_X = A \cap X$, and $A' = A \backslash \{uA_ZA_X\}$. We use the chain rule to define σ_1, σ_2:

$$I_h(A; B|C) = I_h(uA_ZA_XA'; B|C) = I_h(\underbrace{uA_Z; B|C}_{\stackrel{\text{def}}{=}\sigma_1}) + I(\underbrace{A_XA'; B|uA_ZC}_{\stackrel{\text{def}}{=}\sigma_2}) \tag{8}$$

We place σ_1 in Σ_1, and σ_2 in Σ_2. We observe that σ_1 covers τ_1 (because $Z = A_ZB_ZC_Z \subseteq A_ZBC$) and σ_2 is saturated. Furthermore, $h(\Sigma_1) + h(\Sigma_2) = h(\Sigma)$. We prove $\Sigma_1 \models_{EI} \tau_1$. By assumption, $\Sigma \models_{EI} \tau_1 = (Z \rightarrow u)$. Let any $\sigma_2 = (A; B|C) \in \Sigma_2$; since $u \in C$, by Lemma 14 we can remove it, obtaining $\Sigma \setminus \{\sigma_2\} \models_{EI} \tau_1$; repeating this process proves $\Sigma_1 \models_{EI} \tau_1$. Finally, we prove $\Sigma_2 \models_{EI} \tau_2$. By assumption, $\Sigma \models_{EI} \tau_2 = (uZ \rightarrow X)$. Let any $\sigma_1 = (uA_Z; B|C) \in \Sigma_1$; since $uA_Z \subseteq uZ$, by Lemma 14 we can remove it, obtaining $\Sigma \setminus \{\sigma_1\} \models_{EI} \tau_2$; repeating this process proves $\Sigma_2 \models_{EI} \tau_2$. ◀

We now complete the proof of Theorem 6 item 2. Let $\tau = (Z \rightarrow X)$, and Σ be saturated. We show, by induction on $|X|$, that if $\mathcal{S}_n \models_{EI} \Sigma \Rightarrow \tau$ then $\Gamma_n \models h(\tau) \leq h(\Sigma)$. If $|X| = 1$, then $X = \{x\}$, $h(x|Z) = I(x; x|Z)$ is elemental, and the claim follows from Th. 9. Otherwise, let u be any variable in X, write $\tau = (Z \rightarrow uX')$, and apply Lemma 15 to $\tau_1 = (Z \rightarrow u)$, $\tau_2 = (Zu \rightarrow X')$, which gives us a partition of Σ into Σ_1, Σ_2. On one hand, $\mathcal{S}_n \models_{EI} \Sigma_1 \Rightarrow \tau_1$, and from Th. 9 we derive $h(\tau_1) \leq h(\Sigma_1)$ (because τ_1 is elemental, and covered by Σ_1); on the other hand $\mathcal{S}_n \models_{EI} \Sigma_2 \Rightarrow \tau_2$ where Σ_2 is saturated, which implies, by induction, $h(\tau_2) \leq h(\Sigma_2)$. The result follows from $h(\tau) = h(\tau_1) + h(\tau_2) \leq h(\Sigma_1) + h(\Sigma_2) = h(\Sigma)$, completing the proof.

5 Relaxation for General CIs: Sometimes Impossible

We consider the relaxation problem for arbitrary Conditional Independence statements. Recall that our golden standard is to check (in)equalities forall entropic functions, $h \in \Gamma_n^*$. As we saw, for MVD+FDs, these (in)equalities coincide with those satisfied by $\mathcal{S}_n$, and with those satisfied by Γ_n. In general, however, they differ. We start with an impossibility result, then prove that relaxation with an arbitrarily small error term always exists. Both results are for the topological closure, $\mathbf{cl}\,(\Gamma_n^*)$. This makes the negative result stronger, but the positive result weaker; it is unlikely for the positive result to hold for Γ_n^*, see [17, Sec.V.(A)] and Appendix A.

▶ **Theorem 16.** *There exists* Σ, τ *with four variables, such that* $\mathbf{cl}(\Gamma_4^*) \models_{EI} (\Sigma \Rightarrow \tau)$ *and* $\mathbf{cl}(\Gamma_4^*) \not\models_{AI} (\Sigma \Rightarrow \tau)$.

For the proof, we adapt an example by Kaced and Romashchenko [17, Inequality ($\mathcal{I}5'$) and Claim 5], built upon an earlier example by Matúš [25]. Let Σ and τ be the following:

$$\Sigma = \{(C;D|A),(C;D|B),(A;B),(B;C|D)\} \qquad \tau = (C;D) \tag{9}$$

We first prove that, for any $\lambda \geq 0$, there exists an entropic function h such that:

$$I_h(C;D) > \lambda \cdot (I_h(C;D|A) + I_h(C;D|B) + I_h(A;B) + I_h(B;C|D)) \tag{10}$$

Indeed, consider the distribution shown in Fig. 1 (c) (from [17]). By direct calculation, $I_h(C;D) = \varepsilon + O(\varepsilon^2) = \Omega(\varepsilon)$, while $I_h(C;D|A) = I_h(C;D|B) = I_h(A;B) = 0$ and $I_h(B;C|D) = O(\varepsilon^2)$ and we obtain Eq.(10) by choosing ε small enough. Next, we prove $\mathbf{cl}(\Gamma_n^*) \models_{EI} (\Sigma \Rightarrow \tau)$. Matúš [25] proved the following[5] $\forall h \in \Gamma_n^*$ and $\forall k \in \mathbb{N}$:

$$I_h(C;D) \leq I_h(C;D|A) + \frac{k+3}{2} I_h(C;D|B) + I_h(A;B) + \frac{k-1}{2} I_h(B;C|D) + \frac{1}{k} I_h(B;D|C) \tag{11}$$

