Inconsistency-tolerant Reasoning over Linear Probabilistic Knowledge Bases

Nico Potyka[a], Matthias Thimm[b]

[a]University of Osnabrück, Germany
[b]University of Koblenz-Landau, Germany

Abstract

We consider the problem of reasoning under uncertainty in the presence of inconsistencies. Our knowledge bases consist of linear probabilistic constraints that, in particular, generalize many probabilistic-logical knowledge representation formalisms. We first generalize classical probabilistic models to inconsistent knowledge bases by considering a notion of minimal violation of knowledge bases. Subsequently, we use these generalized models to extend two classical probabilistic reasoning problems (the probabilistic entailment problem and the model selection problem) to inconsistent knowledge bases. We show that our approach satisfies several desirable properties and discuss some of its computational properties.

Keywords: probabilistic reasoning, inconsistency-tolerant reasoning, probabilistic logic

---

1. Introduction

One of the main challenges in Knowledge Representation and Reasoning research is the handling of *uncertain* and *inconsistent* information. Unreliable sensor data, distorted communication channels, and other noisy data sources demand special treatment of information in order to produce reliable and robust results. Approaches for representing and reasoning with uncertainty can be roughly divided into formalisms for *qualitative uncertainty* and formalisms for *quantitative uncertainty*. The former comprise the large class of non-monotonic logics (Gabbay et al., 1994), i. e., logics that do not generally satisfy the property of *monotonicity* of classical logics (conclusions are preserved under the addition of new information). In order to allow for uncertain reasoning, these formalisms often include non-monotonic rules that do not necessarily hold in all cases. The archetype of such a rule is the *default rule* from default logic (Reiter, 1980). Some further examples of formalisms following this line are, e. g., answer set programming (Gelfond and Leone, 2002), conditional logics (Nute and Cross, 2002), defeasible logics (Nute, 1994), and computational models of argumentation (Baroni et al., 2011). In contrast, formalisms for *quantitative uncertainty* allow a finer-grained representation of uncertainty. For example, in probabilistic logics (Nilsson, 1986; Paris, 1994; Lukasiewicz and Kern-Isberner, 1999; Hansen and Jaumard, 2000), classical formulas in a knowledge base can be annotated with probabilities or intervals of probabilities and inferences can be drawn with probabilistic degrees of certainty.

In many situations, our knowledge is not only uncertain, but also inconsistent. The notion of *inconsistency* refers (usually) to multiple pieces of information and represents a conflict among those, i. e., they cannot hold at the same time. For instance, the two statements "John is on vacation in California" and "John is at home in New York" represent inconsistent information and in order to draw meaningful conclusions from a knowledge base containing these statements, this conflict has to be resolved or dealt with. In applications such as decision-support systems, knowledge bases are usually created by adding the formalized knowledge of different experts and different data sources and conflicts are hardly avoidable.

Several approaches have been developed for dealing with inconsistencies. Some of the qualitative formalisms mentioned before, such as computational models of argumentation, can also be applied for this purpose. Other examples include paraconsistent logics (Priest, 1991; Béziau et al., 2007; Arieli et al., 2011) which are formalisms based on classical logic that allow reasoning with inconsistent information. Approaches to inconsistency measurement (Grant and Hunter, 2006) can be used to analyze the severity of inconsistencies and to provide help in consolidating them. The fields of belief revision (Hansson, 2001) and belief merging (Cholvy and Hunter, 1997; Konieczny and Pérez, 1998) deal with the particular case of inconsistencies in dynamic settings. Usually, when new observations are made in a dynamic environment these observations can contradict with previously held beliefs and old beliefs have to be forgotten. There are also non-classical approaches to belief revision and belief merging such as answer set programming, cf. e. g. (Slota and Leite, 2012) or various probabilistic approaches (Kern-Isberner and Rödder, 2003; Adamcik, 2014; Wilmers, 2015).

The notions of uncertainty and inconsistency are orthogonal to each other. For example, a piece of information may be uncertain such as "Tomorrow it will rain with probability 0.9". If this is the only belief an agent possesses, it is consistent (although vague). Multiple pieces of information may be inconsistent such as "The bird Tweety flies" and "the bird Tweety does not fly". While these two statements are contradictory, each one represents a certain statement of its own. Of course, multiple pieces of information can also be both uncertain and inconsistent such as "Tomorrow it will rain with probability 0.9" and "tomorrow it will rain with probability 0.6". These beliefs are individually uncertain and taken together also inconsistent.

In this paper, our knowledge may be both *uncertain* and *inconsistent*. We represent uncertain knowledge by linear probabilistic constraints that in particular cover many probabilistic logical formalisms as considered in e. g. (Nilsson, 1986; Lukasiewicz, 1999; Fisseler, 2008; Kern-Isberner and Thimm, 2010). Inconsistencies occur in this framework when multiple constraints cannot be satisfied jointly by a probability distribution. To deal with inconsistencies, we consider probability distributions that in some sense minimally violate the knowledge rather than satisfy it. Based on this idea, we generalize two fundamental reasoning problems from probabilistic logics to inconsistent knowledge bases. These generalizations have the same solutions as the classical problem if the knowledge base is consistent and satisfy several reasonable properties in the general case. In particular, they feature two particularly interesting robustness properties. First, they are robust with respect to inconsistent knowledge that is independent of the query if we measure violation appropriately (we will make this precise later). Second, they behave continuously in the sense that if a knowledge base is close to a consistent knowledge base, then the query results will be close to the results under the consistent knowledge base. In this sense, minor inconsistencies in a knowledge base cannot yield major changes in the derived probabilities. We call these properties *Independence* and *Consistent Blaschke Continuity* and will make them precise in Sections 4 and 5. As we explain in Section 6, solving the generalized problems is barely harder than solving the classical problems in the sense that they belong to the same class of optimization problems and their size does only grow by a constant factor.

In this paper, we generalize, unify and extend the approach and results of our previous works (Potyka, 2014; Potyka and Thimm, 2014, 2015) that were in particular restricted to propositional probabilistic logics with point probabilities. The main contributions of this work are

1. We generalize the notion of models of a linear probabilistic knowledge base in the presence of inconsistencies (Section 3).
2. We generalize the classical probabilistic entailment problem in our framework and investigate its properties (Section 4).
3. We generalize the classical model selection problem in our framework and investigate its properties (Section 5).

Necessary preliminaries are presented in Section 2. We discuss related work in Section 7, and conclude with a summary in Section 8. AppendixA summarizes the notation used in the paper.

2. Preliminaries

2.1. Linear Probabilistic Knowledge Bases

We consider knowledge bases that represent probabilistic-logical relationships by means of a finite set of random variables $\mathcal{X} = \{X_1, \ldots, X_n\}$. For instance, in propositional probabilistic logics, $\mathcal{X}$ corresponds to the propositional variables of the language. In relational probabilistic logics, $\mathcal{X}$ corresponds to the ground atoms of the language. We will give some more detailed examples as we move on. Each $X \in \mathcal{X}$ has a finite domain $\mathrm{dom}(X) = \{x_1, \ldots, x_k\}$ of values it can take. An *atom over* $\mathcal{X}$ is an expression of the form $X = x$, where $X \in \mathcal{X}$ and $x \in \mathrm{dom}(X)$. A *formula over* $\mathcal{X}$ is either an atom or a *composite formula* that is built up by connecting atoms with the logical junctors $\neg, \wedge, \vee$ in the usual way. We denote the set of all formulas over $\mathcal{X}$ by $\mathcal{L}(\mathcal{X})$. If $\mathrm{dom}(X) = \{0, 1\}$, we call X a boolean variable and abbreviate $X = 1$ by X and $X = 0$ by $\neg X$. We will use this notation in particular if X is a relational atom. A *possible world* ω over $\mathcal{X}$ is a variable assignment, that is, a function that maps each variable to an element from its domain. In propositional probabilistic logics, ω corresponds to the possible truth assignments to the variables in the language. In relational logics, ω corresponds to the possible Herbrand interpretations (truth assignments to the ground atoms of the language). We will assume that the variables in $\mathcal{X}$ are ordered in some way so that we can write variable assignments more concisely. If ω assigns $x_i \in \mathrm{dom}(X_i)$ to X_i for $1 \leq i \leq n$, we write $\omega = (x_1, \ldots, x_n)$. We denote the set of all possible worlds by Ω. We say that ω satisfies the atom $X = x$ iff $\omega(X) = x$. Satisfaction of composite formulas is defined in the usual recursive way. If ϕ is a formula and ω satisfies ϕ, we write $\omega \models \phi$, and call ω a *model* of ϕ. The set of all models of ϕ is denoted by $\mathsf{Mod}(\phi)$.

In order to talk about the probability of formulas, we consider probability distributions $P : \Omega \to [0, 1]$ over the finite set of possible worlds (recall that a finite probability distribution is a non-negative function that satisfies $\sum_{\omega \in \Omega} P(\omega) = 1$). We overload the symbol $P(\cdot)$ to also denote the induced probability measure over $\mathcal{L}(\mathcal{X})$. More formally, for a formula ϕ, we let $P(\phi) = \sum_{\omega \in \mathsf{Mod}(\phi)} P(\omega)$. We will consider knowledge bases consisting of linear probabilistic constraints.

Definition 1 (Linear Probabilistic Constraint, Satisfaction)**.** A *linear probabilistic constraint* c over $\mathcal{X}$ is an expression of the form

$$c : h_0 + h_1\pi(\phi_1) + \ldots + h_m\pi(\phi_m) \leq 0, \tag{1}$$

where $\phi_1, \ldots, \phi_m$ are formulas over $\mathcal{X}$, $h_0, \ldots, h_m \in \mathbb{Q}$ are rational numbers, and π is a meta-logical symbol.

A probability distribution P (over the possible worlds over $\mathcal{X}$) satisfies c, written as $P \models c$ iff

$$h_0 + h_1P(\phi_1) + \ldots + h_mP(\phi_m) \leq 0. \tag{2}$$

Linear constraints of the above form cover in particular classical probabilistic logic (Nilsson, 1986) and several relational probabilistic logics as considered in (Lukasiewicz, 1999; Fisseler, 2008; Kern-Isberner and Thimm, 2010) for instance.

Example 1 (Basic Constraints)**.** The simple probabilistic constraint "The probability of ϕ is at most p" for some formula ϕ and $x \in [0, 1]$ can be represented as a linear constraint via

$$c_1 : \ -p + \pi(\phi) \leq 0$$

Similarly, the constraint "The probability of ϕ is at least p" can be represented as

$$c_2 : p - \pi(\phi) \leq 0$$

and "The probability of ϕ is exactly p" by taking both c_1 and c_2 into account. Furthermore, the constraint "The conditional probability of ϕ given ψ is undefined (that is, ψ has probability 0) or is at least p" can be represented as

$$c_3 : p\pi(\psi) - \pi(\psi \wedge \phi) \leq 0. \tag{3}$$

Note that this equation implies $P(\phi \mid \psi) \geq p$ for any P with $P \models c_3$ and $P(\psi) > 0$. We cannot express the condition $P(\psi) > 0$ in our framework. However, whether or not we enforce $P(\psi) > 0$, often does not make any significant difference for the reasoning problems that we are interested in, see (Potyka, 2016) for a thorough discussion. For notational convenience we abbreviate c_3 with $\pi(\phi \mid \psi) \geq p$ and write analogously $\pi(\phi \mid \psi) \leq p$ and $\pi(\phi \mid \psi) = p$. □

In the following, it will often be convenient to consider indicator functions for formulas ϕ over $\mathcal{X}$. The indicator function $1_\phi : \Omega \to \{0,1\}$ corresponding to ϕ yields $1_\phi(\omega) = 1$ if and only if $\omega \models \phi$ and 0 otherwise. Using indicator functions, we can rewrite the satisfaction condition of constraints as follows.

Lemma 1. *$P \models c$ iff*

$$\sum_{\omega \in \Omega} a_\omega P(\omega) \leq 0,$$

where $a_\omega = h_0 + \sum_{j=1}^{m} 1_{\phi_j}(\omega) h_j$.

Proof. $P \models c$ iff

$$\begin{aligned}
0 &\geq h_0 + h_1 P(\phi_1) + \ldots + h_m P(\phi_m) \\
&= h_0 \sum_{\omega \in \Omega} P(\omega) + h_1 \sum_{\omega \in \mathsf{Mod}(\phi_1)} P(\omega) + \ldots + h_m \sum_{\omega \in \mathsf{Mod}(\phi_m)} P(\omega) \\
&= h_0 \sum_{\omega \in \Omega} P(\omega) + h_1 \sum_{\omega \in \Omega} 1_{\phi_1}(\omega) P(\omega) + \ldots + h_m \sum_{\omega \in \Omega} 1_{\phi_m}(\omega) P(\omega) \\
&= \sum_{\omega \in \Omega} P(\omega)(h_0 + \sum_{j=1}^{m} 1_{\phi_j}(\omega) h_j).
\end{aligned}$$

□

Let us fix an arbitrary order of the possible worlds in Ω. Then we can identify a probability distribution P with the $|\Omega|$-dimensional column vector whose i-th component contains the probability of the i-th world $P(\omega_i)$. For a linear probabilistic constraint c, let r_c denote the row vector whose i-th component contains a_{ω_i} as in Lemma 1. Then Lemma 1 says that $P \models c$ iff $r_c P \leq 0$.

Definition 2 (Linear Probabilistic Knowledge Base, Satisfaction, Constraint Matrix)**.** A *linear probabilistic knowledge base* (or *knowledge base* for short) $\mathcal{K} = \{c_1, \ldots, c_m\}$ is a finite set of linear probabilistic constraints. A probability distribution P satisfies $\mathcal{K}$, written as $P \models \mathcal{K}$, iff P satisfies all $c \in \mathcal{K}$. The *constraint matrix corresponding to $\mathcal{K}$* is the $m \times n$ matrix

$$A_{\mathcal{K}} = \begin{pmatrix} r_{c_1} \\ \ldots \\ r_{c_m} \end{pmatrix}$$

As before, we abbreviate the set of all models of $\mathcal{K}$ as $\mathsf{Mod}(\mathcal{K}) = \{P \mid P \models \mathcal{K}\}$. A knowledge base $\mathcal{K}$ is *consistent* if there is a probability distribution P with $P \models \mathcal{K}$, otherwise it is *inconsistent*.

The following corollary follows immediately from Lemma 1.

Corollary 1. *$P \models \mathcal{K}$ iff $A_{\mathcal{K}} P \leq 0$.*

We conclude this part with an example that illustrates the expressivity of our formalism.

Example 2 (Constraints from Relational Probabilistic Logics)**.** Many relational probabilistic logics are based on the idea of regarding relational formulas as templates for propositional formulas, see, e.g., (Lukasiewicz, 1999; Fisseler, 2008; Loh et al., 2010). The general idea is to consider a finite set of constants *Consts* and a finite set of predicate symbols *Preds*. Knowledge bases consist of relational probabilistic conditionals like $(\phi \mid \psi)[l, u]$, where ϕ and ψ are relational formulas and l, u are probabilities such that $l \leq u$. The intuitive statement is that the probability of ϕ given that ψ holds is between l and u. A probability distribution P over the Herbrand interpretations (or equivalently, a joint distribution over the ground atoms regarded as random variables) of *Consts* and *Preds* satisfies the conditional $(\phi \mid \psi)[l, u]$ if for all possible ground instances $(\phi' \mid \psi')[l, u]$ of $(\phi \mid \psi)[l, u]$ over *Consts*, we have that either $P(\psi') = 0$ or $P(\phi' \mid \psi') \in [l, u]$ (note that this translates to two linear constraints of the form (3) for each ground conditional). If our relational conditional is $(Flies(X) \mid Bird(X))[0.8, 1]$ for instance and *Consts* $= \{a, b\}$, then we have to check the condition for the ground conditionals $(Flies(a) \mid Bird(a))[0.8, 1]$ and $(Flies(b) \mid Bird(b))[0.8, 1]$. This basic semantics can be refined by taking constraints over the possible ground instances and different grounding operators into account (Fisseler, 2008; Loh et al., 2010). We will use this semantics in some examples later and refer to it as the *grounding semantics*. Note that for grounding semantics, each conditional in the logical knowledge base yields several linear probabilistic constraints (as explained before, the number of constraints corresponding to the conditional is twice the number of its ground instances).

Grounding semantics are quite restrictive in the sense that they demand that a probability interval must hold for all ground instances. However, sometimes it is more natural to demand only that some algebraic combination of the probabilities of all ground instances lie in this interval. Two semantics that follow this idea have been proposed in (Kern-Isberner and Thimm, 2010). While a discussion of these semantics is out of the scope of this paper, we note that the *aggregating* *semantics* from (Kern-Isberner and Thimm, 2010) also yields linear probabilistic constraints, see the proof of (Potyka, 2015b), Proposition 3.9 for a detailed derivation. Interestingly, each conditional under aggregating semantics yields only one linear probabilistic constraint. Even though these constraints contain stronger correlations between ground atoms than the corresponding constraints under grounding semantics, computing reasoning results under aggregating semantics is often significantly faster than computing the corresponding results under grounding semantics. Actually, the aggregating semantics can be used to approximate reasoning results under grounding semantics, see (Potyka, 2016) for a detailed discussion. □

2.2. Reasoning over Linear Probabilistic Knowledge Bases: Entailment and Model Selection

Given a consistent knowledge base $\mathcal{K}$, we are interested in deriving meaningful probabilities for conditional queries. A *(conditional) query* is an expression of the form

$$(\phi \mid \psi),$$

where $\phi, \psi \in \mathcal{L}(\mathcal{X})$ are formulas over $\mathcal{X}$. There are two major approaches to derive meaningful probabilities for a query $(\phi \mid \psi)$ from the models of a consistent knowledge base $\mathcal{K}$. The first approach is to derive upper and lower bounds on the conditional probability of ϕ given ψ with respect to all models of $\mathcal{K}$ (Nilsson, 1986; Hansen and Jaumard, 2000). This approach is often referred to as the *probabilistic entailment problem*. The second approach is a two-stage process. One selects a best model that satisfies the knowledge base and then uses this model to compute the conditional probability of ϕ given ψ (Paris, 1994; Kern-Isberner, 2001). We refer to the first subproblem of the second approach as the *probabilistic model selection problem* and to the second as the *probabilistic inference problem*. We will be mainly concerned with the first stage here, i. e., with the probabilistic model selection problem. The probabilistic inference problem can always be solved naively (even though not very efficiently) by iterating over the probabilities of all worlds to compute the conditional probability. We now give a formal definition of the two reasoning problems that we are interested in.

Definition 3 (Probabilistic Entailment Problem, Entailment relation $\models^{pe}$)**.** Given a consistent knowledge base $\mathcal{K}$ over $\mathcal{X}$ and a query $(\phi \mid \psi), \phi, \psi \in \mathcal{L}(\mathcal{X})$, the *probabilistic entailment problem* is to compute lower and upper bounds on the probability of ϕ given ψ among the models of $\mathcal{K}$, that is, to solve the optimization problems

$$\begin{array}{ll} \min\limits_{P \in \mathsf{Mod}(\mathcal{K})} / \max\limits_{P \in \mathsf{Mod}(\mathcal{K})} & \dfrac{P(\phi \wedge \psi)}{P(\psi)}, \\ \textit{subject to} & P(\psi) > 0. \end{array}$$

If the probabilistic entailment problem can be solved, and the minimum and maximum are l and u, we write $\mathcal{K} \models^{pe} (\phi \mid \psi)[l, u]$.

Definition 4 (Probabilistic Model Selection Problem)**.** Given a consistent knowledge base $\mathcal{K}$ over $\mathcal{X}$ and a cost function $\mathcal{C}$ mapping probability distributions to $\mathbb{R}$, the *probabilistic model selection problem* is to compute a model of minimal cost, that is, to solve the optimization problem

$$\arg \min_{P \in \mathsf{Mod}(\mathcal{K})} \mathcal{C}(P).$$

The probabilistic entailment problem can be solved by fractional programming techniques (Charnes and Cooper, 1962; Hailperin, 1986) if there is a $P \in \mathsf{Mod}(\mathcal{K})$ that satisfies $P(\psi) > 0$. Since the model sets of linear probabilistic knowledge bases are guaranteed to be compact and convex (due to linearity), the probabilistic model selection problem can be solved by convex programming techniques whenever the cost function is strictly convex (concave) and continuous.

Example 3. We illustrate the probabilistic entailment problem by means of an example from (Potyka et al., 2015). Suppose that we want to design an intelligent agent that has to watch some pets. Sometimes the pets attack each other and our agent has to separate them. We consider a probabilistic conditional knowledge base $\mathcal{K}$ under grounding semantics as explained in Example 2. We consider constants $(\{bully, sylvester, tweety\}$ and the predicate symbols $\{Bird, Cat, Dog, Attacks, LT\}$ where $Bird, Cat, Dog$ are unary predicates denoting whether a term is a bird, a cat, or a dog, respectively, $Attacks$ and LT are binary predicates denoting whether one animal attacks another and whether one animal is larger than another animal, respectively. $\mathcal{K}$ contains the following conditionals (all probabilities are point intervals, so we use the notation $[p]$ rather than $[l, u]$ for probability intervals):

$$\begin{aligned}\mathcal{K} = \{&(\mathit{Bird}(X) \wedge \mathit{Cat}(X))[0], (\mathit{Bird}(X) \wedge \mathit{Dog}(X))[0], (\mathit{Cat}(X) \wedge \mathit{Dog}(X))[0],\\ &(\mathit{Attacks}(X, X))[0], (\mathit{LT}(X, Y) \wedge \mathit{LT}(Y, X))[0],\\ &(\mathit{Bird}(\mathit{tweety}))[1], (\mathit{Cat}(\mathit{sylvester})))[1], (\mathit{Dog}(\mathit{bully})))[1]\},\\ &(\mathit{Attacks}(X, Y) \mid \mathit{LT}(Y, X))[0.1],\\ &(\mathit{LT}(X, Y) \mid \mathit{Cat}(X) \wedge \mathit{Bird}(Y))[0.9],\\ &(\mathit{Attacks}(X, Y) \mid \mathit{Cat}(X) \wedge \mathit{Bird}(Y))[0.8]\}\end{aligned}$$

The conditionals express that a bird cannot be a cat, a bird cannot be a dog and a cat cannot be a dog. Furthermore, pets do not attack themselves and larger-than is an asymmetric relation. We also know that tweety is a bird, $\mathit{sylvester}$ is a cat, and bully is a dog. We assume that it is unlikely that a pet attacks a larger animal. Furthermore, we suppose that cats are usually larger than birds and that cats like attacking birds. We have the following entailment results under grounding semantics.

$$\begin{aligned}&\mathcal{K} \models^{pe} (\mathit{Attacks}(\mathit{tweety}, \mathit{sylvester}))[0.09, 0.19],\\ &\mathcal{K} \models^{pe} (\mathit{Attacks}(\mathit{tweety}, \mathit{tweety}))[0, 0],\\ &\mathcal{K} \models^{pe} (\mathit{Attacks}(\mathit{tweety}, \mathit{bully}))[0, 1],\\ &\mathcal{K} \models^{pe} (\mathit{Attacks}(\mathit{sylvester}, \mathit{tweety}))[0.8, 0.8].\end{aligned}$$

3. Generalized Models

We know from Corollary 1 that $P \models \mathcal{K}$ iff $A_{\mathcal{K}} P \leq 0$. Hence, if $\mathcal{K}$ is inconsistent, the system of inequalities $A_{\mathcal{K}} P \leq 0$ has no solution. Instead, we can search for solutions that violate the inequalities in a minimal way (Potyka, 2014; Bona and Finger, 2015). This can be done by replacing the constraints $A_{\mathcal{K}} P \leq 0$ with $A_{\mathcal{K}} P \leq \epsilon$ for some non-negative vector ϵ and minimizing the size of ϵ with respect to some vector norm.

Definition 5 (Vector norm)**.** A *vector norm* $\|.\|$ is a function $\|.\| : \mathbb{R}^n \to \mathbb{R}$ which satisfies

1. $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$.
2. $\|cx\| = |c|\|x\|$ for all $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$.
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

A vector norm is called *continuous* when it is a continuous function in the usual sense. An important class of vector norms is the class of *p-norms*. A p-norm is a vector norm $\|.\|_p : \mathbb{R}^n \to \mathbb{R}$ such that $\|x\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$, where $p \geq 1$. We assume that p is a natural number here in order to avoid additional computational issues. The most interesting special cases include the Manhattan norm $\|x\|_1 = \sum_{i=1}^n |x_i|$, the Euclidean norm $\|x\|_2 = \sqrt[p]{\sum_{i=1}^n x_i^2}$, and the maximum norm (the limit for $p \to \infty$) $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$. We will be in particular interested in norms that can be derived from p-norms as explained in the following definition.

Definition 6 (p-norm-related norm)**.** Let $p \geq 1$ and let $w = (w_i)$ be a sequence of positive real weights $w_i > 0$. For all $x \in \mathbb{R}^n$ the *w-weighted p-norm* is defined by

$$\|x\|_{p,w} = \sqrt[p]{\sum_{i=1}^n w_i |x_i|^p}$$

The *raised w-weighted p-norm* is defined by

$$\|x\|_{p,w}^p = \sum_{i=1}^n w_i |x_i|^p$$

The *w-weighted ∞-norm* is defined by

$$\|x\|_{\infty,w} = \max\{|w_1 \cdot x_1|, \cdots, |w_n \cdot x_n|\}$$

A *p-norm-related norm* $\|x\|_p^w$ is a norm that is either a w-weighted p-norm, a raised w-weighted p-norm or, in the case $p = \infty$, a w-weighted ∞-norm.

Remark 1. Of course, in practice, one will usually define only a finite set of weights for a given knowledge base. Note that this can formally be regarded as the special case, where all subsequent entries in the weight series are 0.

Sometimes there are constraints in our knowledge base that have a special status and should not be violated at all. Therefore, we introduce a second knowledge base that we call a set of integrity constraints. The set of integrity constraints is supposed to be consistent. Then, instead of minimizing ϵ with respect to all probability distributions, we consider only those that satisfy our integrity constraints (Potyka and Thimm, 2015).

Definition 7 (Minimal Violation Value with respect to $\mathcal{IC}$ and $\|.\|$)**.** Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases with corresponding constraint matrices $A_{\mathcal{K}}$ (of size $m \times n$), $A_{\mathcal{IC}}$ and let $\mathcal{IC}$ be consistent. Let $\|.\|$ be some continuous vector norm. The *minimal violation value of* $\mathcal{K}$ *with respect to* $\|.\|$ and integrity constraints $\mathcal{IC}$ is defined by

$$\begin{aligned} \min_{(x,\epsilon)\in\mathbb{R}^{n+m}} \quad & \|\epsilon\| \\ \textit{subject to} \quad & A_{\mathcal{K}}\, x \leq \epsilon, \\ & A_{\mathcal{IC}}\, x \leq 0, \\ & \sum_{i=1}^{n} x_i = 1, \\ & x \geq 0, \\ & \epsilon \geq 0. \end{aligned} \tag{4}$$

We denote the minimal violation value by $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K})$.

If $\mathcal{IC}$ and $\|.\|$ are clear from context or not important for the discussion, we will just omit the corresponding sub- and superscript to improve readability.

Proposition 1. *Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases and let $\mathcal{IC}$ be consistent. Let $\|.\|$ be some continuous vector norm. Then the minimal violation value $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K})$ is well-defined and non-negative. In particular, $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K}) = 0$ if and only if $\mathcal{K} \cup \mathcal{IC}$ is consistent.*

Proof. Let us first note that the feasible region of (4) is always non-empty. To see this, consider an arbitrary $P' \in \mathsf{Mod}(\mathcal{IC})$ (note that P' exists by consistency of $\mathcal{IC}$) and let $\epsilon' \in \mathbb{R}^m$ such that $\epsilon_i = |(A_{\mathcal{K}} P')_i|$ (ϵ_i contains the magnitude of the i-th component of the matrix-vector-product $A_{\mathcal{K}} P'$). Then (P', ϵ') is a feasible solution. In particular, the feasible region is defined by linear equality and inequality constraints and therefore convex and closed.

By non-negativity of vector norms, the objective function value of feasible solutions is bounded from below by 0. Therefore, continuity of $\|.\|$ implies the existence of a feasible solution that takes the minimum. Hence, $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K})$ is well-defined.

Non-negativity of $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K})$ follows from non-negativity of vector norms.

Definiteness of vector norms implies that $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K}) = 0$ iff $\epsilon = 0$. Since $\epsilon = 0$ iff $A_{\mathcal{K}}\, x \leq 0$ for some probability vector x that satisfies $\mathcal{IC}$, Corollary 1 implies that $\mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K}) = 0$ if and only if $\mathcal{K} \cup \mathcal{IC}$ is consistent. $\square$

Those probability distributions that minimally violate the knowledge base, will be called the generalized models of the knowledge base (Potyka and Thimm, 2014).

Definition 8 (Generalized models with respect to $\|.\|$)**.** Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases such that $\mathcal{IC}$ is consistent. Let $\|.\|$ be some continuous vector norm. The set of *generalized models of $\mathcal{K}$ with respect to the integrity constraints $\mathcal{IC}$ and* $\|.\|$ is defined by

$$\mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K}) = \{P \in \mathsf{Mod}(\mathcal{IC}) \mid \exists \epsilon \in \mathbb{R}^m : (P, \epsilon) \textit{ is an optimal solution of } (4)\}.$$

Again, if $\mathcal{IC}$ and $\|.\|$ are clear from context or not important for the discussion, we will omit the corresponding sub- and superscript to improve readability. The following proposition states some basic properties of generalized models. First, generalized models always exist and the set of generalized models satisfies some technical properties that will be useful for some proofs. Second, the generalized models respect the integrity constraints, that is, each generalized model is also a classical model of the integrity constraints. Third, if our knowledge base is consistent with the integrity constraints, the generalized models will coincide with the classical models. In particular, if $\mathcal{K}$ is consistent and $\mathcal{IC} = \emptyset$, the generalized models will coincide with the classical models.

Proposition 2.

1. $\mathsf{GMod}(\mathcal{K})$ *is always non-empty, convex and compact.*
2. $\mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K}) \subseteq \mathsf{Mod}(\mathcal{IC})$.
3. *If $\mathcal{K} \cup \mathcal{IC}$ is consistent, then* $\mathsf{GMod}_{\mathcal{IC}}(\mathcal{K}) = \mathsf{Mod}(\mathcal{K} \cup \mathcal{IC})$.

Proof. 1. We already showed in the proof of Proposition 1 that an optimal solution of (4) exists. Hence, the set of optimal solutions of (4) is non-empty. We can see from (4) that the set of optimal solutions correspond to the vectors $(x, \epsilon) \in \mathbb{R}^{n+m}$ that satisfy

$$\begin{aligned}
&\|\epsilon\| \leq \mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K}), \\
&A_{\mathcal{K}}\, x \leq \epsilon, \\
&A_{\mathcal{IC}}\, x \leq 0, \\
&\sum_{i=1}^{n} x_i = 1, \\
&x \geq 0, \\
&\epsilon \geq 0.
\end{aligned}$$

This is a system of convex continuous inequality constraints and one linear equality constraint. Therefore, the set of optimal solutions is convex and closed. Due to the first ($\|\epsilon\| \leq \mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$), fourth ($\sum_{i=1}^n x_i = 1$) and fifth ($x \geq 0$) constraint, the set of optimal solutions is also bounded and therefore compact. $\mathsf{GMod}(\mathcal{K})$ is the projection of the set of optimal solutions on the first n components ((x, ϵ) is mapped to x). Since the projection is linear and continuous, and the set of optimal solutions is non-empty, convex and compact, $\mathsf{GMod}(\mathcal{K})$ is non-empty, convex and compact.

2. If $P \in \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$, then there is an ϵ such that (P, ϵ) is an optimal solution of (4) and by the second constraint of (4), P satisfies $\mathcal{IC}$, i.e., $P \in \mathsf{Mod}(\mathcal{IC})$.

3. If $\mathcal{K} \cup \mathcal{IC}$ is consistent, we know from Proposition 1 that $\mathcal{I}_{\mathcal{IC}}(\mathcal{K}) = 0$. Each $P \in \mathsf{Mod}(\mathcal{K} \cup \mathcal{IC})$ satisfies $A_{\mathcal{K}}P \leq 0$ and therefore yields an optimal solution $(P, 0)$ of (4). Hence, $\mathsf{Mod}(\mathcal{K} \cup \mathcal{IC}) \subseteq \mathsf{GMod}_{\mathcal{IC}}(\mathcal{K})$. Conversely, if $P \in \mathsf{GMod}_{\mathcal{IC}}(\mathcal{K})$, then $A_{\mathcal{K}}P \leq 0$ (because of $\mathcal{I}_{\mathcal{IC}}(\mathcal{K}) = 0$ and definiteness of vector norms) and $A_{\mathcal{IC}}P \leq 0$ (because of the second constraint of (4)). Hence, also $\mathsf{GMod}_{\mathcal{IC}}(\mathcal{K}) \subseteq \mathsf{Mod}(\mathcal{K} \cup \mathcal{IC})$. $\square$

4. Generalized Entailment

4.1. Basic Definitions and Results

The framework developed so far allows us to generalize the classical probabilistic entailment problem (see Definition 3) to the inconsistent case as follows.

Definition 9 (Generalized Probabilistic Entailment Problem, Entailment relation $\models^{gpe}$)**.** Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ such that $\mathcal{IC}$ is consistent. Let $(\phi \mid \psi)$ be a query with $\phi, \psi \in \mathcal{L}(\mathcal{X})$, and let $\|.\|$ be some continuous vector norm. The *generalized probabilistic entailment problem* is to compute lower and upper bounds on the probability of ϕ given ψ among the generalized models of $\mathcal{K}$, that is, to solve the optimization problems

$$\begin{aligned} \min_{P \in \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})} \Big/ \max_{P \in \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})} \quad & \frac{P(\phi \wedge \psi)}{P(\psi)}, \\ \textit{subject to} \quad & P(\psi) > 0. \end{aligned} \tag{5}$$

If the generalized probabilistic entailment problem can be solved, and the minimum and maximum are l and u, we write $\mathcal{K} \models^{gpe}_{\mathcal{IC},\|\cdot\|} (\phi \mid \psi)[l, u]$.

If $\mathcal{IC}$ and $\|.\|$ are not important or clear from the context, we will omit the lower indices to enhance readability. The objective function and the strict inequality constraint $P(\psi) > 0$ look problematic from a computational perspective. However, we can apply similar ideas like in (Charnes and Cooper, 1962) to derive equivalent convex programs. For each formula $F \in \mathcal{L}(\mathcal{X})$, we let a_F denote the row vector whose i-th component is 1 if and only if the i-th world satisfies F. Then for each probability distribution P over $\mathcal{L}(\mathcal{X})$

$$a_F P = \sum_{\omega \in \Omega} 1_F(\omega) P(\omega) = P(F).$$

We can now derive the following equivalent convex programs.

Lemma 2. (5) *is equivalent to the convex programs*

$$\begin{aligned}
\min_{(x,\epsilon,s) \in \mathbb{R}^{n+m+1}} / \max_{(x,\epsilon,s) \in \mathbb{R}^{n+m+1}} \quad & a_{\phi \wedge \psi} x && (6) \\
\textit{subject to} \quad & A_{\mathcal{K}}\, x \leq s \cdot \epsilon, \\
& \|\epsilon\| \leq \mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K}), \\
& A_{\mathcal{IC}}\, x \leq 0, \\
& a_{\psi} x = 1, \\
& \sum_{i=1}^{n} x_i = s, \\
& x \geq 0, \\
& \epsilon \geq 0, \\
& s \geq 0.
\end{aligned}$$

Proof. (6) has a linear objective function and the constraints consist of linear equalities and inequalities and a convex inequality. Therefore, (6) is a convex program.

It remains to show that (6) is equivalent to (5). We will show that to each feasible solution of (6), there is a feasible solution of (5) that yields the same objective function value and vice versa.