The inequality obviously holds for $\mathbf{cl}(\Gamma_n^*)$ too. The EI follows by taking $k \to \infty$. Inequality (11) is almost a relaxation of the implication (9): the only extra term is the last term, which can be made arbitrarily small by increasing k. Our second result generalizes this:

▶ **Theorem 17.** *Let* Σ, τ *be arbitrary CIs, and suppose* $\mathbf{cl}(\Gamma_n^*) \models \Sigma \Rightarrow \tau$*. Then, for every* $\varepsilon > 0$ *there exists* $\lambda > 0$ *such that, forall* $h \in \mathbf{cl}(\Gamma_n^*)$*:*

$$h(\tau) \leq \lambda \cdot h(\Sigma) + \varepsilon \cdot h(\Omega) \tag{12}$$

Intuitively, the theorem shows that every EI can be relaxed in $\mathbf{cl}(\Gamma_n^*)$, if one allows for an error term, which can be made arbitrarily small. We notice that the converse of the theorem always holds: if $h(\Sigma) = 0$, then (12) implies $h(\tau) \leq \varepsilon \cdot h(\Omega)$, $\forall \varepsilon > 0$, which implies $h(\tau) = 0$.

Proof of Theorem 17. For the proof we need a brief review of cones [37, 4]. A set $C \subseteq \mathbb{R}^N$ is *convex* if, for any two points $x_1, x_2 \in C$ and any $\theta \in [0,1]$, $\theta x_1 + (1-\theta)x_2 \in C$; and it is called a *cone*, if for every $x \in C$ and $\theta \geq 0$ we have that $\theta x \in C$. The *conic hull* of C, $\mathbf{conhull}(C)$, is the set of vectors of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $x_1, \ldots, x_k \in C$ and $\theta_i \geq 0, \forall i \in [k]$. A cone K is *finitely generated* if $K = \mathbf{conhull}(L)$ for some finite set $L \subset \mathbb{R}^N$, and is *polyhedral* if there exists $u_1, \ldots, u_r \in \mathbb{R}^N$ s.t. $K = \{x \mid u_i \cdot x \geq 0, i \in [r]\}$; a cone is finitely generated iff it is polyhedral. For any $K \subseteq \mathbb{R}^N$, the *dual* is the set $K^* \subseteq \mathbb{R}^N$ defined as:

$$K^* \stackrel{\text{def}}{=} \{y \mid \forall x \in K, x \cdot y \geq 0\} \tag{13}$$

K^* represents the linear inequalities that hold for all $x \in K$, and is always a closed, convex cone (it is the intersection of closed half-spaces). We warn that the $*$ in Γ_n^* does *not* represent the dual; the notation Γ_n^* for entropic functions is by now well established, and we adopt it here too, despite it's clash with the standard notation for the dual cone. The following are known properties of cones (reviewed and proved in the Appendix):

(A) For any set K, $\mathbf{cl}\left(\mathbf{conhull}(K)\right) = K^{**}$.

[5] Matus [25] proved $I(C;D) \leq I(C;D|A) + I(C;D|B) + I(A;B) + I(C;E|B) + \frac{1}{k}I(B;E|C) + \frac{k-1}{2}(I(B;C|D) + I(C;D|B))$. Inequality (11) follows by setting $E = D$.

(B) If L is a finite set, then $\mathbf{conhull}(L)$ is closed.

(C) If K_1 and K_2 are closed, convex cones then: $(K_1 \cap K_2)^* = \Big(\mathbf{cl}\left(\mathbf{conhull}\left(K_1^* \cup K_2^*\right)\right)\Big)$. ◀

Theorem 17 follows from a more general statement about cones:

▶ **Theorem 18.** *Let $K \subseteq \mathbb{R}^N$ be a closed, convex cone, and let $y_1, \ldots, y_m, y$ be $m+1$ vectors in $\mathbb{R}^N$. The following are equivalent:*

(a) *For every $x \in K$, if $x \cdot y_1 \leq 0, \ldots, x \cdot y_m \leq 0$ then $x \cdot y \leq 0$.*

(b) *For every $\varepsilon > 0$ there exists $\theta_1, \ldots, \theta_m \geq 0$ and an error vector $e \in \mathbb{R}^N$ such that $||e||_\infty < \varepsilon$ and, for every $x \in K$, $x \cdot y \leq \theta_1 x \cdot y_1 + \cdots + \theta_m x \cdot y_m + x \cdot e$.*

Proof. Let $L \stackrel{\text{def}}{=} \{-y_1, -y_2, \ldots, -y_m\}$. Statement (a) is equivalent to $-y \in (K \cap L^*)^*$. Consider statement (b). It asserts $\forall \varepsilon > 0, \exists ||e||_\infty < \varepsilon$ such that

$$\exists \theta_1 \geq 0, \ldots, \exists \theta_k \geq 0, \underbrace{\forall x \in K, x \cdot y \leq x \cdot (\sum \theta_i y_i + e)}_{-y+\sum \theta_i y_i + e \in K^*}$$

In other words, $-y + e \in \mathbf{conhull}(K^* \cup L)$ and, since this must hold for arbitrarily small $||e||_\infty$, statement (b) is equivalent to $-y \in \mathbf{cl}\left(\mathbf{conhull}(K^* \cup L)\right)$. We prove equivalence of (a) and (b):

$$\begin{aligned}
(K \cap L^*)^* &= \mathbf{cl}\left(\mathbf{conhull}\left(K^* \cup L^{**}\right)\right) && \text{Item (C)} \\
&= \mathbf{cl}\left(\mathbf{conhull}\left(K^* \cup \mathbf{cl}\left(\mathbf{conhull}(L)\right)\right)\right) && \text{Item (A)} \\
&= \mathbf{cl}\left(\mathbf{conhull}\left(K^* \cup \mathbf{conhull}(L)\right)\right) && \text{Item (B)} \\
&= \mathbf{cl}\left(\mathbf{conhull}(K^* \cup L)\right) && \text{Def. of } \mathbf{conhull}(-)
\end{aligned}$$ ◀

We now prove Theorem 17, using the fact that $K \stackrel{\text{def}}{=} \mathbf{cl}(\Gamma_n^*)$ is a closed cone [40]. Let $\Sigma = \{\sigma_1, \ldots, \sigma_m\}$. Associate to each term $\sigma_i = (B_i; C_i | A_i)$ the vector $y_i \in \mathbb{R}^{2^n}$ such that, forall $h \in \mathbb{R}^{2^n}$, $h \cdot y_i = I_h(B_i; C_i | A_i) = h(A_i B_i) + h(A_i C_i) - h(A_i B_i C_i) - h(C_i)$ (i.e. y_i has two coordinates equal to $+1$, and two equal to -1), for $i = 1, m$. Denote by y the similar vector associated to τ. To prove Theorem 17, let $\varepsilon > 0$. By assumption, $\mathbf{cl}(\Gamma_n^*) \models \Sigma \Rightarrow \tau$, thus condition (a) of Th. 18 holds, and this implies condition (b), where we choose e such that $||e||_\infty < \varepsilon/2^n$. Then, condition (b) becomes:

$$h(\tau) = h \cdot y \leq \sum_i \theta_i h \cdot y_i + h \cdot e = \sum_i \theta_i h(\sigma_i) + \sum_{W \subseteq [n]} |e_W| h(W) \leq \lambda h(\Sigma) + \varepsilon h(\Omega)$$

where $\lambda = \max_i \theta_i$. This completes the proof of Theorem 17.

6 Restricted Axioms

The characterization of the entropic cone $\mathbf{cl}(\Gamma_n^*)$ is currently an open problem [40]. In other words, there is no known decision procedure capable of deciding whether an exact or approximate implication holds for all entropic functions. In this section, we consider implications that can be inferred using only the Shannon inequalities (e.g., (2), and (3)), and thus hold for all polymatroids $h \in \Gamma_n$. Several tools exists (e.g. ITIP or XITIP [39]) for checking such inequalities.

This study is important for several reasons. First, by restricting to Shannon inequalities we obtain a sound, but in general incomplete method for deciding implications. All axioms for reasoning about MVD, FD, or semi-graphoid axioms[6] [2, 27, 12] are, in fact, based on Shannon inequalities. Second, under some syntactic restrictions, they are also complete; as we saw, they are complete for MVD and/or FDs, for saturated constraints and/or conditionals, and also for marginal constraints [12]. Third, Shannon inequalities are complete for reasoning for a different class of constraints, called *measure-based constraints*, which were introduced by Sayrafi et al. [32] (where Γ_n is denoted by $\mathcal{M}_{\mathtt{SI}}$) and shown to have a variety of applications.

We start by showing that every exact implication of CIs can be relaxed over Γ_n. This result was known, e.g. [17]; we re-state and prove it here for completeness.

▶ **Theorem 19.** *Let Σ, τ be arbitrary CIs. If $\Gamma_n \models_{EI} \Sigma \Rightarrow \tau$, then there exist $\lambda \geq 0$, s.t. $\Gamma_n \models h(\tau) \leq \lambda \cdot h(\Sigma)$. In other words, CIs admit relaxation over Γ_n.*

Proof. (Sketch) We set $K = \Gamma_n$ in Th. 18. Then K is polyhedral, hence K^* is finitely generated. Therefore, in the proof of Th. 18, the set $K^* \cup L$ is finitely generated, hence **conhull** $(K^* \cup L)$ is closed, therefore there is no need for an error vector e in Statement (b) of Th. 18, and, hence, no need for ε in AI (12) ◀

It follows that Shannon inequalities are incomplete for proving the implication $\Sigma \Rightarrow \tau$, where Σ, τ are given by Eq. (9). This is a "non-Shannon" exact implication, i.e. it holds only in **cl** (Γ_n^*), but fails in Γ_n, otherwise it would admit a relaxation. The explanation is that Matus' inequality (11) is a non-Shannon inequality. (The first example of a non-Shannon inequality is due to Yeung and Zhang [42].) Next, we turn our attention to the size of the factor λ. We prove a lower bound of 3:

▶ **Theorem 20** ([9])**.** *The following inequality holds for all polymatroids $h \in \Gamma_n$:*

$$h(Z) \leq I_h(A;B|C) + I_h(A;B|D) + I_h(C;D|E) + I_h(A;E) + 3h(Z|A) + 2h(Z|B) \tag{14}$$

but the inequality fails if any of the coefficients $3, 2$ are replaced by smaller values. In particular, denoting τ, Σ the terms on the two sides of Eq.(14), the exact implication $\Gamma_n \models_{EI} \Sigma \Rightarrow \tau$ holds, and does not have a 1-relaxation.