Let P be a feasible solution of (5). Note that this in particular means that $P(\psi) > 0$. Let $s = \frac{1}{P(\psi)}$ and $x = sP$, that is, x is the vector from $\mathbb{R}^n$ whose i-th component contains the probability of the i-th world with respect to P scaled by s. Since P is a generalized model of $\mathcal{K}$, there is an $\epsilon \in \mathbb{R}^m$ such that (P, ϵ)

is an optimal solution of (4). Then (x, ϵ, s) clearly satisfies the sixth to eighth constraint. By feasibility of (P, ϵ) with respect to (4), we get for the first constraint that

$$A_{\mathcal{K}}\, x = sA_{\mathcal{K}}\, P \leq s\epsilon.$$

For the second constraint, we have

$$\|\epsilon\| = \mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K}).$$

For the third constraint, we get

$$A_{\mathcal{IC}}\, x = sA_{\mathcal{K}}\, P \leq s0 \leq 0.$$

For the fourth constraint, we have

$$a_\psi x = sa_\psi P = \frac{1}{P(\psi)} P(\psi) = 1.$$

Finally, we get for the fifth constraint that

$$\sum_{i=1}^{n} x_i = s \sum_{i=1}^{n} P(\omega_i) = s.$$

Hence, (x, ϵ, s) satisfies all constraints and is indeed a feasible solution of (6). For the objective function value, we get

$$a_{\phi\wedge\psi} x = \frac{1}{P(\psi)} a_{\phi\wedge\psi} P = \frac{P(\phi \wedge \psi)}{P(\psi)},$$

hence P and (x, ϵ) yield indeed the same objective function value.

Conversely, let (x, ϵ, s) be an optimal solution of (6). We let the corresponding probability distribution P be defined by the probability vector $\|x\|_1^{-1} x$, that is, P is obtained by normalizing x by dividing by the scalar $\|x\|_1 = \sum_{i=1}^{n} x_i = s$ (the last equality follows from the fifth constraint of (6)). In particular, $P(\psi) = \|x\|_1^{-1} a_\psi x = \|x\|_1^{-1} 1 > 0$ (the last equality follows from the constraint $a_\psi x = 1$ and feasibility of x) and we can check as before that P is a generalized model of $\mathcal{K}$ and hence a feasible solution of (5). For instance,

$$A_{\mathcal{K}}\, P = \|x\|_1^{-1} A_{\mathcal{K}}\, x \leq s^{-1} s\epsilon = \epsilon.$$

In particular, the normalizing scalar cancels out in the objective function, i.e.,

$$\frac{P(\phi \wedge \psi)}{P(\psi)} = \frac{a_{\phi \wedge \psi} P}{a_\psi P} = \frac{\|x\|_1^{-1} a_{\phi \wedge \psi} x}{\|x\|_1^{-1} a_\psi x} = \frac{a_{\phi \wedge \psi} x}{1},$$

where the last equality follows again from the constraint $a_\psi x = 1$ and feasibility of x. Hence, again P and (x, ϵ, s) yield the same objective function value. □

The following proposition explains in which cases the generalized entailment problem has a solution.

Proposition 3 (Solvability of Generalized Probabilistic Entailment Problem)**.** *Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ such that $\mathcal{IC}$ is consistent. Let $(\phi \mid \psi), \phi, \psi \in \mathcal{L}(\mathcal{X})$ be a query and let $\|.\|$ be some continuous vector norm. Minimum and maximum of* (5) *exist if and only if there is a $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ such that $P(\psi) >$* 0.

Proof. (5) has a feasible solution if and only if there is a $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ such that $P(\psi) > 0$. From the proof of Lemma 2 we know that this is the case if and only if (6) has a feasible solution and that both optimization problems yield the same results. Since (6) is a convex program, minimum and maximum do indeed exist in the case of feasibility. □

Since $\mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ is always non-empty, the generalized probabilistic entailment problem often yields answers to queries even if the knowledge base $\mathcal{K}$ is inconsistent. In case that $P(\psi) = 0$ for all $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$, we can return the empty interval $[1, 0]$ to indicate that the condition of the query is considered impossible with respect to $\mathcal{K}$ (Lukasiewicz, 2001).

Before investigating some common-sense properties of generalized probabilistic entailment, let us look at some reasoning examples to get some intuition for what happens under different norms. All examples will be modeled by a relational logic under grounding semantics as explained in Example 2.

Example 4 (Nixon Diamond)**.** Let us consider the *Nixon diamond*. We believe that quakers are usually pacifists while republicans are usually not. However, we know that Nixon was both a quaker and a republican. Our knowledge base contains the following probabilistic conditionals:

$$(\mathit{Pacifist}(X) \mid \mathit{Quaker}(X))[0.9],$$
$$(\mathit{Pacifist}(X) \mid \mathit{Republican}(X))[0.1],$$
$$(\mathit{Quaker}(\mathit{nixon}) \wedge \mathit{Republican}(\mathit{nixon}))[1].$$

Table 1 shows the generalized entailment results when using different p-norms. When using the 1-norm, the belief that Nixon is both a quaker and a republican

Query	1-norm	2-norm	∞-norm
$(\mathit{Pacifist}(nixon))$	$[0.1, 0.9]$	$[0.376, 0.624]$	$[0.376, 0.624]$
$(\mathit{Quaker}(nixon))$	$[1]$	$[0.744, 1]$	$[0.724, 1]$
$(\mathit{Republican}(nixon))$	$[1]$	$[0.744, 1]$	$[0.724, 1]$
$(\mathit{Pacifist}(X) \mid \mathit{Quaker}(X))$	$[0.1, 0.9]$	$[0.517, 0.615]$	$[0.519, 0.624]$
$(\mathit{Pacifist}(X) \mid \mathit{Republican}(X))$	$[0.1, 0.9]$	$[0.385, 0.483]$	$[0.376, 0.481]$

Table 1: Generalized entailment results (rounded to 3 digits) for probabilistic Nixon diamond.

is maintained. On the other hand, our beliefs about the relationship between pacifists, quakers and republicans are weakened substantially. In contrast, when using the 2- and ∞-norm, our beliefs in Nixon being a quaker and a republican are weakened, but it is still regarded more likely (> 0.5) that a quaker is a pacifist and a republican is not. The reason for this behavior is most likely that the 1-norm does not care about the extent of individual violations of constraints, but only about the overall violation of the knowledge base. In contrast, when using a p-norm for $p > 1$, higher violations of constraints are penalized more heavily due to the exponent p. Therefore, with increasing p, the violation of constraints is often more distributed among conflicting constraints. □

Example 5 (Combining Probabilistic Classifiers)**.** Suppose we want to incorporate a classification task in our reasoning system and trained several probabilistic classifiers using methods like logistic regression. Let us use the linear constraint $\pi(L = l) = \rho$ to express that l is the right label with probability ρ. In order to use these results for probabilistic reasoning, there are at least two kinds of inconsistencies we have to deal with. First, it is highly unlikely that all classifiers compute exactly the same probability for all labels and classical probabilistic reasoning approaches cannot deal with even minor deviations. For instance, if we get $\pi(L = l) = 0.9$ for the first and $\pi(L = l) = 0.89$ for the second classifier, the knowledge base would be inconsistent and classical probabilistic entailment was not applicable. Second, multiclass classification is often performed by training a binary classifier for each label. So if we have three labels, it might well be that we get $\pi(L = l_1) = 0.9$, $\pi(L = l_2) = 0.7$, $\pi(L = l_3) = 0.2$ for a single classifier. This information is contradictory because the probabilities of the labels have to sum to 1. Of course, we could just renormalize the values or introduce a boolean variable for each label, but we can also resolve both types of inconsistencies automatically by applying generalized probabilistic entailment. Let us consider the following knowledge base, where each row corresponds to the results of a single classifier (formally, we regard this knowledge base as a multiset):

$$\begin{array}{lll}
\pi(L = l_1) = 0.9 & \pi(L = l_2) = 0.7 & \pi(L = l_3) = 0.2 \\
\pi(L = l_1) = 0.8 & \pi(L = l_2) = 0.9 & \pi(L = l_3) = 0.5 \\
\pi(L = l_1) = 0.9 & \pi(L = l_2) = 0.9 & \pi(L = l_3) = 0.4
\end{array}$$

Table 2 shows the generalized entailment results when using different p-norms. The 1-norm yields again very conservative results, whereas we end up with point

Query	1-norm	2-norm	∞-norm
$(L = l_1)$	$[0.1, 0.8]$	$[0.511]$	$[0.467]$
$(L = l_2)$	$[0, 0.7]$	$[0.478]$	$[0.467]$
$(L = l_3)$	$[0, 0.2]$	$[0.111]$	$[0.067]$

Table 2: Generalized entailment results (rounded to 3 digits) for probabilistic classification example.

probabilities for the 2- and ∞-norm. Intuitively, generalized probabilistic entailment merged and normalized the probabilities in an automatic fashion. Note that we could further control the influence of different classifiers by setting their weights according to our preferences. □

Example 6 (Diagnosis)**.** Let us consider a medical toy example. Suppose we have some data about patients who most likely suffered from a cold and we want to use this data to revise our beliefs about our diagnosis and the symptoms of a cold. We assume that the symptoms have been recorded reliably and so we model this information by integrity constraints of the form (*has_symptom(patient, symptom)*)$[x]$, where $x \in \{0, 1\}$ dependent on whether the symptom was present or not. Table 3 shows the values of x for all patients and all symptoms. Our knowledge base contains a rule of the form (*has_disease(Patient,cold)*)$[1]$ for each patient from Table 3. By putting these rules into the knowledge base rather than into the set of integrity constraints, we incorporate our beliefs that all patients suffered from a cold without making this information irrefutable. We also add a rule (*induces_symptom(cold, Symptom)*)$[1]$ for each symptom from Table 3. Finally, we add the general rule

$$(\textit{has\_symptom(P,S)} \mid \textit{has\_disease(P,D)} \land \textit{induces\_symptom(D,S)})[1]$$

	alice	bob	charles	dora	eliza
fever	1	0	1	0	0
headache	1	1	1	1	0
cough	1	1	0	1	1

Table 3: Integrity constraints for diagnosis example.

that states that if a patient has a disease D and this disease induces symptom S, then the patient will have symptom S. Table 4 shows some query results for different norms. Recall that our knowledge base contains the assumptions that all

Query	1-norm	2-norm	∞-norm
(*has_disease(alice,cold)*)	$[1]$	$[1]$	$[0.667, 1]$
(*has_disease(bob,cold)*)	$[1]$	$[0.769]$	$[0.667]$
(*has_disease(charles,cold)*)	$[0, 1]$	$[0.667]$	$[0.667]$
(*has_disease(dora,cold)*)	$[1]$	$[0.769]$	$[0.667]$
(*has_disease(eliza,cold)*)	$[0, 1]$	$[0.615]$	$[0.667]$
(*induces_symptom(cold,fever)*)	$[0]$	$[0.461]$	$[0.667]$
(*induces_symptom(cold,headache)*)	$[0, 1]$	$[0.692]$	$[0.667]$
(*induces_symptom(cold,cough)*)	$[0, 1]$	$[0.667]$	$[0.667]$

Table 4: Generalized entailment results (rounded to 3 digits) for diagnosis example.

patients suffer from a cold, that a cold induces all symptoms and that a patient who suffers from a disease that induces a certain symptom has to feature this symptom. So we might expect that the probability that a patient suffers from a cold decreases if the patient does not feature all of the symptoms. This can be best observed from the 2-norm. In particular, the probability for Alice, who features all symptoms, remains 1, whereas the probability for Eliza, who features only one symptom, is the lowest. The probability for Charles is also decreased compared to the other to patients who feature 2 symptoms. The reason is most likely that he suffers from fever, which only 2 of 5 patients do, whereas he does not suffer from cough, which all other patients do. This is also reflected in the probabilities that a symptom is induced by a cold, where fever has a significantly lower probability for the 2-norm. Whereas the results for 1- and ∞-norm can also be explained by their special behaviour (not preferring small violations over large ones for the 1-norm and only caring about the maximum violation for the ∞-norm), we note that the 2-norm yields the most intuitive results. $\square$

4.2. Properties

We will now consider some common-sense properties of generalized probabilistic entailment. The first result states that generalized probabilistic entailment indeed generalizes probabilistic entailment in the sense that the results coincide for consistent knowledge bases.

Proposition 4 (Consistency)**.** *If $\mathcal{K} \cup \mathcal{IC}$ is consistent, the generalized probabilistic entailment problem coincides with the probabilistic entailment problem. That is,*

$$\mathcal{K} \models^{gpe}_{\mathcal{IC},\|.\|} (\phi \mid \psi)[l, u] \quad \textit{iff} \quad (\mathcal{K} \cup \mathcal{IC}) \models^{pe} (\phi \mid \psi)[l, u].$$

Proof. The claim follows immediately from item 3 of Proposition 2 because it says that the feasible regions of the generalized probabilistic entailment problem and the probabilistic entailment problem are equal if $\mathcal{K} \cup \mathcal{IC}$ is consistent. $\square$

Next, we consider some properties that are related to the intuition that reasoning results should be *independent* of knowledge that is not related to the query. In order to make this precise, we need to restrict and to extend possible worlds. If ω is a possible world wrt. $\mathcal{X}$ and $\mathcal{X}' \subseteq \mathcal{X}$ then $\omega|_{\mathcal{X}'}$ denotes the restriction of ω to $\mathcal{X}'$. Conversely, we can combine possible worlds over disjoint sets to possible worlds of the union of these sets. For instance, if we let $\omega_1 = \omega|_{\mathcal{X}'}$ and $\omega_2 = \omega|_{\mathcal{X} \setminus \mathcal{X}'}$ then ω_1 and ω_2 can be combined to ω and we write $\omega = (\omega_1, \omega_2)$. The following lemma states some simple rules for computing with independent probability distributions in our framework.

Lemma 3 (Product Distribution, Marginal Distribution)**.** *Let $\mathcal{X}_1, \mathcal{X}_2$ be disjoint sets of random variables, let $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ and let $\Omega_1, \Omega_2, \Omega$ denote the corresponding sets of possible worlds.*

1. *Let P_1, P_2 be probability distributions over $\mathcal{X}_1, \mathcal{X}_2$. Then $P : \Omega \to [0, 1]$ defined by $P(\omega) = P_1(\omega|_{\mathcal{X}_1}) P_2(\omega|_{\mathcal{X}_2})$ for all $\omega \in \Omega$ is a probability distribution over $\mathcal{X}$. In particular, for all $\phi_1 \in \mathcal{L}(\mathcal{X}_1)$, we have $P(\phi_1) = P_1(\phi_1)$.*
2. *Let P be a probability distribution over $\mathcal{X}$. Then $P_1 : \Omega_1 \to [0, 1]$ defined by $P_1(\omega_1) = \sum_{\omega_2 \in \Omega_2} P(\omega_1, \omega_2)$ for all $\omega_1 \in \Omega_1$ is a probability distribution over $\mathcal{X}_1$. In particular, for all $\phi_1 \in \mathcal{L}(\mathcal{X}_1)$, we have $P(\phi_1) = P_1(\phi_1)$.*

Proof. 1. For all $\phi_1 \in \mathcal{L}(\mathcal{X}_1)$, we have

$$P(\phi_1) = \sum_{(\omega_1,\omega_2)\in\mathsf{Mod}(\phi_1)} P_1(\omega_1)P_2(\omega_2) = \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2} 1_\phi(\omega_1)P_1(\omega_1)P_2(\omega_2)$$
$$= \underbrace{\sum_{\omega_1\in\Omega_1} 1_\phi(\omega_1)P_1(\omega_1)}_{=P_1(\phi_1)} \underbrace{\sum_{\omega_2\in\Omega_2} P_2(\omega_2)}_{=1}.$$

Hence, $P(\phi_1) = P_1(\phi_1)$. In particular, $\sum_{\omega\in\Omega} P(\omega) = P(\top) = P_1(\top) = 1$ and by non-negativity of P_1, P_2, P is non-negative. Hence, P is indeed a probability distribution over $\mathcal{X}$.

2. For all $\phi_1 \in \mathcal{L}(\mathcal{X}_1)$, we have

$$P_1(\phi_1) = \sum_{\omega_1\in\mathsf{Mod}(\phi_1)}\sum_{\omega_2\in\Omega_2} P(\omega_1,\omega_2) = \sum_{\omega_1\in\Omega_1} 1_\phi(\omega_1) \sum_{\omega_2\in\Omega_2} P(\omega_1,\omega_2)$$
$$= \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2} 1_\phi(\omega_1,\omega_2)P(\omega_1,\omega_2) = P(\phi_1).$$

We can check that P_1 is a probability distribution over $\mathcal{X}_1$ like in item 1. $\square$

We will denote the *product distribution* P from item 1 of Lemma 3 by $P_1 \odot P_2$ and the *marginal distribution* P_1 from item 2 by $P|_{\mathcal{X}_1}$. Note that by symmetry, we can consider the marginal distribution $P|_{\mathcal{X}_2}$ analogously.

Lemma 3 implies that generalized entailment is language-invariant, that is, adding new random variables to the language does not change inference results.

Corollary 2 (Language Invariance)**.** *Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ and let $\mathcal{IC}$ be consistent. Let $\|.\|$ be some continuous vector norm and let $(\phi \mid \psi), \phi, \psi \in \mathcal{L}(\mathcal{X})$ be a query over $\mathcal{L}(\mathcal{X})$. Let $\mathcal{X}'$ be a set of random variables with $\mathcal{X} \cap \mathcal{X}' = \emptyset$. If*

- $\mathcal{K} \models^{gpe}_{\mathcal{IC},\|.\|} (\phi \mid \psi)[l, u]$ *holds in $\mathcal{L}(\mathcal{X})$ and*
- $\mathcal{K} \models^{gpe}_{\mathcal{IC},\|.\|} (\phi \mid \psi)[l', u']$ *holds in $\mathcal{L}(\mathcal{X} \cup \mathcal{X}')$,*

then $l = l'$ and $u = u'$.

Proof. Consider a model P of $\mathcal{K}$ with respect to $\mathcal{L}(\mathcal{X})$. Let P' be the uniform distribution over $\mathcal{L}(\mathcal{X}')$. Then Lemma 3 implies that the product distribution $P^* = P \odot P'$ will agree with P on $\mathcal{L}(\mathcal{X})$. Therefore, P^* is a model of $\mathcal{K}$ with respect to $\mathcal{L}(\mathcal{X} \cup \mathcal{X}')$ that yields the same probability for the query. Therefore, $l' \leq l$ and $u \leq u'$.

Conversely, let P' be a model of $\mathcal{K}$ with respect to $\mathcal{L}(\mathcal{X} \cup \mathcal{X}')$. Then Lemma 3 implies that the marginal distribution $P'|_{\mathcal{X}}$ agrees with P' on $\mathcal{L}(\mathcal{X})$. Hence, $P'|_{\mathcal{X}}$ is a model of $\mathcal{K}$ with respect to $\mathcal{L}(\mathcal{X})$ that yields the same probability for the query. Therefore, also $l \leq l'$ and $u' \leq u$ and the claim follows. □

In a similar spirit, we might expect that if our knowledge base can be decomposed into independent parts and if our query depends only on one part, then the result should depend only on this one part. As it turns out, we have to further restrict norms to satisfy this desideratum.

Definition 10 ((Weakly) Dimension-Consistent Vector Norm)**.** A vector norm $\|.\|$ is called *dimension-consistent* iff for all $x_1, y_1 \in \mathbb{R}^{k_1}, x_2, y_2 \in \mathbb{R}^{k_2}$, we have that

1. $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \| < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|$ implies that $\|x_1\| < \|y_1\|$ or $\|x_2\| < \|y_2\|$.
2. $\|x_1\| < \|y_1\|$ and $\|x_2\| \leq \|y_2\|$ implies that $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \| < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|$.

If 2. only holds when both inequalities in the condition are strict, i.e.,

- $\|x_1\| < \|y_1\|$ and $\|x_2\| < \|y_2\|$ implies that $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \| < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|$,

we call $\|.\|$ *weakly dimension-consistent*.

The ∞-norm is an example of a norm that satisfies only the weaker version of item 2.

Example 7. We have $\| \begin{pmatrix} 0 \end{pmatrix} \|_\infty < \| \begin{pmatrix} 1 \end{pmatrix} \|_\infty$ and $\| \begin{pmatrix} 1 \end{pmatrix} \|_\infty \leq \| \begin{pmatrix} 1 \end{pmatrix} \|_\infty$ but $\| \begin{pmatrix} 0 \\ 1 \end{pmatrix} \|_\infty = 1 = \| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \|_\infty$.

The ∞-norm is weakly dimension-consistent and for $p < \infty$, p-norm-related norms are dimension-consistent as we show shortly. However, this needs not to be true when we allow adding weights in an arbitrary manner as the following example illustrates.

Example 8. Consider weights w with $w_1 = 1$, $w_2 = 3$, $w_3 = 1$. Let $x_1 = (11)$, $y_1 = (10)$, $x_2 = (10, 11)^T$, $y_2 = (11, 10)^T$. Then

$$\| \begin{pmatrix} 11 \\ 10 \\ 11 \end{pmatrix} \|_{1,w} = 52 < 53 = \| \begin{pmatrix} 10 \\ 11 \\ 10 \end{pmatrix} \|_{1,w}, \quad \textit{but}$$

$$\| \begin{pmatrix} 11 \end{pmatrix} \|_{1,w} = 11 > 10 = \| \begin{pmatrix} 10 \end{pmatrix} \|_{1,w} \textit{ and } \| \begin{pmatrix} 10 \\ 11 \end{pmatrix} \|_{1,w} = 43 > 41 = \| \begin{pmatrix} 11 \\ 10 \end{pmatrix} \|_{1,w}.$$

The problem in the example is, of course, that we weight the components differently as we decompose the vector. However, in practice, we associate weights with probabilistic constraints rather than with dimensions and change the weight vector appropriately when we consider subsets of the knowledge base. We could take account of this formally by considering a weighting function, rather than a weight vector, but in order to keep our notation simple, we will not do so. The results that we present for (weakly) dimension-consistent vector norms in the following are also true for weighted p-norm-related norms as long as the weights for subvectors are adapted appropriately. We explain this precisely in the following lemma.

Lemma 4. *Let $x_1, y_1 \in \mathbb{R}^{k_1}$, $x_2, y_2 \in \mathbb{R}^{k_2}$ and let $\|x\|_p^w$ be a w-weighted p-norm-related norm. Let $w^{(1)} = (w_1, \dots, w_{k_1}, 0, \dots)$, $w^{(2)} = (w_{k_1+1}, \dots, w_{k_1+k_2}, 0, \dots)$ ($w^{(1)}$ starts with the first k_1 elements of w and $w^{(2)}$ with the following k_2 elements, all following elements are 0). Then*

1. $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|_p^w < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|_p^w$ *implies that* $\|x_1\|_p^{w^{(1)}} < \|y_1\|_p^{w^{(1)}}$ *or* $\|x_2\|_p^{w^{(2)}} < \|y_2\|_p^{w^{(2)}}$.
2. $\|x_1\|_p^{w^{(1)}} < \|y_1\|_p^{w^{(1)}}$ *and* $\|x_2\|_p^{w^{(2)}} \leq \|y_2\|_p^{w^{(2)}}$ *implies that* $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|_p^w < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|_p^w$ *for* $p < \infty$.
3. $\|x_1\|_p^{w^{(1)}} < \|y_1\|_p^{w^{(1)}}$ *and* $\|x_2\|_p^{w^{(2)}} < \|y_2\|_p^{w^{(2)}}$ *implies that* $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|_p^w < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|_p^w$ *for* $p = \infty$.

Proof. 1. First, consider the case that $\|x\|_p^w$ is a raised w-weighted p-norm $\|x\|_{p,w}^p = \sum_{i=1}^n w_i |x_i|^p$. We prove the claim's contraposition. Suppose that both $\|x_1\|_{p,w^{(1)}}^p \geq$

$\|y_1\|^p_{p,w^{(1)}}$ and $\|x_2\|^p_{p,w^{(2)}} \geq \|y_2\|^p_{p,w^{(2)}}$. Then

$$\begin{aligned}\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|^p_{p,w} &= \sum_{i=1}^{k_1} w_i \cdot |x_{1,i}|^p + \sum_{i=1}^{k_2} w_{k_1+i} \cdot |x_{2,i}|^p = \|x_1\|^p_{p,w^{(1)}} + \|x_2\|^p_{p,w^{(2)}} \\ &\geq \|y_1\|^p_{p,w^{(1)}} + \|y_2\|^p_{p,w^{(2)}} = \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|^p_{p,w}.\end{aligned}$$

Next consider a w-weighted p-norm $\|x\|_{p,w}$. Note that $\|x\|_{p,w} = \sqrt[p]{\|x\|^p_{p,w}}$. Hence, by monotonicity of the root function, $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|_{p,w} < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|_{p,w}$ implies $\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|^p_{p,w} < \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|^p_{p,w}$. From our previous result, we can conclude $\|x_1\|^p_{p,w^{(1)}} < \|y_1\|^p_{p,w^{(1)}}$ or $\|x_2\|^p_{p,w^{(2)}} < \|y_2\|^p_{p,w^{(2)}}$. Taking the p-th root, we can conclude that $\|x_1\|_{p,w^{(1)}} < \|y_1\|_{p,w^{(1)}}$ or $\|x_2\|_{p,w^{(2)}} < \|y_2\|_{p,w^{(2)}}$ as desired.

Finally consider a w-weighted ∞-norm $\|x\|_{\infty,w} = \max\{|w_1 \cdot x_1|, \dots, |w_n \cdot x_n|\}$. We prove the claim's contraposition again. Suppose that both $\|x_1\|_{\infty,w^{(1)}} \geq \|y_1\|_{\infty,w^{(1)}}$ and $\|x_2\|_{\infty,w^{(2)}} \geq \|y_2\|_{\infty,w^{(2)}}$. Then

$$\begin{aligned}&\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|_{\infty,w} = \max\{|w_1 \cdot x_1|, \dots, |w_n \cdot x_n|\} \\ &= \max\{\max\{|w_1 \cdot x_1|, \dots, |w_{k_1} \cdot x_{k_1}|\}, \max\{|w_{k_1+1} \cdot x_{k_1+1}|, \dots, |w_{k_1+k_2} \cdot x_{k_1+k_2}|\}\} \\ &= \max\{\|x_1\|_{\infty,w^{(1)}}, \|x_2\|_{\infty,w^{(2)}}\} \\ &\geq \max\{\|y_1\|_{\infty,w^{(1)}}, \|y_2\|_{\infty,w^{(2)}}\} = \max\{|w_1 \cdot y_1|, \dots, |w_n \cdot y_n|\} \\ &= \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|_{\infty,w}.\end{aligned}$$

2. For the case that $\|x\|^w_p$ is a raised w-weighted p-norm $\|x\|^p_{p,w} = \sum_{i=1}^n w_i |x_i|^p$, $\|x_1\|^p_{p,w^{(1)}} < \|y_1\|^p_{p,w^{(1)}}$ and $\|x_2\|^p_{p,w^{(2)}} \leq \|y_2\|^p_{p,w^{(2)}}$ implies

$$\begin{aligned}\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|^p_{p,w} &= \sum_{i=1}^{k_1} w_i \cdot |x_{1,i}|^p + \sum_{i=1}^{k_2} w_{k_1+i} \cdot |x_{2,i}|^p = \|x_1\|^p_{p,w^{(1)}} + \|x_2\|^p_{p,w^{(2)}} \\ &< \|y_1\|^p_{p,w^{(1)}} + \|y_2\|^p_{p,w^{(2)}} = \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|^p_{p,w}.\end{aligned}$$

For weighted $p-$norms the claim follows analogously to 1.

3. Consider a w-weighted ∞-norm $\|x\|_{\infty,w} = \max\{|w_1 \cdot x_1|, \dots, |w_n \cdot x_n|\}$. $\|x_1\|_{\infty,w^{(1)}} < \|y_1\|_{\infty,w^{(1)}}$ and $\|x_2\|_{\infty,w^{(2)}} < \|y_2\|_{\infty,w^{(2)}}$ implies

$$\begin{aligned}
&\| \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \|_{\infty,w} = \max\{|w_1 \cdot x_1|, \dots, |w_n \cdot x_n|\} \\
&= \max\{\max\{|w_1 \cdot x_1|, \dots, |w_{k_1} \cdot x_{k_1}|\}, \max\{|w_{k_1+1} \cdot x_{k_1+1}|, \dots, |w_{k_1+k_2} \cdot x_{k_1+k_2}|\}\} \\
&= \max\{\|x_1\|_{\infty,w^{(1)}}, \|x_2\|_{\infty,w^{(2)}}\} \\
&< \max\{\|y_1\|_{\infty,w^{(1)}}, \|y_2\|_{\infty,w^{(2)}}\} = \max\{|w_1 \cdot y_1|, \dots, |w_n \cdot y_n|\} \\
&= \| \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \|_{\infty,w}.
\end{aligned}$$

□

Remark 2. Note that an unweighted p-norm-related norm corresponds to the special case of a (1)-weighted p-norm-related norm (that is, the weight of each dimension is 1). In this case, the lemma does indeed say that unweighted p-norm-related norms for $p < \infty$ are dimension-consistent and that the ∞-norm is weakly dimension-consistent in the sense of Definition 10.

In the following lemma, we state some matrix computation rules for independent knowledge bases.

Lemma 5 (Matrix Multiplication with Product Distribution and Marginal Distribution)**.** *Let $\mathcal{X}_1, \mathcal{X}_2$ be disjoint sets of random variables, let $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ and let $\Omega_1, \Omega_2, \Omega$ denote the corresponding sets of possible worlds. Let $\mathcal{K}_1$ be a knowledge base over $\mathcal{L}(\mathcal{X}_1)$ and let $\mathcal{K}_2$ be a knowledge base over $\mathcal{L}(\mathcal{X}_2)$.*

1. *Let P_1, P_2 be probability distributions over $\mathcal{X}_1, \mathcal{X}_2$ and let $P = P_1 \odot P_2$ denote the corresponding product distribution over $\mathcal{X}$. Then*

$$A_{(\mathcal{K}_1 \cup \mathcal{K}_2)} P = \begin{pmatrix} A_{\mathcal{K}_1} P_1 \\ A_{\mathcal{K}_2} P_2 \end{pmatrix}.$$

2. *Let P be a probability distribution over $\mathcal{X}$ and let $P_i = P|_{\mathcal{X}_i}$ for $i = 1, 2$. Then*

$$A_{(\mathcal{K}_1 \cup \mathcal{K}_2)} P = \begin{pmatrix} A_{\mathcal{K}_1} P_1 \\ A_{\mathcal{K}_2} P_2 \end{pmatrix}.$$

Proof. 1. Let $r_1, \dots, r_{k_1}$ denote the rows of $A_{(\mathcal{K}_1 \cup \mathcal{K}_2)}$ that correspond to the constraints in $\mathcal{K}_1$ and let $r_{k_1+1}, \dots r_{k_1+k_2}$ denote the remaining rows that correspond to the constraints in $\mathcal{K}_2$. Then for all $i \in 1, \dots, k_1 + k_2$, we have

$$r_i P = \sum_{\omega \in \Omega} a^i_\omega P(\omega) = \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} (h^i_0 + \sum_{j=1}^{m_i} 1_{\phi^i_j}(\omega_1, \omega_2) h^i_j) P_1(\omega_1) P_2(\omega_2)$$

as explained in Lemma 1. For $i \in \{1, \dots, k_1\}$, we have

$$\begin{aligned} r_i P &= \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} (h^i_0 + \sum_{j=1}^{m_i} 1_{\phi^i_j}(\omega_1, \omega_2) h^i_j) P_1(\omega_1) P_2(\omega_2) \\ &= \sum_{\omega_1 \in \Omega_1} (h^i_0 + \sum_{j=1}^{m_i} 1_{\phi^i_j}(\omega_1) h^i_j) P_1(\omega_1) \sum_{\omega_2 \in \Omega_2} P_2(\omega_2) \\ &= \sum_{\omega_1 \in \Omega_1} (h^i_0 + \sum_{j=1}^{m_i} 1_{\phi^i_j}(\omega_1) h^i_j) P_1(\omega_1), \end{aligned}$$

where we used the fact that $1_{\phi^i_j}(\omega_1, \omega_2) = 1_{\phi^i_j}(\omega_1)$ because the formulas in $\mathcal{K}_1$ depend only on ω_1. If we let s_i denote the i-th row of $A_{\mathcal{K}_1}$, we get again from Lemma 1 that

$$s_i P_1 = \sum_{\omega_1 \in \Omega_1} (h^i_0 + \sum_{j=1}^{m_i} 1_{\phi^i_j}(\omega_1) h^i_j) P_1(\omega_1) = r_i P.$$

In the same way, we can show that for $i \in \{k_1 + 1, \dots, k_1 + k_2\}$, we have $r_i P = s'_i P_2$, where s'_i denotes the $(i - k_1)$-th row of $A_{\mathcal{K}_2}$. Therefore,

$$A_{(\mathcal{K}_1 \cup \mathcal{K}_2)} P = \begin{pmatrix} r_1 P \\ \dots \\ r_{k_1} P \\ r_{k_1+1} P \\ \dots \\ r_{k_1+k_2} P \end{pmatrix} = \begin{pmatrix} s_1 P_1 \\ \dots \\ s_{k_1} P_1 \\ s'_1 P_2 \\ \dots \\ s'_{k_2} P_2 \end{pmatrix} = \begin{pmatrix} A_{\mathcal{K}_1} P_1 \\ A_{\mathcal{K}_2} P_2 \end{pmatrix}.$$

2. The proof is similar to the proof of item 1 and is therefore left out. □

We can now explain precisely how we can decompose independent knowledge bases and their generalized models.

Lemma 6 (Decomposability of Independent Knowledge Bases)**.** *Let $\mathcal{X}_1, \mathcal{X}_2$ be disjoint sets of random variables, let $\mathcal{K}_i$ denote a linear probabilistic knowledge base over $\mathcal{L}(\mathcal{X}_i)$ and let $\mathcal{IC}_i$ be a consistent knowledge base over $\mathcal{L}(\mathcal{X}_i)$ for $i = 1, 2$. Let $\|.\|$ be some continuous, weakly dimension-consistent vector norm.*

1. *If P_i is a probability distribution over $\mathcal{X}_i$ that is a generalized model of $\mathcal{K}_i$ for $i = 1, 2$, then $P = P_1 \odot P_2$ is a generalized model of $(\mathcal{K}_1 \cup \mathcal{K}_2)$.*

If $\|.\|$ is dimension-consistent (not only weakly), we also have

2. *If P is a probability distribution over $\mathcal{X}_1 \cup \mathcal{X}_2$ that is a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$, then $P_1 = P|_{\mathcal{X}_1}$ is a generalized model of $\mathcal{K}_1$.*

Proof. 1. According to Lemma 3, P is a probability distribution that coincides with P_i on $\mathcal{L}(\mathcal{X}_i)$. Therefore, P in particular satisfies both $\mathcal{IC}_1$ and $\mathcal{IC}_2$ since P_i satisfies $\mathcal{IC}_i$. It remains to show that P minimally violates $\mathcal{K}_1 \cup \mathcal{K}_2$. For the sake of contradiction, suppose this is not the case. Then consider some generalized model P' of $\mathcal{K}_1 \cup \mathcal{K}_2$. Since P is not a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$, we have that $\| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} P' \| < \| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} P \|$. Let $P'_i = P'|_{\mathcal{X}_i}$ for $i = 1, 2$. Then $\| \begin{pmatrix} A_1 P'_1 \\ A_2 P'_2 \end{pmatrix} \| < \| \begin{pmatrix} A_1 P_1 \\ A_2 P_2 \end{pmatrix} \|$ according to Lemma 5 and our assumption that P is not a generalized model. Then dimension-consistency implies that $\|A_1 P'_1\| < \|A_1 P_1\|$ or $\|A_2 P'_2\| < \|A_2 P_2\|$. Without loss of generality assume that $\|A_1 P'_1\| < \|A_1 P_1\|$. Then this contradicts the fact that P_1 is a generalized model of $\mathcal{K}_1$. Hence, P minimally violates $\mathcal{K}_1 \cup \mathcal{K}_2$ and is indeed a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$ that yields the same objective function value.