We have checked the two claims in the theorem using the ITIP[7] tool. For the positive result, we also provide direct (manual) proof in the full version of this paper [18]. Since some EIs relax only with $\lambda \geq 3$, the next question is, how large does λ need to be? We prove this upper bound in [18]:

▶ **Theorem 21.** *If $\Gamma_n \models \Sigma \Rightarrow \tau$ then $\Gamma_n \models \tau \leq (2^n)! \cdot h(\Sigma)$. In other words, every implication of CIs admits a $(2^n)!$-relaxation over Γ_n.*

7 Restricted Models

In this section we restrict ourselves to models of uniform 2-tuple distributions. Recall that their entropic functions are the step functions, $\mathcal{S}_n$. Denoting their conic hull by $\mathcal{P}_n \stackrel{\text{def}}{=}$ **conhull** $(\mathcal{S}_n)$, we prove here that all EI's admit a 1-relaxation on $\mathcal{P}_n$. This study has two motivations. First, it leads to a complete, but unsound procedure for implication. A

[6] Semi-graphoid axioms restricted to "strictly positive" distributions, which fail Γ_n^*.
[7] `http://user-www.ie.cuhk.edu.hk/~ITIP/`

model checking system may verify an EI or AI by checking it on all 2-tuple distributions. As we saw in Sec. 4 this procedure is sound and complete for saturated or conditional CI's, but it may be unsound in general, for example the inequality $I_h(X;Y|Z) \leq I_h(X;Y)$ holds for all step functions, but fails on the "parity function" in Fig. 1 (b). Second, this model checking procedure is sound and complete in an important application, namely for checking *differential constraints* in market basket analysis [33]. Differential constraints are more general than the CIs we discussed so far, yet we prove here that they, too, admit a 1-relaxation in $\mathcal{P}_n$. Thus, our relaxation result has immediate application to market basket constraints.

Consider a set of items $\Omega = \{X_1, \ldots, X_n\}$, and a set of baskets $\mathcal{B} = \{b_1, \ldots, b_N\}$ where every basket is a subset $b_i \subseteq \Omega$. The *support function* $f : 2^\Omega \to \mathbb{N}$ assigns to every subset $W \subseteq \mathcal{B}$ the number of baskets in $\mathcal{B}$ that contain the set W: $f(W) = |\{i \mid i \in [N], W \subseteq b_i \in \mathcal{B}\}|$. A constraint $f(W) = f(WX)$ asserts that every basket that contains W also contains X. Sayrafi and Van Gucht [33] define the *density* of a function $f : 2^\Omega \to \mathbb{N}$ as $d_f(W) \stackrel{\text{def}}{=} \sum_{Z:W\subseteq Z}(-1)^{|Z-W|}f(Z)$; we show below this equals the number of baskets $b_i \in \mathcal{B}$ s.t. $W = b_i$. Then, they define a *differential constraint* to be a statement of the form $d_f(W) = 0$, for some $W \subseteq \Omega$, and study the implication problem of differential constraints.

We now explain the connection to step functions $\mathcal{S}_n$; for the purpose of this discussion we consider h_Ω to be a step function, which is $h_\Omega \equiv 0$ (Sec 2.2). Fix $i \in [N]$ and consider the single basket $b_i \in \mathcal{B}$. Define f_{b_i} to be the support function for the singleton set $\{b_i\}$, that is $f_{b_i}(W) = 1$ if $W \subseteq b_i$ and 0 otherwise. It follows that $h_{b_i}(W) \stackrel{\text{def}}{=} 1 - f_{b_i}(W)$ is precisely the step function at b_i. The support function for $\mathcal{B} = \{b_1, \ldots, b_N\}$ is $f = \sum_{i\in[N]} f_{b_i} = N - h$, where $h \stackrel{\text{def}}{=} \sum_{i\in[N]} h_{b_i} \in \mathcal{P}_n$. Thus, any support function f gives rise to a polymatroid $h \stackrel{\text{def}}{=} N - f \in \mathcal{P}_n$. By linearity, their densities are related by $d_f = d_N - d_h$, where d_N is the density of the constant function N: $d_N(W) = N \cdot \sum_{Z:W\subseteq Z\subseteq\Omega}(-1)^{|Z-W|}$, thus $d_N(\Omega) = N$ and $d_N(W) = 0$ for $W \subsetneq \Omega$; in particular, $d_f(W) = -d_h(W)$ for $W \subsetneq \Omega$. Conversely, any $h = \sum_{U\subsetneq\Omega} c_U h_U \in \mathcal{P}_n$, where $c_U \geq 0$, and any $N \geq \sum_U c_U$ gives rise to a set of baskets $\mathcal{B}$ of size N, where each set $U \subsetneq \Omega$ occurs exactly c_U times and Ω occurs exactly $N - \sum_U c_U$ times, such that the support function of $\mathcal{B}$ is $f = |\mathcal{B}| - h$. Therefore, the implication problem of differential constraints studied in [33] is equivalent to the implication problem for $\mathcal{P}_n$. We prove that the latter admits a 1-relaxation. We start with a lemma (proof in Appendix):

▶ **Lemma 22.** *Fix a function $h : 2^\Omega \to \mathbb{R}$ s.t. $h(\emptyset) = 0$. Then $h = \sum_{Z\subsetneq\Omega}(-d_h(Z)) \cdot h_Z$. In other words, the step functions h_Z form a basis for the vector space $\{h \in R^{2^n} \mid h(\emptyset) = 0\}$.*

Fix a step function, $h = h_W$. By the Lemma, h_W admits a unique decomposition $h_W = \sum_{Z\subsetneq\Omega}(-d_{h_W}(Z))h_Z$; it follows that $d_{h_W}(Z) = -1$ when $Z = W$ and $d_{h_W}(Z) = 0$ otherwise. In particular, $d_h \leq 0$ forall $h \in \mathcal{P}_n$. Fix a set of baskets $\mathcal{B} = \{b_1, \ldots, b_N\}$, and let f be its support function. We prove that $d_f(Z)$ is equal to the number of baskets b_i s.t. $Z = b_i$; in particular $d_f \geq 0$. Indeed, for $Z = \Omega$ this follows from the definition of the differential d_f, while for $Z \subsetneq \Omega$ we use the fact that $f = N - \sum_i h_{b_i}$ and $d_f(Z) = -\sum_i d_{h_{b_i}}(Z)$.