2. According to Lemma 3, P_1 is a probability distribution that coincides with P on $\mathcal{L}(\mathcal{X}_1)$. Therefore, P_1 in particular satisfies $\mathcal{IC}_1$ since P satisfies $\mathcal{IC}_1$. It remains to show that P_1 minimally violates $\mathcal{K}_1$. For the sake of contradiction, suppose this is not the case. Consider some generalized model P'_1 of $\mathcal{K}_1$. Since P_1 is not a generalized model of $\mathcal{K}_1$, we have that $\|A_1 P'_1\| < \|A_1 P_1\|$. Let $P_2 = P|_{\mathcal{X}_2}$ and let P'_2 be a generalized model of $\mathcal{K}_2$. Then $\|A_2 P'_2\| \leq \|A_2 P_2\|$ and $P' = P'_1 \odot P'_2$ is a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$ according to item 1. Dimension-consistency implies that $\|A_{\mathcal{K}_1 \cup \mathcal{K}_2} P'\| = \| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \begin{pmatrix} P'_1 \\ P'_2 \end{pmatrix} \| < \| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \| = \|A_{\mathcal{K}_1 \cup \mathcal{K}_2} P\|$, but this contradicts the assumption that P is a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$. □

For dimension-consistent vector norms, we can decompose the reasoning problem whenever the knowledge base decomposes into independent parts as explained in the following proposition.

Proposition 5 (Independence)**.** *Let $\mathcal{X}_1, \mathcal{X}_2$ be disjoint sets of random variables, let $\mathcal{K}_i$ denote a linear probabilistic knowledge base over $\mathcal{L}(\mathcal{X}_i)$ and let $\mathcal{IC}_i$ be a consistent knowledge base over $\mathcal{L}(\mathcal{X}_i)$ for $i = 1, 2$. Let $\|.\|$ be some continuous,* *dimension-consistent vector norm and let* $(\phi_1 \mid \psi_1), \phi_1, \psi_1 \in \mathcal{L}(\mathcal{X}_1)$ *be a query over* $\mathcal{L}(\mathcal{X}_1)$*. If*

- $\mathcal{K}_1 \models^{gpe}_{\mathcal{IC}_1, \|.\|} (\phi_1 \mid \psi_1)[l_1, u_1]$ *holds in* $\mathcal{L}(\mathcal{X}_1)$ *and*
- $(\mathcal{K}_1 \cup \mathcal{K}_2) \models^{gpe}_{(\mathcal{IC}_1 \cup \mathcal{IC}_2), \|.\|} (\phi_1 \mid \psi_1)[l, u]$ *holds in* $\mathcal{L}(\mathcal{X}_1 \cup \mathcal{X}_2)$*,*

then $l_1 = l$ *and* $u_1 = u$*.*

Proof. We prove the claim by showing that to each feasible solution of the generalized entailment problem with respect to $\mathcal{K}_1, \mathcal{IC}_1$ and $\mathcal{L}(\mathcal{X}_1)$ there corresponds a feasible solution of the generalized entailment problem with respect to $(\mathcal{K}_1 \cup \mathcal{K}_2)$, $(\mathcal{IC}_1 \cup \mathcal{IC}_2)$ and $\mathcal{L}(\mathcal{X}_1 \cup \mathcal{X}_2)$ with the same objective function value and vice versa.

Suppose that P_1 is a probability distribution over $\mathcal{X}_1$ that is a generalized model of $\mathcal{K}_1$. Let P_2 be some probability distribution over $\mathcal{X}_2$ that is a generalized model of $\mathcal{K}_2$. Then the distribution $P = P_1 \odot P_2$ over $\mathcal{X}_1 \cup \mathcal{X}_2$ is a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$ (Lemma 6) that yields the same objective function value like P_1 (Lemma 3).

Conversely, suppose that P is a generalized model of $\mathcal{K}_1 \cup \mathcal{K}_2$. Then we can show analogously as before that $P_1 = P|_{\mathcal{X}_1}$ is a generalized model of $\mathcal{K}_1$ that yields the same objective function value. □

Remark 3. For weakly dimension-consistent vector norms like the maximum norm, the proof still shows that $[l_1, u_1] \subseteq [l, u]$. However, $[l, u]$ can be a larger interval if the minimal violation value of $\mathcal{K}_2$ is larger than the one of $\mathcal{K}_1$. In this case, the constraints for $\mathcal{K}_1$ will be relaxed more strongly in $\mathcal{K}_1 \cup \mathcal{K}_2$ than when only considering $\mathcal{K}_1$ on its own.

Independence guarantees robustness of generalized entailment in the sense that classical probabilistic entailment results are maintained if the query is independent of the inconsistent information in the knowledge base.

Corollary 3 (Consistent Independence)**.** *Let* $\mathcal{X}_1, \mathcal{X}_2$ *be disjoint sets of random variables, let* $\mathcal{K}_i$ *denote a linear probabilistic knowledge base over* $\mathcal{L}(\mathcal{X}_i)$ *and let* $\mathcal{IC}_i$ *be a consistent knowledge base over* $\mathcal{L}(\mathcal{X}_i)$ *for* $i = 1, 2$*. Let* $\|.\|$ *be some continuous, dimension-consistent vector norm and let* $(\phi_1 \mid \psi_1), \phi_1, \psi_1 \in \mathcal{L}(\mathcal{X}_1)$ *be a query over* $\mathcal{L}(\mathcal{X}_1)$*. If* $\mathcal{K}_1$ *is consistent,*

- $(\mathcal{K}_1 \cup \mathcal{IC}) \models^{pe} (\phi_1 \mid \psi_1)[l_1, u_1]$ *holds in* $\mathcal{L}(\mathcal{X}_1)$ *and*

- $(\mathcal{K}_1 \cup \mathcal{K}_2) \models^{gpe}_{(\mathcal{IC}_1 \cup \mathcal{IC}_2),\|\cdot\|} (\phi_1 \mid \psi_1)[l,u]$ *holds in* $\mathcal{L}(\mathcal{X}_1 \cup \mathcal{X}_2)$,

then $l_1 = l$ *and* $u_1 = u$.

Proof. According to Proposition 5, $(\mathcal{K}_1 \cup \mathcal{K}_2) \models^{gpe}_{(\mathcal{IC}_1 \cup \mathcal{IC}_2),\|\cdot\|} (\phi_1 \mid \psi_1)[l,u]$ implies that $\mathcal{K}_1 \models^{gpe}_{\mathcal{IC}_1,\|\cdot\|} (\phi_1 \mid \psi_1)[l,u]$ holds in $\mathcal{L}(\mathcal{X}_1)$. Then consistency of $\mathcal{K}_1$ and Proposition 4 imply that $(\mathcal{K}_1 \cup \mathcal{IC}) \models^{pe} (\phi_1 \mid \psi_1)[l,u]$, i. e., $l_1 = l$ and $u_1 = u$. □

Consistent Independence relies on the assumption that the norm is dimension-consistent, not only weakly dimension-consistent. Indeed, the maximum norm does not satisfy this robustness property as we demonstrate in the following example.

Example 9. Consider the knowledge base $\mathcal{K}_1 = \{(a)[0.5]\}$. $\mathcal{K}_1$ is consistent and $\mathcal{K}_1 \models^{pe} (a)[0.5]$. The knowledge base $\mathcal{K}_2 = \{(b)[0.6], (b)[0.8]\}$ is inconsistent and independent of $\mathcal{K}_1$. So we would expect that the generalized entailment results do not change. However, we have $\big(\mathcal{K}_1 \cup \mathcal{K}_2\big) \models^{gpe}_{\emptyset,\|\cdot\|_\infty} (a)[0.4, 0.6]$. The reason for this behavior is that the maximum norm only looks at the maximal violation of a probabilistic constraint. Since adding $\mathcal{K}_2$ makes the violation value positive, the conditional $(a)[0.5]$ can be relaxed after adding $\mathcal{K}_2$ when using the maximum norm. Hence, the maximum norm does not satisfy Consistent Independence.

It would be desirable if generalized entailment behaved continuously in the sense that minor changes in the probabilities stated in the knowledge base could yield only minor changes in the derived probabilities. However, this is not even true for probabilistic entailment as the following example by Paris shows, see (Paris, 1994), Example 3.25[1].

Example 10. Consider a disease d, a symptom s and a possible complication c. Let $\mathcal{K}$ be defined via

$$\begin{array}{llll} \mathcal{K} = \{ & \pi(d \mid s) = 0.75, & \pi(d \mid \neg s) = 0.25, & \pi(\neg c \wedge d \mid s) = 0.15, \\ & \pi(\neg c \mid d \wedge \neg s) = 0.6, & \pi(c \mid d \wedge s) = 0.8, & \pi(c \wedge d \mid \neg s) = 0.1 \quad \} \end{array}$$

$\mathcal{K}$ is consistent and, for instance, $\mathcal{K} \models^{pe} (\neg s)[0,1]$. However, if we construct $\mathcal{K}'$ from $\mathcal{K}$ by replacing $\pi(c \wedge d \mid \neg s) = 0.1$ with $\pi(c \wedge d \mid \neg s) = 0.0999$, we have $\mathcal{K} \models^{pe} (\neg s)[0,0]$.

[1] The example was originally proposed in P. Courtney, Doctoral thesis, Manchester University, Manchester, U.K., 1992.

To exclude such discontinuities, Paris defined convergence of consistent knowledge bases by means of the Blaschke metric (Paris, 1994).

Definition 11 (Blaschke Metric)**.** Let $S_1, S_2 \subseteq \mathbb{R}^n$ be convex sets. S_1, S_2 have Blaschke distance d, $\|S_1, S_2\|_B = d$, iff d is the smallest value such that for every $x_1 \in S_1$, there is a $x_2 \in S_2$ such that $\|x_1 - x_2\|_2 \leq d$ and vice versa. More strictly speaking, we let d be the infimum of

$$\{\delta \mid \forall x_1 \in S_1 \exists x_2 \in S_2 : \|x_1 - x_2\|_2 \leq \delta \textit{ and } \forall x_2 \in S_2 \exists x_1 \in S_1 : \|x_1 - x_2\|_2 \leq \delta\}.$$

If there is no such δ, we let $d = \infty$.

Instead of measuring the distance between modified knowledge bases by the difference in the probabilities, Paris then measures the distance by comparing the induced sets of models. However, since inconsistent knowledge bases always induce the empty set of models, this topology becomes meaningless when considering inconsistent knowledge bases, see (Potyka, 2015b), Observation 6.30, for a detailed discussion. The natural extension to our framework is to replace the models with the generalized models.

Definition 12 (Generalized Blaschke Distance between Knowledge bases)**.** The *generalized Blaschke distance between knowledge bases* is defined as

$$\|\mathcal{K}_1, \mathcal{K}_2\|_B = \|\mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K}_1), \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K}_2)\|_B$$

for all knowledge bases $\mathcal{K}_1, \mathcal{K}_2$.

The generalized Blaschke distance is directly defined from the Blaschke metric and therefore inherits symmetry and triangle inequality. It is not definite, however. For instance, all equivalent consistent knowledge bases have the same set of (generalized) models and so their distance will be 0. The generalized Blaschke distance is therefore only a pseudometric.

Definition 13 (Convergence of Knowledge Bases)**.** Let $(\mathcal{K}_i)$ be a sequence of knowledge bases. We say $(\mathcal{K}_i)$ converges to $\mathcal{K}$ (with respect to $\mathcal{IC}$ and $\|.\|$), $\mathcal{K}_i \rightarrow^{\|\cdot\|}_{\mathcal{IC}} \mathcal{K}$, iff for each $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $i \geq N$ implies $\|\mathcal{K}_1, \mathcal{K}_2\|_B < \epsilon$.

Our definition generalizes Paris' topology over knowledge bases in the sense that whenever $(\mathcal{K}_i)$ converges to a consistent knowledge base $\mathcal{K}$ with respect to

Paris' notion of convergence, we also have $\mathcal{K}_i \to_{\emptyset}^{\|\cdot\|} \mathcal{K}$, see Potyka (2015b), Proposition 6.32.

Let us first note that if two knowledge bases are close with respect to the generalized Blaschke distance, then the probabilities entailed by these knowledge bases are also close.

Lemma 7. *For all $\epsilon > 0$, and for all linear probabilistic knowledge bases $\mathcal{K}_1, \mathcal{K}_2$ over a language $\mathcal{L}(\mathcal{X})$ with n possible worlds, $\|\mathcal{K}_1, \mathcal{K}_2\|_B < \frac{\epsilon}{\sqrt{n}}$ implies that for all $P_1 \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K}_1)$, there is a $P_2 \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K}_2)$ such that $|P_1(F) - P_2(F)| < \epsilon$ for all formulas $F \in \mathcal{L}(\mathcal{X})$.*

Proof. We know from Real Analysis that $\|x\|_1 \leq \sqrt{n}\|x\|_2$ for all $x \in \mathbb{R}^n$. Note also that for all probability distributions P, P' and formulas F, we have $|P(F) - P'(F)| \leq \sum_{\omega \in \mathsf{Mod}(F)} |P(\omega) - P'(\omega)| \leq \|P - P'\|_1$. Therefore, $\|\mathcal{K}_1, \mathcal{K}_2\|_B < \frac{\epsilon}{\sqrt{n}}$ implies that for all $P_1 \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K}_1)$, there is a $P_2 \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K}_2)$ such that $|P(F) - P'(F)| \leq \|P - P'\|_1 \leq \sqrt{n}\|P - P'\|_2 < \epsilon$. □

Note that by symmetry the lemma is also true if we switch the roles of P_1 and P_2. An immediate consequence is that Generalized Probabilistic Entailment is continuous for queries with tautological condition. We call these *unconditional queries*.

Corollary 4 (Blaschke Continuity for Unconditional Queries)**.** *Let $(\mathcal{K}_i)$ be a sequence of knowledge bases such that $\mathcal{K}_i \to_{\mathcal{IC}}^{\|\cdot\|} \mathcal{K}$ and let (ϕ) be an unconditional query. If*

- $\mathcal{K} \models_{\mathcal{IC},\|\cdot\|}^{gpe} (\phi)[l, u]$ *and*
- $\mathcal{K}_i \models_{\mathcal{IC},\|\cdot\|}^{gpe} (\phi)[l_i, u_i]$ *for $i \in \mathbb{N}$*

then $l_i \to l$ and $u_i \to u$ (in the usual sense).

Proof. Since $\mathcal{K}_i \to_{\mathcal{IC}}^{\|\cdot\|} \mathcal{K}$, for each $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $i \geq N$ implies $\|\mathcal{K}_1, \mathcal{K}_2\|_B < \frac{\epsilon}{\sqrt{n}}$. Therefore, Lemma 7 implies that for all $P_i \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K}_i)$ ($P \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})$), there is a $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})$ ($P_i \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K}_i)$) such that $|P_i(\phi) - P(\phi)| < \epsilon$. Therefore, the lower and upper bounds on the probability of ϕ with respect to $\mathcal{K}$ and $\mathcal{K}_i$ must converge as $i \to \infty$. That is, $l_i \to l$ and $u_i \to u$. □

For conditional probability queries, we impose a technical precondition that excludes some difficult boundary cases. It basically states that if $\mathcal{K}_i$ is sufficiently close to $\mathcal{K}$ then the probability of the condition of the query can be bounded away from 0 - at least for some probability distributions that take the extremal conditional probabilities.

Proposition 6 (Blaschke Continuity for Positive Conditional Queries)**.** *Let $(\mathcal{K}_i)$ be a sequence of knowledge bases such that $\mathcal{K}_i \rightarrow^{\|\cdot\|}_{\mathcal{IC}} \mathcal{K}$ and let $(\phi \mid \psi)$ be some query. If*

- $\mathcal{K} \models^{gpe}_{\mathcal{IC},\|\cdot\|} (\phi \mid \psi)[l, u]$ *and*
- $\mathcal{K}_i \models^{gpe}_{\mathcal{IC},\|\cdot\|} (\phi \mid \psi)[l_i, u_i]$ *for $i \in \mathbb{N}$*

and there is an $\epsilon_0 > 0$ and an $N_0 \in \mathbb{N}$ such that for all $i > N_0$, the lower and upper bounds on the conditional probability of ϕ given ψ for $\mathcal{K}_i$ are taken by some $P^l_i, P^u_i \in \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K}_i)$ with $P^l_i(\psi) \geq \epsilon_0$ and $P^u_i(\psi) \geq \epsilon_0$, then $l_i \rightarrow l$ and $u_i \rightarrow u$ (in the usual sense).

Proof. For ease of notation, let $G = \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$ and $G_i = \mathsf{GMod}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K}_i)$.

Let us show that $l_i \rightarrow l$. Let $P \in G$ be a minimal point, i.e., $\frac{P(\phi \wedge \psi)}{P(\psi)} = l$. Consider an arbitrarily small ϵ such that $0 < \epsilon < P(\psi)$. and let $\delta < \frac{\epsilon P(\psi)}{4\sqrt{n}}$. Then $\|G, G_i\|_B < \delta$ implies that for all $P \in G$ with $P(\psi) > 0$ there must be a $P' \in G_i$ with $P'(\psi) > 0$ such that

$$\begin{aligned}
|\frac{P(\phi \wedge \psi)}{P(\psi)} - \frac{P'(\phi \wedge \psi)}{P'(\psi)}| &= |\frac{P(\phi \wedge \psi)P'(\psi) - P'(\phi \wedge \psi)P(\psi)}{P(\psi)P'(\psi)}| \\
&\leq |\frac{P(\phi \wedge \psi)(P(\psi) + \delta) - (P(\phi \wedge \psi) - \delta)P(\psi)}{P(\psi)\frac{1}{2}P(\psi)}| \\
&= 2|\frac{P(\phi \wedge \psi)\delta + \delta P(\psi)}{P(\psi)^2}| \\
&\leq 2\delta|\frac{2P(\psi)}{P(\psi)^2}| = |\frac{4\delta}{P(\psi)}| < \epsilon.
\end{aligned}$$

In the second line, we used the fact that $\delta < \frac{\epsilon P(\psi)}{4} < \frac{P(\psi)}{2}$, which implies $P'(\psi) > P(\psi) - \delta > \frac{1}{2}P(\psi)$. Hence, if i is sufficiently large, $l_i \leq l + \epsilon$.

Let us now show that $l_i \geq l - \epsilon$ also holds if i is sufficiently large. Let us assume that $i > N_0$ and that i is so large that $\|G, G_i\|_B < \delta$ for $\delta < \frac{\epsilon\epsilon_0}{4\sqrt{n}}$. Then, in particular, $\delta < \frac{\epsilon P_i^l(\psi)}{4\sqrt{n}}$ for all such sufficiently large i and for each such P_i^l there must be a $P_i \in G$ such that $|\frac{P_i^l(\phi \wedge \psi)}{P_i^l(\psi)} - \frac{P_i(\phi \wedge \psi)}{P_i(\psi)}| < \epsilon$ (this can be seen as above). Hence, if i is sufficiently large, $l \leq l_i + \epsilon$.