The quantity $I_h(y_1; y_2; \cdots; y_m|W) \stackrel{\text{def}}{=} -\sum_{Z:W\subseteq Z\subseteq\{y_1,\ldots,y_m\}}(-1)^{|Z-W|}h(Z)$ is called the *conditional multivariate mutual information*, thus, $-d_h(W)$ is a saturated conditional multivariate mutual information. We show in the full paper [18] that $-d_h(W)$ is precisely the I-measure of an atom in I-measure theory [40].

Once we have motivated the critical role of the negated densities $-d_h(W)$, we define an *I-measure constraint* to be an arbitrary sum $\sigma = -\sum_i d_h(W_i)$; the exact constraint is the assertion $\sigma = 0$, while an approximate constraint asserts some bound, $\sigma \leq c$. The differential constraints [33] are special cases of I-measure constraints. Any CI constraint is also a special case of an I-measure, for example $h(Y|X) = -\sum_{W:X\subseteq W, Y\not\subseteq W} d_h(W)$, and $I_h(Y;Z|X) = -\sum_{W:X\subseteq W, X\not\subseteq W, Y\not\subseteq W} d_h(W)$. Since $d_h \leq 0$ for $h \in \mathcal{P}_n$, it follows that all I-measure constraints are ≥ 0. We prove[8]:

▶ **Theorem 23.** *Exact implications of I-measure constraints admit a 1-relaxation in $\mathcal{P}_n$.*

Proof. Consider an implication $\Sigma \Rightarrow \tau$ where all constraints in Σ, τ are I-measure constraints. Let $\tau = -\sum_i d_h(W_i)$. Then, for every i, there exists some constraint $\sigma = -\sum_j d_h(W_j) \in \Sigma$ such that $W_i = W_j$ for some j, proving the theorem. If not, then for the step function $h \stackrel{\text{def}}{=} h_{W_i}$ we have $h(\sigma) = 0$ forall $\sigma \in \Sigma$, yet $h(\tau) = 1$, contradicting the assumption $P_n \models \Sigma \Rightarrow \sigma$. ◀

▶ **Example 24.** Consider Example 4.3 in [33]: $d_1 = f(A) + f(ABCD) - f(ABC) - f(ACD)$, $d_2 = f(C) - f(CD)$, and $d = f(AB) - f(ABD)$. Sayrafi and Van Gucht prove $d_1 = d_2 = 0$ implies $d = 0$ for all support functions f. The quantity d_1 represents the number of baskets that contain A, but do not contain BC nor CD, while d_2 is the number of baskets that contain C but not D. Our theorem converts the exact implication into an inequality as follows. Denote by $\sigma_1 \stackrel{\text{def}}{=} I_h(BC;CD|A), \sigma_2 \stackrel{\text{def}}{=} h(D|C), \tau \stackrel{\text{def}}{=} h(D|AB)$, $P_n \models (\sigma_1 = \sigma_2 = 0 \Rightarrow \tau = 0)$ relaxes to $P_n \models \sigma_1 + \sigma_2 \geq \tau$, which translates into $d_1 + d_2 \leq d$ forall support functions f.

8 Discussion and Future Work

Number of Repairs. A natural way to measure the degree of a constraint in a relation instance R is by the number of repairs needed to enforce the constraint on R. In the case of a key constraint, $X \to Y$, where $XY = \Omega$, our information-theoretic measure is naturally related to the number of repairs, as follows. If $h(Y|X) = c$, where h is the entropy of the empirical distribution on R, then one can check $|R|/|\Pi_X(R)| \leq 2^c$. Thus, the number of repairs $|R| - |\Pi_X(R)|$ is at most $(2^c - 1)|\Pi_X(R)|$. We leave for future work an exploration of the connections between number of repairs and information theoretic measures.

Small Model Property. We have proven in Sec. 4 that several classes of implications (including saturated CIs, FDs, and MVDs) have a "small model" property: if the implication holds for all uniform, 2-tuple distributions, then it holds in general. In other words, it suffices to check the implication on the step functions $\mathcal{S}_n$. One question is whether this small model property continues to hold for other tractable classes of implications in the literature. For example, Geiger and Pearl [12] give an axiomatization (and, hence, a decision procedure) for *marginal* CIs. However, marginal CIs do not have the same small model property. Indeed, the implication $(X \perp Y)\&(X \perp Z) \Rightarrow (X \perp YZ)$ holds for all uniform 2-tuple distributions (because $I_h(X;YZ) \leq I_h(X;Y) + I_h(X;Z)$ holds for all step functions), however it fails for the "parity distribution" in Fig.1(b). We leave for future work an investigation of the small model property for other classes of constraints.

Proof Techniques. Since we had to integrate concepts from both database theory and information theory, we had to make a choice of which proof techniques to favor. In particular, $\mathcal{P}_n$, the cone closure of the step functions, is better known in information theory as the *set of entropic functions with a non-negative I-measure*. After trying both alternatives, we have chosen to favor the step functions in most of the proofs, because of their connection to 2-tuple relations. We explain in the full paper [18] the connection to the I-measure, and include the proof of Th. 11, which is easier to express in that language.

[8] A version of this proof based on I-measure theory appears in the full version of the paper [18].