Hence, $l - \epsilon \leq l_i \leq l + \epsilon$ for all $\epsilon > 0$ and sufficiently large i and so $l_i \to l$. The argumentation for $u_i \to u$ is analogous. □

Blaschke Continuity gives us a second robustness property: If our knowledge base is close to a consistent knowledge base, then the generalized entailment results will be close to the probabilistic entailment results with respect to the consistent knowledge base. We make this more precise in the following corollary.

Corollary 5 (Consistent Blaschke Continuity)**.** *Let $\mathcal{K}$, $\mathcal{IC}$ be knowledge bases such that $(\mathcal{K} \cup \mathcal{IC})$ is consistent. Let $(\mathcal{K}_i)$ be a sequence of knowledge bases such that $\mathcal{K}_i \to_{\mathcal{IC}}^{\|\cdot\|} \mathcal{K}$ and let $(\phi \mid \psi)$ be some query. If*

- $(\mathcal{K} \cup \mathcal{IC}) \models^{pe} (\phi \mid \psi)[l, u]$ *and*
- $\mathcal{K}_i \models^{gpe}_{\mathcal{IC},\|\cdot\|} (\phi \mid \psi)[l_i, u_i]$ *for* $i \in \mathbb{N}$

and the other conditions from Proposition 6 are true, then $l_i \to l$ and $u_i \to u$ (in the usual sense).

Proof. From Consistency (Proposition 4), we know that $\mathcal{K} \models^{gpe}_{\mathcal{IC},\|\cdot\|} (\phi \mid \psi)[l, u]$. From this, the claim follows with Proposition 6. □

5. Generalized Model Selection

Let us now consider the probabilistic model selection problem. Analogously to the probabilistic entailment problem, we can extend the probabilistic model selection problem to the inconsistent case by replacing the models with the generalized models.

Definition 14 (Generalized Probabilistic Model Selection Problem)**.** Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ such that $\mathcal{IC}$ is consistent. Given a *cost function* $\mathcal{C}$ mapping probability distributions to $\mathbb{R}$ and some continuous vector norm $\|.\|$, the *generalized probabilistic model selection problem* is to compute a generalized model of minimal cost, that is, to solve the optimization problem

$$\arg \min_{P \in \mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}} \mathcal{C}(P).$$

We will restrict our attention to cost functions that yield a unique solution. The following proposition gives a simple sufficient condition and applies in particular to maximum entropy reasoning (minimizing negative entropy), minimizing the least-squares-error to a given prior and minimizing relative entropy to a positive prior (priors with zero probabilities require a little more care for relative entropy minimization, see the discussion of absolute continuity and prior-consistency in (Kern-Isberner, 2001; Potyka, 2015b)).

Proposition 7 (Solvability of Generalized Probabilistic Model Selection Problem)**.** *Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ such that $\mathcal{IC}$ is consistent. Let $\mathcal{C}$ be a strictly convex continuous cost function and let $\|.\|$ be some continuous vector norm. Then the generalized probabilistic model selection problem has a unique solution.*

Proof. We know from Proposition 2 that $\mathsf{GMod}_{\mathcal{IC}}^{\|\cdot\|}$ is non-empty, convex and compact. Since minimizing a strictly convex continuous function over such a set guarantees the existence of a unique solution, the claim follows. □

Generalized probabilistic model selection satisfies common-sense properties similarly to generalized probabilistic entailment as we will show in the remainder of this section. If the best model with respect to $\mathcal{C}$ and $\|.\|$ is uniquely determined, we will denote it by $\mathcal{M}_{\mathcal{C}}^{\|\cdot\|}(\mathcal{K}, \mathcal{IC})$. Similarly, we will denote by $\mathcal{M}_{\mathcal{C}}(\mathcal{K})$ the best model with respect to the probabilistic model selection problem.

Proposition 8 (Consistency)**.** *Let $\mathcal{C}$ be a strictly convex continuous cost function and let $\|.\|$ be some vector norm. If $\mathcal{K} \cup \mathcal{IC}$ is consistent, the generalized model selection problem coincides with the probabilistic model selection problem. That is,*

$$\mathcal{M}_{\mathcal{C}}^{\|\cdot\|}(\mathcal{K}, \mathcal{IC}) = \mathcal{M}_{\mathcal{C}}(\mathcal{K} \cup \mathcal{IC}).$$

Proof. The claim follows from item 3 of Proposition 2 because it guarantees that the feasible regions of the generalized probabilistic entailment problem and the probabilistic entailment problem are equal if $\mathcal{K} \cup \mathcal{IC}$ is consistent. □

Remark 4. For general cost functions, we can state that the best solutions with respect to the generalized model selection problem coincide with the best solutions with respect to the probabilistic model selection problem whenever $\mathcal{K} \cup \mathcal{IC}$ is consistent (the same proof applies). Note also that we do not need any assumptions on $\|.\|$ because by definiteness of vector norms, each generalized model must be a probabilistic model and vice versa.

When using the negative entropy $-H(P) = \sum_{\omega \in \Omega} P(\omega) \cdot \log(P(\omega))$ as cost function, we get an independence result similar to generalized entailment.

Proposition 9 (ME-Independence)**.** *Let $\mathcal{C} = -H$ be the negative entropy and let $\|.\|$ be some continuous, dimension-consistent vector norm. Let $\mathcal{X}_1, \mathcal{X}_2$ be disjoint sets of random variables, let $\mathcal{K}_i$ denote a linear probabilistic knowledge base over $\mathcal{L}(\mathcal{X}_i)$ and let $\mathcal{IC}_i$ be a consistent knowledge base over $\mathcal{L}(\mathcal{X}_i)$ for $i = 1, 2$. Let P_i^* denote the best generalized model over $\mathcal{L}(\mathcal{X}_i)$ with respect to $\mathcal{K}_i$ and let P^* denote the best generalized model over $\mathcal{L}(\mathcal{X}_1 \cup \mathcal{X}_2)$ with respect to $\mathcal{K}_1 \cup \mathcal{K}_2$. Then $P^* = P_1^* \odot P_2^*$ and $P_i^* = P^*|_{\mathcal{X}_i}$.*

Proof. Let us first note that P^*, P_1^* and P_2^* are all well-defined by Proposition 7. In particular, we know from Lemma 6 that $P_1^* \odot P_2^*$ and $P^*|_{\mathcal{X}_i}$ are indeed generalized models. Assume that $P^* \neq P_1^* \odot P_2^*$. Then $-H(P^*) < -H(P_1^* \odot P_2^*)$. We have that

$$\begin{aligned}
-H(P_1^* \odot P_2^*) &= \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} P_1^*(\omega_1) \cdot P_2^*(\omega_2) \cdot \log(P_1^*(\omega_1) \cdot P_2^*(\omega_2)) \\
&= \big(\sum_{\omega_2 \in \Omega_2} P_2^*(\omega_2) \big) \cdot \big(\sum_{\omega_1 \in \Omega_1} P_1^*(\omega_1) \cdot \log P_1^*(\omega_1) \big) \\
&\quad + \big(\sum_{\omega_1 \in \Omega_1} P_1^*(\omega_1) \big) \cdot \big(\sum_{\omega_2 \in \Omega_2} P_2^*(\omega_2) \cdot \log P_2^*(\omega_2) \big) \\
&= -H(P_1^*) - H(P_2^*).
\end{aligned}$$

From the *Independence Bound for Entropy* (see Yeung (2008), Theorem 2.39), we have $-H(P^*) \geq -H(P^*|_{\mathcal{X}_1}) - H(P^*|_{\mathcal{X}_2})$. Therefore,

$$-H(P^*|_{\mathcal{X}_1}) - H(P^*|_{\mathcal{X}_2}) \leq -H(P^*) < -H(P_1^* \odot P_2^*) = -H(P_1^*) - H(P_2^*).$$

But then $-H(P^*|_{\mathcal{X}_1}) < -H(P_1^*)$ or $-H(P^*|_{\mathcal{X}_2}) < -H(P_2^*)$, which contradicts optimality of P_1^* and P_2^*. Hence, $P^* = P_1^* \odot P_2^*$ must be true. In particular, $P_i^* = (P_1^* \odot P_2^*)|_{\mathcal{X}_i} = P^*|_{\mathcal{X}_i}$. □

We get the following corollary that states that the generalized ME-model coincides with the classical ME-model on consistent, independent subsets of the knowledge base.

Corollary 6 (Consistent ME-Independence)**.** *Let $\mathcal{C} = -H$ be the negative entropy and let $\|.\|$ be some continuous, dimension-consistent vector norm. Let $\mathcal{X}_1, \mathcal{X}_2$ be* *disjoint sets of random variables, let $\mathcal{K}_i$ denote a linear probabilistic knowledge base over $\mathcal{L}(\mathcal{X}_i)$ and let $\mathcal{IC}_i$ be a consistent knowledge base over $\mathcal{L}(\mathcal{X}_i)$ for $i = 1, 2$. Assume further that $\mathcal{K}_1 \cup \mathcal{IC}_1$ is consistent. Let P_1^* denote the classical maximum entropy model over $\mathcal{L}(\mathcal{X}_i)$ with respect to $\mathcal{K}_1$ and let P^* denote the best generalized model over $\mathcal{L}(\mathcal{X}_1 \cup \mathcal{X}_2)$ with respect to $\mathcal{K}_1 \cup \mathcal{K}_2$. Then $P_1^* = P^*|_{\mathcal{X}_1}$.*

Proof. The claim follows from Proposition 8 and Proposition 9. □

Remark 5. Let us emphasize again that dimension-consistency is important for our Independence properties. When using the Maximum norm, the generalized model can be different from the ME model after adding independent inconsistent knowledge. The reason is again that the increased inconsistency value now allows violating the constraints in the consistent knowledge base similar to Example 9.

As another corollary, we get again a language invariance property.

Corollary 7 (ME Language Invariance)**.** *Let $\mathcal{C} = -H$ be the negative entropy and let $\|.\|$ be some continuous, dimension-consistent vector norm. Let $\mathcal{K}$ denote a linear probabilistic knowledge base over $\mathcal{L}(\mathcal{X})$. Let $\mathcal{X}'$ denote another set of random variables and let P^* and Q^* denote the generalized best models of $\mathcal{K}$ over $\mathcal{L}(\mathcal{X})$ and $\mathcal{L}(\mathcal{X} \cup \mathcal{X}')$. Then $P^*(\phi) = Q^*(\phi)$ for all formulas $\phi \in \mathcal{L}(\mathcal{X})$.*

Proof. We know from Proposition 9 that $P^* = Q^*|_{\mathcal{X}}$. From this, the claim follows with Lemma 3. □

Generalized model selection is again Blaschke continuous if we consider strictly convex continuous cost functions and continuous vector norms. The key ideas of the following theorem's proof are taken from Paris' proof of continuity of classical ME reasoning, see (Paris, 1994), Theorem 7.5.

Proposition 10 (Blaschke Continuity)**.** *Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ such that $\mathcal{IC}$ is consistent. Let $\mathcal{C}$ be a strictly convex continuous cost function and let $\|.\|$ be some continuous vector norm. Let $(\mathcal{K}_i)$ be a sequence of knowledge bases such that $\mathcal{K}_i \rightarrow_{\mathcal{IC}}^{\|.\|} \mathcal{K}$. Let (P_i^*) denote the corresponding sequence of best generalized models with respect to $(\mathcal{K}_i)$ and let P^* be the best generalized model with respect to $\mathcal{K}$. Then (P_i^*) converges component-wise to P^*.*

Proof. For ease of notation, we let $G = \mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ and $G_i = \mathsf{GMod}_{\mathcal{IC}}^{\|.\|}(\mathcal{K}_i)$. Our proof takes the key ideas from Theorem 7.5 in (Paris, 1994), which shows a similar continuity result for consistent knowledge bases. We show that for each $\epsilon > 0$, there is a $\delta > 0$ such that $\|G, G_i\|_b < \delta$ implies $\|P^* - P_i^*\|_2 < \epsilon$.

Consider the set

$$S = \{P \in G \mid \|P^* - P\|_2 \geq \frac{\epsilon}{2}\}$$

of probability functions in G having distance at least $\frac{\epsilon}{2}$ to P^*. By continuity of the euclidean distance and compactness of G, S is compact. Since $\mathcal{C}$ is continuous, the minimum

$$\nu = \min \{\mathcal{C}(P) - \mathcal{C}(P^*) \mid P \in S\}$$

does exist and is greater than zero because P^* is the unique best model with respect to $\mathcal{C}$.

Note that $\mathcal{C}$ is even uniformly continuous because the set of all probability distributions is compact. Therefore, we can find a $\delta > 0$ such that for all probability distributions P_1, P_2, $\|P_1 - P_2\|_2 < \delta$ implies that $|\mathcal{C}(P_1) - \mathcal{C}(P_2)| < \min \{\frac{\epsilon}{2}, \frac{\nu}{2}\}$. In particular, we can assume that $\delta < \frac{\epsilon}{2}$. Then, if $\|G, G_i\|_B < \delta$, there is a $P \in G$ with $\|P - P_i^*\|_2 < \delta$ and a $P_i \in G_i$ with $\|P_i - P^*\|_2 < \delta$. Then

$$\mathcal{C}(P^*) > \mathcal{C}(P_i) - \frac{\nu}{2} \geq \mathcal{C}(P_i^*) - \frac{\nu}{2},$$
$$\mathcal{C}(P_i^*) > \mathcal{C}(P) - \frac{\nu}{2} \geq \mathcal{C}(P^*) - \frac{\nu}{2}$$

and therefore, $|\mathcal{C}(P_i^*) - \mathcal{C}(P^*)| < \frac{\nu}{2}$. Hence, we can conclude that

$$|\mathcal{C}(P) - \mathcal{C}(P^*)| \leq |\mathcal{C}(P) - \mathcal{C}(P_i^*)| + |\mathcal{C}(P_i^*) - \mathcal{C}(P^*)| < \frac{\nu}{2} + \frac{\nu}{2} = \nu.$$

But by definition of ν this means that $P \in G \setminus S$ and therefore $\|P - P^*\|_2 < \frac{\epsilon}{2}$. Hence,

$$\|P^* - P_i^*\|_2 \leq \|P^* - P\|_2 + \|P - P_i^*\|_2 < \frac{\epsilon}{2} + \delta < \epsilon.$$

□

In particular, if the limit of the knowledge bases is consistent, the generalized best models will converge to the classical best model of the limit.

Corollary 8 (Consistent Blaschke Continuity)**.** *Let $\mathcal{K}, \mathcal{IC}$ be knowledge bases over $\mathcal{X}$ such that $(\mathcal{K} \cup \mathcal{IC})$ is consistent. Let $\mathcal{C}$ be a strictly convex continuous cost function and let $\|.\|$ be some continuous vector norm. Let $(\mathcal{K}_i)$ be a sequence of knowledge bases such that $\mathcal{K}_i \rightarrow^{\|.\|}_{\mathcal{IC}} \mathcal{K}$. Let (P_i^*) denote the corresponding sequence of best generalized models with respect to $(\mathcal{K}_i)$ and let P^* be the best (classical) model with respect to $\mathcal{K}$. Then (P_i^*) converges component-wise to P^*.*

Proof. The claim follows from Proposition 8 and Blaschke Continuity. □

6. Computational Issues

In this section, we will look at solving our generalized reasoning problems in somewhat more detail. We know from Lemma 2 and the proof of Proposition 7 that both the generalized probabilistic entailment problem and the generalized probabilistic model selection problem (with appropriately restricted cost function and norm) can be solved by convex optimization techniques. There are solvers for convex problems that run in time cubic in the number of optimization variables and are guaranteed to converge to the global optimum (actually each local optimum is globally optimal in a convex optimization problem) (Boyd and Vandenberghe, 2004). In this section, we discuss some interesting special cases that can be solved more efficiently.

6.1. Generalized Probabilistic Entailment

Let us start with the generalized probabilistic entailment problem. Solving this problem is a two-stage process. We first compute the minimal violation measure $\mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$ by solving optimization problem (4) and then solve the optimization problem (2). We will show that in several interesting cases, the problems correspond to linear programs. In practice, such problems can usually be solved in linear time in the number of optimization variables when using the Simplex algorithm (even though the worst-case runtime can be exponential) (Matousek and Gärtner, 2007). However, it is important to remember that the optimization variables correspond to possible worlds over random variables. Therefore, the number of optimization variables is exponential in the number of random variables.

We compute $\mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$ by solving (4), which is a convex optimization problem. However, the only non-linear term is the objective function. For the 1- and ∞-norm, we can linearize the objective function in order to obtain a linear program.

Proposition 11. *When using a (raised) w-weighted 1-norm or a w-weighted ∞-norm, $\mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$ can be computed by linear programming.*

Proof. In order to compute $\mathcal{I}^{\|\cdot\|}_{\mathcal{IC}}(\mathcal{K})$, we have to solve the optimization problem

$$\begin{aligned}
\min_{(x,\epsilon)\in\mathbb{R}^{n+m}} \quad & \|\epsilon\| \\
\textit{subject to} \quad & A_{\mathcal{K}}\, x \leq \epsilon, \\
& A_{\mathcal{IC}}\, x \leq 0, \\
& \sum_{i=1}^{n} x_i = 1, \\
& x \geq 0, \\
& \epsilon \geq 0.
\end{aligned} \tag{7}$$

If we use a (raised) w-weighted 1-norm, our objective function is $\sum_{i=1}^{m} |w_i \cdot \epsilon_i|$. However, since we have the constraint $\epsilon \geq 0$ and w contains only positive entries, $|w_i \cdot \epsilon_i| = w_i \cdot \epsilon_i$ for all feasible solutions. Hence, the objective function can be equivalently written as $\sum_{i=1}^{m} w_i \epsilon_i$ and we have a linear program.