Bounds on the factor λ. In the early stages of this work we conjectured that all CIs in Γ_n admit 1-relaxation, until we discovered the counterexample in Th. 20, where $\lambda = 3$. On the other hand, the only general upper bound is $(2^n)!$. None of them is likely to be tight. We leave for future work the task of finding tighter bounds for λ.

References

1 William Ward Armstrong and Claude Delobel. Decomposition and Functional Dependencies in Relations. *ACM Trans. Database Syst.*, 5(4):404–430, 1980. doi:10.1145/320610.320620.

2 Catriel Beeri, Ronald Fagin, and John H. Howard. A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations. In *Proceedings of the 1977 ACM SIGMOD International Conference on Management of Data, Toronto, Canada, August 3-5, 1977.*, pages 47–61, 1977. doi:10.1145/509404.509414.

3 Tobias Bleifuß, Susanne Bülow, Johannes Frohnhofen, Julian Risch, Georg Wiese, Sebastian Kruse, Thorsten Papenbrock, and Felix Naumann. Approximate Discovery of Functional Dependencies for Large Datasets. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*, pages 1803–1812, 2016. doi:10.1145/2983323.2983781.

4 Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004. doi:10.1017/CBO9780511804441.

5 Toon Calders and Jan Paredaens. Axiomatization of Frequent Sets. In Jan Van den Bussche and Victor Vianu, editors, *Database Theory — ICDT 2001*, pages 204–218, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

6 Terence Chan. Recent Progresses in Characterising Information Inequalities. *Entropy*, 13(2):379–401, 2011. doi:10.3390/e13020379.

7 Xu Chu, Ihab F. Ilyas, Paolo Papotti, and Yin Ye. RuleMiner: Data quality rules discovery. In *IEEE 30th International Conference on Data Engineering, Chicago, ICDE 2014, IL, USA, March 31 - April 4, 2014*, pages 1222–1225, 2014. doi:10.1109/ICDE.2014.6816746.

8 Mehmet M. Dalkilic and Edward L. Roberston. Information Dependencies. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '00, pages 245–253, New York, NY, USA, 2000. ACM. doi:10.1145/335168.336059.

9 Randall Dougherty, Christopher F. Freiling, and Kenneth Zeger. Linear rank inequalities on five or more variables. *CoRR*, abs/0910.0284, 2009. arXiv:0910.0284.

10 Ronald Fagin. Horn clauses and database dependencies. *J. ACM*, 29(4):952–985, 1982. doi:10.1145/322344.322347.

11 Abram Friesen and Pedro Domingos. The Sum-Product Theorem: A Foundation for Learning Tractable Models. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1909–1918, New York, New York, USA, June 2016. PMLR. URL: http://proceedings.mlr.press/v48/friesen16.html.

12 Dan Geiger and Judea Pearl. Logical and Algorithmic Properties of Conditional Independence and Graphical Models. *The Annals of Statistics*, 21(4):2001–2021, 1993. URL: http://www.jstor.org/stable/2242326.

13 Chris Giannella and Edward L. Robertson. On approximation measures for functional dependencies. *Inf. Syst.*, 29(6):483–507, 2004. doi:10.1016/j.is.2003.10.006.

14 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.

15 Christian Herrmann. Corrigendum to "On the undecidability of implications between embedded multivalued database dependencies" [Inform. and Comput. 122(1995) 221-235]. *Inf. Comput.*, 204(12):1847–1851, 2006. doi:10.1016/j.ic.2006.09.002.

16 Ihab F. Ilyas and Xu Chu. Trends in Cleaning Relational Data: Consistency and Deduplication. *Foundations and Trends in Databases*, 5(4):281–393, 2015. doi:10.1561/1900000045.

17 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.

18 Batya Kenig and Dan Suciu. Integrity Constraints Revisited: From Exact to Approximate Implication. *CoRR*, abs/1812.09987, 2018. arXiv:1812.09987.

19 Juha Kontinen, Sebastian Link, and Jouko Väänänen. Independence in Database Relations. In Leonid Libkin, Ulrich Kohlenbach, and Ruy de Queiroz, editors, *Logic, Language, Information, and Computation*, pages 179–193, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

20 Sebastian Kruse and Felix Naumann. Efficient Discovery of Approximate Dependencies. *PVLDB*, 11(7):759–772, 2018. doi:10.14778/3192965.3192968.

21 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.

22 Chen Li, editor. *Proceedings of the Twenty-fourth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 13-15, 2005, Baltimore, Maryland, USA*. ACM, 2005. URL: http://dl.acm.org/citation.cfm?id=1065167.

23 David Maier. *Theory of Relational Databases.* Computer Science Pr, 1983.

24 Francesco M. Malvestuto. Statistical treatment of the information content of a database. *Inf. Syst.*, 11(3):211–223, 1986. doi:10.1016/0306-4379(86)90029-3.

25 F. Matúš. Infinitely Many Information Inequalities. In *2007 IEEE International Symposium on Information Theory*, pages 41–44, June 2007. doi:10.1109/ISIT.2007.4557201.

26 Alejandro Molina, Antonio Vergari, Nicola Di Mauro, Sriraam Natarajan, Floriana Esposito, and Kristian Kersting. Mixed Sum-Product Networks: A Deep Architecture for Hybrid Domains. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 3828–3835, 2018. URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16865.

27 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.

28 Jean-Philippe Pellet and André Elisseeff. Using Markov Blankets for Causal Structure Learning. *Journal of Machine Learning Research*, 9:1295–1342, 2008. doi:10.1145/1390681.1442776.