If we use a w-weighted ∞-norm, the objective function is $\max\{|w_i \cdot \epsilon_i| \mid 1 \leq i \leq m\}$, which by the same argument as before is equivalent to $\max\{w_i \cdot \epsilon_i \mid 1 \leq i \leq m\}$. Let W be the $m \times m$-diagonal matrix whose diagonal entries correspond to the first m entries in w, that is $W_{i,i} = w_i$ for $i = 1, \dots, m$. Consider the linear program

$$\begin{aligned}
\min_{(x,y)\in\mathbb{R}^{n+1}} \quad & y \\
\textit{subject to} \quad & W A_{\mathcal{K}}\, x \leq y \cdot \vec{1}, \\
& A_{\mathcal{IC}}\, x \leq 0, \\
& \sum_{i=1}^{n} x_i = 1, \\
& x \geq 0, \\
& y \geq 0,
\end{aligned} \tag{8}$$

where $\vec{1}$ denotes the m-dimensional vector that contains only ones (each row in the result of $A_{\mathcal{K}}\, x$ has to be less than or equal to y). Note that we only changed the first and last constraint (and reduced the number of optimization variables). We show that this new linear program is equivalent to the original one. First suppose that (x, ϵ) is an optimal solution to the original problem. We let $y = \max\{w_i \cdot \epsilon_i \mid 1 \leq i \leq m\}$. Since, $\epsilon \geq 0$ and w is positive, we have $y \geq 0$. Furthermore,

$WA_{\mathcal{K}}\,x \leq W\epsilon \leq y \cdot \vec{1}$ by feasibility of (x, ϵ) and definition of y. Hence, (x, y) is a solution to the new problem with the same objective function value and the minimum of the new problem bounds the minimum of the original problem from below.

Conversely, suppose that (x, y) is an optimal solution of the new problem. Let $\epsilon' = A_{\mathcal{K}}\,x$. Let ϵ be defined by $\epsilon_i = \max\{\epsilon'_i, 0\}$ (we replace the negative components in ϵ' with 0). We have $\epsilon \geq 0$ and $A_{\mathcal{K}}\,x = \epsilon' \leq \epsilon$. Hence, (x, ϵ) is a feasible solution for the original problem. For the objective function, we get $\|\epsilon\|_{\infty,w} = \max\{|w_i \cdot \epsilon_i| \mid 1 \leq i \leq m\} = \max\{|(WA_{\mathcal{K}}\,x)_i| \mid 1 \leq i \leq m\} \leq y$. For the sake of contradiction, suppose that $\|\epsilon\|_{\infty,w} < y$. Then $WA_{\mathcal{K}}\,x < y \cdot \vec{1}$. However, then we could find a $y' \in \mathbb{R}$ such that $WA_{\mathcal{K}}\,x \leq y' \cdot \vec{1} < y \cdot \vec{1}$. In particular, $y' < y$, contradicting optimality of (x, y). Hence, we must have $\|\epsilon\|_{\infty,w} = y$ and the minimum of the original problem bounds the minimum of the new problem from below. Hence, both problems must have the same minimum. □

Remark 6. Note in particular that in the linear programming formulation, the number of optimization variables can only decrease (the number remains equal for the 1-norm and decreases by $m - 1$ for the ∞-norm).

Note that when we want to perform some form of generalized probabilistic reasoning, we have to compute the minimal violation value only once. Afterwards, we can reuse it to answer different queries (the minimal violation value depends only on the knowledge base and not on the query). After having computed the minimal violation value, the generalized probabilistic entailment problem corresponds to a linear program for all p-norm-related norms as we will show now.

First note that each optimal solution of (4) is a pair (P, ϵ) consisting of a probability distribution and a *violation vector*. Even though all these vectors have the same length (they all satisfy $\|\epsilon\| = \mathcal{I}^{\|.\|}_{\mathcal{IC}}(\mathcal{K})$), there can be different violation vectors in general. However, if we use particular vector norms, the violation vector is indeed unique. This is in particular true for p-norm-related norms as the following lemma shows.

Lemma 8 (Unique Violation Vector for p-norm-related norms with $1 < p < \infty$)**.** *Let $\|.\|^w_p$ be some p-norm-related norm with $1 < p < \infty$. Then for all optimal solutions $(P_1, \epsilon_1), (P_2, \epsilon_2)$ of* (4)*, we have that $\epsilon_1 = \epsilon_2$.*

Proof. As $\|.\|^w_p$ is a vector norm, we have $\|x+y\|_p \leq \|x\|_p + \|y\|_p$ for all $x, y \in \mathbb{R}^n$ due to item 3 of Definition 5. In particular, if $1 < p < \infty$ and $\|.\|^w_p = \|.\|_p$ is a

p-norm, equality holds if and only if there is a non-negative constant $\lambda \geq 0$ such that $x = \lambda y$, see (Steele, 2004), Problem 9.3.

Let us first consider the case that $\|.\|_p^w = \|.\|_p$ is a p-norm. We know from the proof of Proposition 2 that the set of optimal solutions of (4) is convex. Therefore, the convex combination $(0.5P_1 + 0.5P_2, 0.5\epsilon_1 + 0.5\epsilon_2)$ is also an optimal solution. Optimality implies that $\|0.5\epsilon_1 + 0.5\epsilon_2\|_p = \mathcal{I}_{\mathcal{IC}}^{\|.\|_p}(\mathcal{K}) = \|\epsilon_1\|_p = \|\epsilon_2\|_p$. Hence, $0.5\|\epsilon_1 + \epsilon_2\|_p = \|0.5\epsilon_1 + 0.5\epsilon_2\|_p = 0.5\|\epsilon_1\|_p + 0.5\|\epsilon_2\|_p$, i. e., $\|\epsilon_1 + \epsilon_2\|_p = \|\epsilon_1\|_p + \|\epsilon_2\|_p$ and equality holds for the Minkowski inequality. Since we assume $1 < p < \infty$, we can conclude $\epsilon_1 = \lambda\epsilon_2$. From $\|\epsilon_1\|_p = \|\epsilon_2\|_p$, it follows that $\|\epsilon_1\|_p = \|\lambda\epsilon_2\|_p = |\lambda|\|\epsilon_1\|_p$, i.e., $|\lambda| = 1$. Since $\lambda \geq 0$, we have $\lambda = 1$ and therefore $\epsilon_1 = \epsilon_2$.

Now let $\|.\|_p^w = \|.\|_{p,w}$ be a weighted p-norm. For all $n \in \mathbb{N}$, let $W_n^{\frac{1}{p}}$ be the diagonal matrix whose diagonal entries correspond to the p-th root of the first n entries in w, that is, $W_{i,i} = w_i^{\frac{1}{p}}$ for $i = 1, \ldots, n$. Then

$$\|x\|_{p,w} = \sqrt[p]{\sum_{i=1}^{n} w_i |x_i|^p} = \sqrt[p]{\sum_{i=1}^{n} |w_i^{\frac{1}{p}} x_i|^p} = \|W_n^{\frac{1}{p}} x\|_p \tag{9}$$

for all $x \in \mathbb{R}^n$. Hence $\|\epsilon_i\|_{p,w} = \|W_m^{\frac{1}{p}}\epsilon_i\|_{p,w}$ for $i = 1, 2$ and as before, we can conclude that $W_m^{\frac{1}{p}}\epsilon_1 = \lambda W_m^{\frac{1}{p}}\epsilon_2$. Since w contains only positive entries, $W_m^{\frac{1}{p}}$ is a diagonal matrix with positive diagonal entries and therefore invertible. Hence, we can conclude that $\epsilon_1 = \lambda\epsilon_2$ and from here like above that $\epsilon_1 = \epsilon_2$.

Finally, let $\|.\|_p^w = \|x\|_{p,w}^p$ be a raised w-weighted p-norm. Note that $\|x\|_{p,w}^p$ is just a weighted p-norm $\|.\|_{p,w}$ raised to the power of p. This is a monotone transformation, since all arguments are non-negative. Therefore, the optimal solutions remain unchanged an the claim follows from the previous result. □

The lemma gives us a simple characterization of $\mathsf{GMod}_{\mathcal{IC}}^{\|.\|_p}(\mathcal{K})$ in terms of a system of linear equations and inequalities as explained in the following corollary.

Corollary 9. *Let $\|.\|_p^w$ be some p-norm-related norm with $1 < p < \infty$, let $\mathcal{K}$ be some knowledge base and let ϵ be the corresponding minimal violation vector. Then $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|.\|_p}(\mathcal{K})$ if and only if $A_{\mathcal{K}}P \leq \epsilon$ and $A_{\mathcal{IC}}P \leq 0$.*

Proof. If $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|.\|_p}(\mathcal{K})$, there is an ϵ' such that (P, ϵ') is an optimal solution of (4). But then $\epsilon' = \epsilon$ by Lemma 8. Hence, $A_{\mathcal{K}}P \leq \epsilon$ and $A_{\mathcal{IC}}P \leq 0$ by feasibility of (P, ϵ'). If conversely, $A_{\mathcal{K}} P \leq \epsilon$ and $A_{\mathcal{IC}} P \leq 0$, then (P, ϵ) is a feasible solution of (4). In particular, (P, ϵ) is an optimal solution by definition of ϵ. Hence, $P \in \mathsf{GMod}_{\mathcal{IC}}^{\|.\|_p}(\mathcal{K})$. $\square$

We can now give the linear programming formulation of the generalized probabilistic entailment problem for p-norm-related norms.

Proposition 12 (Generalized Probabilistic Entailment Problem for p-norms)**.** *The generalized probabilistic entailment problem for p-norm-related norms can be solved by linear programming whenever the corresponding minimal violation vector is known.*

Proof. We know from Lemma 2 that the generalized probabilistic entailment problem can be solved by the following convex programs

$$
\begin{aligned}
\min_{(x,\epsilon,s) \in \mathbb{R}^{n+m+1}} / \max_{(x,\epsilon,s) \in \mathbb{R}^{n+m+1}} \quad & a_{\phi \wedge \psi} x && (10) \\
\textit{subject to} \quad & A_{\mathcal{K}}\, x \leq s \cdot \epsilon, \\
& \|\epsilon\|_p^w \leq \mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K}), \\
& A_{\mathcal{IC}}\, x \leq 0, \\
& a_\psi x = 1, \\
& \sum_{i=1}^n x_i = s, \\
& x \geq 0, \\
& \epsilon \geq 0, \\
& s \geq 0.
\end{aligned}
$$

Note that the only non-linear term is the second constraint $\|\epsilon\|_p^w \leq \mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$. However, $\|\epsilon\|_p^w \leq \mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ can equivalently be written as $\|\epsilon\|_p^w = \mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ because $\mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ is by definition the minimum value that ϵ can take.

In the case that $1 < p < \infty$, we know from Lemma 8 that the minimal violation vector ϵ is unique. As Corollary 9 states, this allows us to characterize the generalized models by the linear constraints $A_{\mathcal{K}} x \leq \epsilon$ and $A_{\mathcal{IC}}\, x \leq 0$. Therefore, we can replace the constraints $A_{\mathcal{K}}\, x \leq s \cdot \epsilon$ and $\|\epsilon\|_p^w \leq \mathcal{I}_{\mathcal{IC}}^{\|.\|}(\mathcal{K})$ in the original problem from Lemma 2 with the single constraint $A_{\mathcal{K}} x \leq s \cdot \epsilon$. This yields the following linear program:

$$
\begin{aligned}
\min_{(x,\epsilon,s)\in\mathbb{R}^{n+m+1}} \Big/ \max_{(x,\epsilon,s)\in\mathbb{R}^{n+m+1}} \quad & a_{\phi\wedge\psi}x \\
\textit{subject to} \quad & A_{\mathcal{K}}x \leq s\cdot\epsilon, \\
& A_{\mathcal{IC}}\,x \leq 0, \\
& a_{\psi}x = 1, \\
& \sum_{i=1}^{n} x_i = s, \\
& x \geq 0, \\
& \epsilon \geq 0, \\
& s \geq 0.
\end{aligned}
\tag{11}
$$

Like in the proof of Lemma 8, we can check that both optimization problems are indeed equivalent.

In the case $p = 1$, we have $\|\epsilon\|_1^w = \sum_{i=1}^m |w_i \cdot \epsilon_i|$. Since ϵ is restricted to be non-negative this can be equivalently written as $\sum_{i=1}^m w_i \cdot \epsilon_i$ and so the constraint $\|\epsilon\|_1^w = \mathcal{I}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})$ is linear, so that the whole optimization problem is linear.

Finally, consider the case $p = \infty$. Using the same argumentation that we used in the proof of Proposition 11, we can show that the constraints $A_{\mathcal{K}}\,x \leq s\cdot\epsilon$ and $\|\epsilon\|_\infty^w \leq \mathcal{I}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})$ can be replaced with the constraint $A_{\mathcal{K}}\,x \leq s\cdot\mathcal{I}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})\cdot\vec{1}$, where $\vec{1}$ again denotes the m-dimensional vector that contains only ones (note that $\mathcal{I}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})$ is the minimal y that linear program (8) in the proof of Proposition 11 can take). This yields the linear program

$$
\begin{aligned}
\min_{(x,\epsilon,s)\in\mathbb{R}^{n+m+1}} \Big/ \max_{(x,\epsilon,s)\in\mathbb{R}^{n+m+1}} \quad & a_{\phi\wedge\psi}x \\
\textit{subject to} \quad & A_{\mathcal{K}}\,x \leq s\cdot\mathcal{I}_{\mathcal{IC}}^{\|\cdot\|}(\mathcal{K})\cdot\vec{1}, \\
& A_{\mathcal{IC}}\,x \leq 0, \\
& a_{\psi}x = 1, \\
& \sum_{i=1}^{n} x_i = s, \\
& x \geq 0, \\
& \epsilon \geq 0, \\
& s \geq 0.
\end{aligned}
\tag{12}
$$

We can again check equivalence to the original optimization problem by using the same argumentation as in the proof of Lemma 8. □

Hence, in order to perform generalized probabilistic entailment with some p-norm-related norm, we have to solve a convex (linear for $p = 1, \infty$) program once in order to compute the minimal violation value. Afterwards, we have to solve a linear program for each query. Compare this with classical probabilistic entailment. When using probabilistic entailment, we typically perform a satisfiability test once (which corresponds to a linear program (Hansen and Jaumard, 2000; Potyka, 2015b)) and afterwards solve linear programs for each query ((Hansen and Jaumard, 2000; Lukasiewicz, 1999; Potyka, 2015b)). The number of optimization variables for the optimization problems in classical probabilistic entailment basically corresponds to the number of possible worlds. In generalized probabilistic entailment the number increases at most by the size of the knowledge base. This is often negligible since the knowledge base is typically small compared to the number of possible worlds. The number of constraints in probabilistic entailment basically corresponds to the size of the knowledge base. The same is true for the generalized probabilistic entailment problem when using p-norm-related norms as we saw in the proof of Proposition 12. Therefore, we can expect the same asymptotic runtime behavior for both classical and generalized probabilistic entailment when using p-norm-related norms.

6.2. Generalized Probabilistic Model Selection

In the last subsection, we saw that we can simplify the computational problem for the generalized probabilistic entailment problem when we restrict to certain norms. For the generalized probabilistic model selection problem, we have to restrict the cost function, too. The only special case that we looked at in more detail is optimum entropy model selection. Entropy maximization subject to linear equality constraints is often simplified by deriving a dual unconstrained problem with the same dimension. This can also be done for generalized optimum entropy model selection when knowledge bases contain only equality constraints. The technical details can be found in the proof appendix of (Potyka and Thimm, 2014) for entropy maximization and in (Potyka, 2015b) for the more general case of relative entropy minimization subject to a prior distribution. However, when considering inequality constraints as in this paper, we usually cannot eliminate all constraints in the corresponding dual problem. In (Kazama and Tsujii, 2005) a dual problem is presented that leaves one non-negativity constraint for each inequality constraint (non-negativity constraints in the primal can be eliminated, however).

We might apply similar ideas to simplify generalized probabilistic model selection under relative entropy to a prior in future work. Independent of all this, we note again that we can expect the same asymptotic runtime behavior for both classical and generalized model selection because the the size of the optimization problems does not change significantly.

7. Related Work

The minimal violation value from Definition 7 was originally introduced as an inconsistency measure in (Potyka, 2014). An inconsistency measure $\mathcal{I}$ is a function that maps a knowledge base to a non-negative real number such that larger values indicate larger inconsistency (Grant and Hunter, 2013). De Bona and Finger extended minimal violation measures to interval probabilities and also noted some relationships between minimal violation measures and Dutch books, see (Bona and Finger, 2015) for more details. For probabilistic logics, several other inconsistency measures have been proposed, see, e. g. (Thimm, 2013; Picado-Muiño, 2011).

In (Potyka and Thimm, 2014) we used generalized models to repair inconsistent knowledge bases using the principle of maximum entropy. In (Potyka and Thimm, 2015) we generalized the propositional probabilistic entailment problem to the inconsistent case. In this paper, we unified and complemented results from (Potyka, 2014), (Potyka and Thimm, 2014) and (Potyka and Thimm, 2015), generalized them to linear probabilistic knowledge bases (that include relational logics as proposed in (Lukasiewicz, 1999; Fisseler, 2008; Kern-Isberner and Thimm, 2010) and probabilistic logics that allow interval probabilities (Bona et al., 2014; Potyka, 2016)) and more flexible norms than p-norms. The latter is in particular important for recent applications that we considered in reasoning with priorities (Potyka, 2015a) and group decision making (Potyka et al., 2016), where one wants to assign different weights to different beliefs. In particular, we corrected a too strong continuity claim made in (Potyka and Thimm, 2014) for generalized model selection under maximum entropy.

Several other methods have been proposed to deal with inconsistent information in probabilistic logic. Daniel considered some ideas that are closely related to generalized reasoning (Daniel, 2009). To deal with inconsistent knowledge, he first defined the notion of a *candidacy function*, which assigns a real number between 0 and 1 to each probability distribution. Intuitively, the candidacy function assigns 1 to all models of the knowledge base. The value decreases with the distance of the probability distribution to the hyperplanes that correspond to linear constraints in the knowledge base. Note that this approach is different from minimal violation values, where we minimize the numerical error of linear constraints (we minimize the norm of the projection $A_{\mathcal{K}}P$). However, the best candidates satisfy similar nice properties like the generalized models, namely they form a compact and convex set, which corresponds to the usual models if $\mathcal{K}$ is consistent. In order to reason with inconsistent knowledge, Daniel replaced the models with the best candidates and selected the best candidate maximizing entropy (this corresponds to another generalization of probabilistic model selection under maximum entropy). He showed that this approach satisfies several principles, which transfer Paris principles for inference processes (Paris, 1994) to inconsistent knowledge bases, see (Daniel, 2009), Proposition 21. These principles include or are closely related to Language Invariance, Independence (Irrelevant Information) and Weak Continuity as considered in this work. To define Continuity, Daniel adapted the Blaschke metric in a similar way as we did, by replacing the models with the best candidates. Computing best candidates is not discussed in greater detail in Daniel (2009), so that it remains unclear how expensive the corresponding optimization problems are and whether there are some computationally attractive choices of the parameters of candidate-based generalized reasoning.