29 Hoifung Poon and Pedro M. Domingos. Sum-Product Networks: A New Deep Architecture. In *UAI 2011, Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, Barcelona, Spain, July 14-17, 2011*, pages 337–346, 2011. URL: https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=2194&proceeding_id=27.

30 Yehoshua Sagiv, Claude Delobel, D. Scott Parker, Jr., and Ronald Fagin. An Equivalence Between Relational Database Dependencies and a Fragment of Propositional Logic. *J. ACM*, 28(3):435–453, July 1981. doi:10.1145/322261.322263.

31 Babak Salimi, Johannes Gehrke, and Dan Suciu. Bias in OLAP Queries: Detection, Explanation, and Removal. In *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*, pages 1021–1035, 2018. doi:10.1145/3183713.3196914.

32 Bassem Sayrafi, Dirk Van Gucht, and Marc Gyssens. The implication problem for measure-based constraints. *Inf. Syst.*, 33(2):221–239, 2008. doi:10.1016/j.is.2007.07.005.

33 Bassem Sayrafi and Dirk Van Gucht. Differential Constraints. In *Proceedings of the Twenty-fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '05, pages 348–357, New York, NY, USA, 2005. ACM. doi:10.1145/1065167.1065213.

34 Yannis Sismanis, Paul Brown, Peter J. Haas, and Berthold Reinwald. GORDIAN: efficient and scalable discovery of composite keys. In *Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006*, pages 691–702, 2006. URL: `http://dl.acm.org/citation.cfm?id=1164187`.

35 Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

36 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.

37 Milan Studený. Convex cones in finite-dimensional real vector spaces. *Kybernetika*, 29(2):180–200, 1993. URL: `http://www.kybernetika.cz/content/1993/2/180`.

38 S. K. Michael Wong, Cory J. Butz, and Dan Wu. On the implication problem for probabilistic conditional independency. *IEEE Trans. Systems, Man, and Cybernetics, Part A*, 30(6):785–805, 2000. `doi:10.1109/3468.895901`.

39 Raymond W. Yeung. A framework for linear information inequalities. *IEEE Trans. Information Theory*, 43(6):1924–1934, 1997. `doi:10.1109/18.641556`.

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

41 Zhen Zhang and Raymond W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Information Theory*, 43(6):1982–1986, 1997. `doi:10.1109/18.641561`.

42 Zhen Zhang and Raymond W. Yeung. On Characterization of Entropy Function via Information Inequalities. *IEEE Trans. Information Theory*, 44(4):1440–1452, 1998. `doi:10.1109/18.681320`.

A Example for Section 5

Theorem 16 states that some EI does not relax to an AI. The example, based on [17], uses 4 random variables, hence it is essentially a statement about $\mathbb{R}^{15}$, making it difficult to visualize its underlying geometry. We give here a simpler counterexample, in $\mathbb{R}^3$, albeit in a vector space unrelated to information theory.

Let $K \stackrel{\text{def}}{=} \{\begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \mid x_1 \geq 0, x_3 \geq 0, x_1x_3 \geq x_2^2\}$ be the cone of semi-positive definite 2×2 matrices. This is known to be a convex cone, also called the *positive semi-definite cone* [4]. Equivalently, we view K as a cone in $\mathbb{R}^3$, namely $K = \{(x_1, x_2, x_3) \mid x_1 \geq 0, x_3 \geq 0, x_1x_3 \geq x_2^2\}$ Then K satisfies the following Exact Implication:

$$\forall(x_1, x_2, x_3) \in K : \quad x_1 \leq 0 \Rightarrow x_2 \leq 0$$

because $x_1 \leq 0$ is equivalent to $x_1 = 0$, implying $x_2^2 \leq 0$ thus $x_2 = 0$. However, K does not satisfy the corresponding Approximate Implication, more precisely the following is false:

$$\exists\lambda > 0, \forall(x_1, x_2, x_3) \in K : \quad x_2 \leq \lambda x_1 \qquad\qquad \text{(this is false)}$$

Indeed, for every choice of $\lambda > 0$, choose $0 < x_1 < 1/\lambda$, and let $x_2 = 1$, $x_3 = 1/x_1$. Then $(x_1, x_2, x_3) \in K$, yet $x_2 > \lambda x_1$.

Instead, Theorem 18 states that, for every $\varepsilon > 0$, there exists $\lambda > 0$, and an error term $\mathbf{e} = (e_1, e_2, e_3)$, with $e_1, e_2, e_3 < \varepsilon$, such that:

$$\forall\mathbf{x} \in K : \qquad\qquad x_2 \leq \lambda x_1 + e_1x_1 + e_2x_2 + e_3x_3$$

In our simple example, this statement is easily verified. Indeed, given $\varepsilon > 0$, define $\lambda \stackrel{\text{def}}{=} 1/\varepsilon$ and $\mathbf{e} \stackrel{\text{def}}{=} (0, 0, \varepsilon/4)$. Then $\lambda x_1 + (e_1x_1 + e_2x_2 + e_3x_3) = \frac{1}{\varepsilon}x_1 + \frac{\varepsilon}{4}x_3 \geq 2\sqrt{\frac{1}{4}x_1x_3} = \sqrt{x_1x_3} \geq$

$|x_2| \geq x_2$ proving the AI. This simple example explains why the error term is necessary in Theorem 18.