Reasoning with inconsistent information has also been studied extensively for classical logics. Some ideas are to consider additional truth values (Priest, 1991; Arieli et al., 2011), to reason with maximally consistent subsets of the knowledge base (Benferhat et al., 1997), or to consider alternative conjunctives (Konieczny et al., 2005).

Another way to deal with inconsistencies is to repair the knowledge base. In (Rödder and Xu, 1999) three approaches have been proposed to restore consistency in a probabilistic knowledge base. Roughly speaking, the first idea is to relax the conditionals' probabilities to intervals, the second to partition the knowledge base into consistent subsets that get merged afterwards. The third approach is similar to minimal violation measures restricted to conditionals with point probabilities, but instead of minimizing $|P(\psi\phi) - pP(\phi)|$ for each ground conditional $(\psi \mid \phi)[p]$, roughly speaking, the log-ratio $\log(\frac{P(\overline{\psi}\phi)}{P(\psi\phi)} \frac{p}{1-p})$ is minimized. The system *Heureka* (Finthammer et al., 2007) implements several heuristic ideas to restore consistency that can be parameterized in different ways. The main ideas are to remove conditionals or to relax probabilistic constraints. A theoretically appealing idea is to repair knowledge bases while minimizing the change in the conditionals' probabilities. This idea has been investigated in (Thimm, 2009, 2013) for the case of changing point probabilities and in (Picado-Muiño, 2011) for the case of relaxing probabilities to intervals. Both frameworks can be seen as special cases of a general AGM-like consistency restoration framework that has been proposed recently (Bona et al., 2016).

For the special case that the knowledge base $\mathcal{K}$ at hand is the union of several consistent knowledge bases, $\mathcal{K} = \biguplus_i \mathcal{K}_i$, several fusion and belief merging approaches have been proposed. The idea in (Kern-Isberner and Rödder, 2003) is basically to relax the constraints and to use the maximum entropy model afterwards to get consistent probabilities for the original knowledge base. Adamcik investigated the problem of merging probabilistic knowledge bases in detail and gave a comprehensive overview of different approaches and their properties (Adamcik, 2014). Some other recent probabilistic belief merging and revision approaches can be found in (Wilmers, 2015; Rens et al., 2016).

8. Summary and Discussion

In this paper, we investigated the generalized probabilistic entailment problem and the generalized probabilistic model selection problem for general linear probabilistic knowledge bases. We unified and extended previous results for propositional probabilistic logics and made them in particular available for relational probabilistic logics as considered in (Lukasiewicz, 1999; Fisseler, 2008; Loh et al., 2010; Kern-Isberner and Thimm, 2010) for instance.

As we showed, our generalized reasoning approaches generalize the corresponding standard approaches for consistent knowledge bases. In particular, they are robust with respect to minor inconsistencies (Consistent Continuity) and if we restrict to dimension-consistent vector norms also with respect to inconsistent knowledge that is independent of the query (Independence). While Daniel proposed a paraconsistent probabilistic reasoning approach with similar nice logical properties (Daniel, 2009), the computational properties of his approach remain unclear (c.f. the discussion in Related Work). In contrast, we can solve our problems by convex programming techniques in general and several interesting special cases can even be solved by linear programming.

However, as knowledge bases become larger, we need additional tools to compute solutions in reasonable time. Whenever we have deterministic integrity constraints, the problem dimension can be reduced significantly, see (Potyka, 2015b), Sections 3.2.2 and 4.2, for a detailed discussion. In general, we can apply column generation techniques to speed up solving the generalized probabilistic entailment problem (Georgakopoulos et al., 1988; Hansen and Perron, 2008; Finger and De Bona, 2011; Cozman and di Ianni, 2013) and ideas similar to belief propagation to speed up solving generalized model selection problems when using optimum entropy as a cost function (Rödder and Meyer, 1996; Schramm and Ertel, 2000).

When considering relational logics, we can often exploit symmetries and apply lifted inference techniques (Poole, 2003; Kersting, 2012; Van den Broeck et al., 2011). In particular, there has been some progress in applying lifted inference techniques to linear programming problems (Mladenov et al., 2012), which might be helpful for solving the generalized entailment problem. Besides these computational issues, future work will be directed towards applications of generalized probabilistic reasoning in domains like decision theory and group decision making (Potyka et al., 2016).

Acknowledgements: We thank the anonymous reviewers for their valuable comments to improve previous versions of this paper.

Adamcik, M., 2014. Collective reasoning under uncertainty and inconsistency. Ph.D. thesis, University of Manchester.

Arieli, O., Avron, A., Zamansky, A., 2011. What is an ideal logic for reasoning with inconsistency?. In: IJCAI. pp. 706–711.

Baroni, P., Caminada, M., Giacomin, M., 2011. An Introduction to Argumentation Semantics. The Knowledge Engineering Review 26 (4), 365–410.

Benferhat, S., Dubois, D., Prade, H., 1997. Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study part 1: The flat case. Studia Logica 58 (1), 17–45.

Béziau, J.-Y., Carnielli, W., Gabbay, D. (Eds.), 2007. Handbook of Paraconsistency. College Publications, London.

Bona, G. D., Cozman, F. G., Finger, M., 2014. Towards classifying propositional probabilistic logics. Journal of Applied Logic 12 (3), 349–368.

Bona, G. D., Finger, M., 2015. Measuring inconsistency in probabilistic logic: rationality postulates and Dutch book interpretation. Artificial Intelligence 227, 140–164.

Bona, G. D., Finger, M., Ribeiro, M. M., Santos, Y. D., Wassermann, R., 2016. Consolidating probabilistic knowledge bases via belief contraction. In: Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning (KR'16). pp. 125–134.

Boyd, S. P., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Charnes, A., Cooper, W. W., 1962. Programming with linear fractional functionals. Naval Research logistics quarterly 9 (3-4), 181–186.

Cholvy, L., Hunter, A., 1997. Information fusion in logic: A brief overview. In: Qualitative and Quantitative Practical Reasoning (ECSQARU'97/FAPR'97). Vol. 1244 of Lecture Notes in Computer Science. Springer, pp. 86–95.

Cozman, F. G., di Ianni, L. F., 2013. Probabilistic satisfiability and coherence checking through integer programming. In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty. Springer, pp. 145–156.

Daniel, L., 2009. Paraconsistent Probabilistic Reasoning. Ph.D. thesis, L'École Nationale Supérieure des Mines de Paris.

Finger, M., De Bona, G., 2011. Probabilistic satisfiability: Logic-based algorithms and phase transition. In: IJCAI. pp. 528–533.

Finthammer, M., Kern-Isberner, G., Ritterskamp, M., 2007. Resolving Inconsistencies in Probabilistic Knowledge Bases. In: KI 2007: Advances in Artificial Intelligence. pp. 114–128.

Fisseler, J., 2008. Toward Markov logic with conditional probabilities. In: Wilson, D. C., Lane, H. C. (Eds.), Proceedings of the 21st International FLAIRS Conference. AAAI Press, pp. 643–648.

Gabbay, D. M., Hogger, C. J., Robinson, J. A., 1994. Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning. Clarendon Press.

Gelfond, M., Leone, N., June 2002. Logic programming and knowledge representation - the a-prolog perspective. Artificial Intelligence 138 (1-2), 3–38.

Georgakopoulos, G., Kavvadias, D., Papadimitriou, C. H., 1988. Probabilistic satisfiability. Journal of complexity 4 (1), 1–11.

Grant, J., Hunter, A., 2006. Measuring Inconsistency in Knowledgebases. Journal of Intelligent Information Systems 27, 159–184.

Grant, J., Hunter, A., 2013. Measuring the good and the bad in inconsistent information. In: Proceedings IJCAI 2013. p. 26322637.

Hailperin, T., 1986. Boole's Logic and Probability, 2nd Edition. Vol. 85 of Studies in Logic and the foundations of mathematics. Elsevier.

Hansen, P., Jaumard, B., 2000. Probabilistic satisfiability. In: Kohlas, J., Moral, S. (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems. Vol. 5 of Handbook of Defeasible Reasoning and Uncertainty Management Systems. Springer Netherlands, pp. 321–367.

Hansen, P., Perron, S., 2008. Merging the local and global approaches to probabilistic satisfiability. International Journal of Approximate Reasoning 47 (2), 125–140.

Hansson, S. O., 2001. A Textbook of Belief Dynamics. Kluwer Academic Publishers, Norwell, MA, USA.

Kazama, J., Tsujii, J., 2005. Maximum entropy models with inequality constraints: A case study on text categorization. Machine Learning 60 (1-3), 159–194.

Kern-Isberner, G., 2001. Conditionals in Nonmonotonic Reasoning and Belief Revision. No. 2087 in LNCS. Springer-Verlag.

Kern-Isberner, G., Rödder, W., 2003. Belief Revision and Information Fusion in a Probabilistic Environment. In: Proceedings of the 16th International FLAIRS Conference, (FLAIRS'03). AAAI Press, pp. 506–510.

Kern-Isberner, G., Thimm, M., 2010. Novel semantical approaches to relational probabilistic conditionals. In: KR 2010. p. 382392.

Kersting, K., 2012. Lifted probabilistic inference. In: ECAI. pp. 33–38.

Konieczny, S., Lang, J., Marquis, P., 2005. Reasoning under inconsistency: the forgotten connective. In: International Joint Conference on Artificial Intelligence 2005 (IJCAI'05). Vol. 19. p. 484.

Konieczny, S., Pérez, R. P., 1998. On the logic of merging. In: Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98). Morgan Kaufmann, p. 488498.

Loh, S., Thimm, M., Kern-Isberner, G., 2010. On the problem of grounding a relational probabilistic conditional knowledge base. In: Proceedings of the 14th International Workshop on Non-Monotonic Reasoning (NMR 2010), Toronto, Canada (May 2010).

Lukasiewicz, T., 1999. Probabilistic deduction with conditional constraints over basic events. JAIR 10, 380–391.

Lukasiewicz, T., Jul. 2001. Probabilistic logic programming with conditional constraints. ACM Trans. Comput. Logic 2 (3), 289–339.

Lukasiewicz, T., Kern-Isberner, G., 1999. Probabilistic Logic Programming under Maximum Entropy. In: Proceedings ECSQARU-99. Vol. 1638. Springer Lecture Notes in Artificial Intelligence, pp. 279–292.

Matousek, J., Gärtner, B., 2007. Understanding and Using Linear Programming. Universitext (1979). Springer.

Mladenov, M., Ahmadi, B., Kersting, K., 2012. Lifted linear programming. In: AISTATS. pp. 788–797.

Nilsson, N. J., February 1986. Probabilistic logic. Artificial Intelligence 28, 71–88.

Nute, D., 1994. Defeasible logic. In: Gabbay, D. M., J., H. C., A., R. J. (Eds.), Handbook of logic in artificial intelligence and logic programming (vol. 3). Oxford University Press, Inc., New York, NY, USA, pp. 353–395.

Nute, D., Cross, C., 2002. Conditional logic. In: Gabbay, D., Guenther, F. (Eds.), Handbook of Philosophical Logic, second edition Edition. Vol. 4. Kluwer Academic Publishers, pp. 1–98.

Paris, J. B., 1994. The Uncertain Reasoner's Companion – A Mathematical Perspective. Cambridge University Press.

Picado-Muiño, D., 2011. Measuring and Repairing Inconsistency in Probabilistic Knowledge Bases. International Journal of Approximate Reasoning.

Poole, D., 2003. First-order probabilistic inference. In: IJCAI. Vol. 3. Citeseer, pp. 985–991.

Potyka, N., 2014. Linear programs for measuring inconsistency in probabilistic logics. In: Proceedings of the 14th International Conf. on Principles of Knowledge Representation and Reasoning (KR'14). pp. 568–577.

Potyka, N., 2015a. Reasoning over linear probabilistic knowledge bases with priorities. In: International Conference on Scalable Uncertainty Management. Springer, pp. 121–136.

Potyka, N., 2015b. Solving reasoning problems for probabilistic conditional logics with consistent and inconsistent information. Ph.D. thesis, Hagen.

Potyka, N., 2016. Relationships between semantics for relational probabilistic conditional logics. In: Computational Models of Rationality, Essays dedicated to Gabriele Kern-Isberner. College Publications, pp. 332–347.

Potyka, N., Acar, E., Thimm, M., Stuckenschmidt, H., 2016. Group decision making via probabilistic belief merging. In: International Joint Conference on Artificial Intelligence 2016 (IJCAI'16). pp. 3623–3629.

Potyka, N., Beierle, C., Kern-Isberner, G., 2015. A concept for the evolution of relational probabilistic belief states and the computation of their changes under optimum entropy semantics. Journal of Applied Logic 13 (4), 414–440.

Potyka, N., Thimm, M., 2014. Consolidation of probabilistic knowledge bases by inconsistency minimization. In: Proceedings of the 21st European Conference on Artificial Intelligence (ECAI'14). pp. 729–734.

Potyka, N., Thimm, M., 2015. Probabilistic reasoning with inconsistent beliefs using inconsistency measures. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15). pp. 3156–3163.

Priest, G., 1991. Minimally inconsistent lp. Studia Logica 50 (2), 321–331.

Reiter, R., 1980. A Logic for Default Reasoning. Artificial Intelligence 13 (1–2), 81–132.

Rens, G., Meyer, T. A., Casini, G., 2016. On revision of partially specified convex probabilistic belief bases. In: 22nd European Conference on Artificial Intelligence (ECAI 2016). pp. 921–929.

Rödder, W., Meyer, C.-H., 1996. Coherent knowledge processing at maximum entropy by spirit. In: Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., pp. 470–476.

Rödder, W., Xu, L., 1999. Entropy-driven inference and inconsistency. In: AISTATS.

Schramm, M., Ertel, W., 2000. Reasoning with probabilities and maximum entropy: the system pit and its application in lexmed. In: Operations Research Proceedings 1999. Springer, pp. 274–280.

Slota, M., Leite, J., 2012. On Semantic Update Operators for Answer-Set Programs. In: Coelho, H., Studer, R., Wooldridge, M. (Eds.), Proceedings of the 19th European Conference on Artificial Intelligence (ECAI'10). IOS Press, pp. 957–962.

Steele, J. M., 2004. The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities. Cambridge University Press.

Thimm, M., 2009. Measuring Inconsistency in Probabilistic Knowledge Bases. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence.

Thimm, M., April 2013. Inconsistency measures for probabilistic logics. Artificial Intelligence 197, 1–24.

Van den Broeck, G., Taghipour, N., Meert, W., Davis, J., De Raedt, L., 2011. Lifted probabilistic inference by first-order knowledge compilation. In: IJCAI. pp. 2178–2185.

Wilmers, G., 2015. A foundational approach to generalising the maximum entropy inference process to the multi-agent context. Entropy 17 (2), 594–645.

Yeung, R. W., 2008. Information theory and network coding. Springer Science & Business Media.

AppendixA. Notation

1_ϕ	Indicator function	p. 6
$\mathcal{L}(\mathcal{X})$	Language over $\mathcal{X}$	p. 4
$\mathsf{Mod}(\phi)$	Set of models of ϕ	p. 4
a_F	Row vector representing formula F	p. 15
$A_\mathcal{K}$	Constraint matrix corresponding to $\mathcal{K}$	p. 6
c	Linear probabilistic constraint	p. 5
$\mathcal{C}$	Cost function on probability distributions	p. 9
$\mathcal{IC}$	Integrity constraints	p. 11
$\mathcal{I}_{\mathcal{IC}}^{\Vert.\Vert}(\mathcal{K})$	Minimal violation value of $\mathcal{K}$	p. 12
$\mathsf{GMod}_{\mathcal{IC}}^{\Vert.\Vert}(\mathcal{K})$	Generalized models of $\mathcal{K}$ wrt. $\mathcal{IC}$ and $\Vert.\Vert$	p. 13
$\mathcal{M}_{\mathcal{C}}^{\Vert.\Vert}(\mathcal{K},\mathcal{IC})$	Best model of $\mathcal{K}$ and $\mathcal{IC}$ wrt. $\mathcal{C}$ and $\Vert.\Vert$	p. 35
$\mathcal{M}_{\mathcal{C}}(\mathcal{K})$	Best model of $\mathcal{K}$ and $\mathcal{IC}$ wrt. probabilistic entailment	p. 35
$\models$	General satisfaction relation	p. 4
$\models^{pe}$	Probabilistic entailment	p. 9
$\models^{gpe}$	Generalized probabilistic entailment	p. 14
$\Vert.\Vert$	Vector norm	p. 10
$\Vert.\Vert_p$	p-norm	p. 10
$\Vert.\Vert_1$	Manhattan norm	p. 10
$\Vert.\Vert_2$	Euclidean norm	p. 10
$\Vert.\Vert_\infty$	Maximum norm	p. 10
$\Vert.\Vert_{p,w}$	w-weighted p-norm	p. 11
$\Vert.\Vert_{p,w}^p$	Raised w-weighted p-norm	p. 11
$\Vert.\Vert_{\infty,w}$	w-weighted ∞-norm	p. 11
$\Vert S_1, S_2\Vert_B$	Blaschke distance between sets S_1, S_2	p. 31
$\Vert\mathcal{K}_1, \mathcal{K}_2\Vert_B$	Generalized Blaschke distance between knowledge bases $\mathcal{K}_1$, $\mathcal{K}_2$	p. 31
ω	Possible world in $\mathcal{L}(\mathcal{X})$	p. 4
$\omega\vert_{\mathcal{X}'}$	Restriction of ω to $\mathcal{X}'$	p. 21
(ω_1, ω_2)	Combination of ω_1 and ω_2	p. 21
Ω	Set of possible worlds	p. 4
P	Probability distribution over Ω	p. 5
$P\vert_{\mathcal{X}'}$	Marginal distribution of P wrt. $\mathcal{X}'$	p. 26
$P_1 \odot P_2$	Production distribution	p. 26
$(\phi \mid \psi)$	Conditional query	p. 8
r_c	Row vector representing constraint c	p. 6

$\mathcal{X} = \{X_1, \ldots, X_n\}$ Random variables p. 4