Geometrically, the error term is necessary whenever the cone $\textbf{conhull}(K^* \cup L)$ used in the proof of Theorem 18 is not closed. In our example $K^* = K$ because the positive, semi-definite cone is self-dual [4], and $L \stackrel{\text{def}}{=} \{(-1,0,0)\}$, and will check that $\textbf{conhull}(K^* \cup L) = \textbf{conhull}(K \cup L)$ is not closed. For that, consider the sequence $\mathbf{y}_n \stackrel{\text{def}}{=} (0,1,1/n)$. On one hand, we have $(n,1,1/n) \in K$, therefore $\mathbf{y}_n = (0,1,1/n) = (n,1,1/n) + n(-1,0,0) \in \textbf{conhull}(K \cup L)$. On the other hand, $\lim_{n\to\infty} \mathbf{y}_n = (0,1,0) \notin \textbf{conhull}(K \cup L)$. To see this, it suffices to notice that every vector in $\textbf{conhull}(K \cup L)$ has the form $(x_1 - \lambda, x_2, x_3)$ for some $(x_1, x_2, x_3) \in K$ and $\lambda \geq 0$, and, therefore, it satisfies the EI $x_3 = 0 \Rightarrow x_2 = 0$; our limit vector $(0,1,0)$ does not satisfy this EI, hence it is not in $\textbf{conhull}(K \cup L)$. This proves the fact that $\textbf{conhull}(K^* \cup L)$ is not closed, and hence taking its closure in the proof of Theorem 18 is necessary, leading to the error term.

B Proof of Cone Properties and Identities from Section 5

We need several known properties of cones; we give the proofs of some of them, for completeness, and refer for the others to [4].

▶ **Theorem 25.** *Let $K, K_1, K_2 \subseteq \mathbb{R}^n$. The following holds.*

1. $K_1 \subseteq K_2 \Rightarrow K_2^* \subseteq K_1^*$
2. $K_1 \subseteq K_2^*$ *iff* $K_1^* \supseteq K_2$.
3. *If* $K \neq \emptyset$ *then* $\textbf{cl}\left(\textbf{conhull}(K)\right) = K^{**}$.
4. $K^* = \left(\textbf{cl}\left(\textbf{conhull}(K)\right)\right)^*$.
5. *If K_1 and K_2 are closed, convex cones then:* $(K_1 \cap K_2)^* = \left(\textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)\right)$.
6. *A cone K is finitely generated iff K^* is finitely generated.*

Proof.

Proof of (1) Let $x \in K_2^*$, and let $y \in K_1$. Since $x \cdot z \geq 0$ for every vector $z \in K_2$, and since $K_2 \supseteq K_1$ then $x \cdot y \geq 0$ as well. Therefore, $x \in K_1^*$.

Proof of (2) Both statements assert $\forall x \in K_1, \forall y \in K_2, x{\cdot}y \geq 0$.

We omit the proofs of (3) and (4) and refer to [4].

Proof of (5) By definition of union we have that: $K_i^* \subseteq \textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)$ for $i \in \{1,2\}$. By item (1) we have that $K_i^{**} \supseteq \left(\textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)\right)^*$ for $i \in \{1,2\}$. Since K_1 and K_2 are closed convex cones then by item (3) it holds that $K_1^{**} = K_1$ and $K_2^{**} = K_2$. Therefore, for $i \in \{1,2\}$ we have that $K_i \supseteq \left(\textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)\right)^*$. From the above we get that $K_1 \cap K_2 \supseteq \left(\textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)\right)^*$. By property (1) we get that:
$(K_1 \cap K_2)^* \subseteq \left(\textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)\right)^{**}$. By property (3) we have that

$$\left(\textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)\right)^{**} = \textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right).$$

Overall, we get that $(K_1 \cap K_2)^* \subseteq \textbf{cl}\left(\textbf{conhull}(K_1^* \cup K_2^*)\right)$.

Proof of (6). Suppose K is finitely generated, $K = \textbf{conhull}(\{x_1, \ldots, x_n\})$. Then $K^* = \{y \mid x_1 \cdot y \geq 0, \ldots, x_n \cdot y \geq 0\}$, hence it is polyhedral by definition. ◀

C Proof of Lemma 22

Recall that $h(\Omega|Z) = h(\Omega) - h(Z)$. Define $\delta_h(W)$ to be the Möbius inverse of $h(\Omega|W)$, in other words:

$$\forall W : \delta_h(W) = \sum_{Z:W\subseteq Z} (-1)^{|Z-W|} h(\Omega|Z) \qquad \forall W : h(\Omega|W) = \sum_{Z:W\subseteq Z} \delta_h(Z) \tag{15}$$

We claim that, for $W \subsetneq \Omega$, $\delta_h(W) = -d_h(W)$. Indeed, $\delta_h(W) = \sum_{Z:W\subseteq Z}(-1)^{|Z-W|}h(\Omega|Z) = h(\Omega)\sum_{Z:W\subseteq Z\subseteq\Omega}(-1)^{|Z-W|} - d_h(W)$ and the claim follows from $\sum_{Z:W\subseteq Z\subseteq\Omega}(-1)^{|Z-W|} = 0$ when $W \subsetneq \Omega$. We prove that $h = \sum_{Z\subsetneq\Omega} \delta_h(Z) \cdot h_Z$, by using the right part of Eq.(15):

$$h(W) = h(\Omega|\emptyset) - h(\Omega|W) = \sum_Z \delta_h(Z) - \sum_{Z:W\subseteq Z} \delta_h(Z) = \sum_{Z:W\not\subseteq Z} \delta_h(W) = \sum_Z \delta_h(Z) \cdot h_Z(W)$$

because $h_Z(W) = 1$ iff $W \not\subseteq Z$, and $h(W) = 0$ otherwise. This proves that the $2^n - 1$ step functions span the vector space $\{h \in R^{2^n} \mid h(\emptyset) = 0\}$; since the latter has dimension $2^n - 1$, it follows that the step functions form a basis.