Probabilistic Knowledge Representation Using Gröbner Basis Theory

Gabriele Kern-Isberner and **Marco Wilhelm**
Technische Universität Dortmund
Dortmund, Germany

Christoph Beierle
FernUniversität Hagen
Hagen, Germany

Abstract

An often used methodology for reasoning with probabilistic conditional knowledge bases is provided by the principle of maximum entropy (so-called MaxEnt principle) that realises an idea of informational economy. In this paper we exploit the fact that MaxEnt distributions can be computed by solving nonlinear equation systems that reflect the conditional logical structure of these distributions. We apply the theory of Gröbner bases that is well known from computational algebra to the polynomial system which is associated with a MaxEnt distribution, in order to obtain results for reasoning with maximum entropy. First, a necessary condition for knowledge bases to be consistent is derived. Furthermore, approaches to answering MaxEnt queries are presented by demonstrating how inferring the MaxEnt probability of a single conditional from a given knowledge base is possible. Finally, we discuss computational methods to establish general MaxEnt inference rules.

1 Introduction

Probability theory provides one of the most powerful and widely used frameworks for uncertain reasoning to date. However, it is not only its flexibility by using a continuous range of possible degrees of uncertainties for knowledge representation – actually, this has also been brought forth often as a point of criticism – but rather its foundations relying on a stringent additive structure and the appealing ways how to combine qualitative and quantitative aspects of uncertainties (e.g., by Bayesian networks (Pearl 1988)) that allows uncertain reasoning of high quality. Knowledge bases of probabilistic conditionals of the kind "*If A, then B with probability x*", formally represented by $(B|A)[x]$, are a concise means to make relevant knowledge explicit, and inductive reasoning methodologies like probabilistic networks, or the principle of maximum entropy (Paris and Vencovská 1990; Jaynes 1983), so-called *MaxEnt principle*, provide convenient frameworks for answering queries.

This paper focusses on the MaxEnt principle as a most appropriate form of commonsense probabilistic reasoning (Paris 1999). It has long been known that MaxEnt distributions arise as unique solutions to nonlinear equation systems (Jaynes 1983). In (Kern-Isberner 2001), a more abstract equation system has been derived that motivated the algebraic theory of conditional structures which, in particular, allows for abstract logical investigations of inference based on MaxEnt distributions (cf., e.g., (Kern-Isberner 1998)) and has been applied to qualitative and first-order environments (cf., e.g., (Kern-Isberner and Thimm 2012)) as well.

In this paper, we present an approach to address computational issues of MaxEnt reasoning that takes profit from these algebraic ways of looking at MaxEnt. We apply the theory of Gröbner bases to find MaxEnt probabilities via polynomial systems. Gröbner bases have been introduced by Bruno Buchberger first and have emerged as an important subfield of computational algebra (cf. (Buchberger 2006)). They allow the representation of solutions of polynomial systems in a normalized form so that questions concerning such solutions can be reduced to investigating properties of the appertaining Gröbner basis. We make use of these algebraic connections to derive a necessary condition for the consistency of a probabilistic knowledge base. Moreover, we show how Gröbner basis techniques help answering queries to probabilistic knowledge bases. More precisely, the probability of any conditional that can be MaxEnt inferred from a given knowledge basis is shown to satisfy a polynomial equation; if linear equations arise then the probability can be computed directly. We elaborate on this latter idea and an even more abstract algebraic representation of MaxEnt distributions to derive automatically symbolic MaxEnt inference rules for the first time. Existing tools for applying the MaxEnt principle, such as SPIRIT (Rödder, Reucher, and Kulmann 2006), are graph based and cannot exploit the structural dependencies of a given probabilistic knowledge base as well as symbolic methods. Dukkipati already showed that Gröbner bases techniques can be used to identify MaxEnt distributions in algebraic varities (cf. (Dukkipati 2009)), but he did not consider conditional knowledge bases nor inferences.

The rest of this paper is organized as follows: In Section 2 basic logical notations are set, and the principle of maximum entropy is recalled. Section 3 gives a short overview of the theory of Gröbner bases. Our approach of applying Gröbner bases to MaxEnt reasoning is presented in Section 4; here, we also prove a necessary condition for knowledge bases to be consistent. Section 5 shows how to answer MaxEnt queries by Gröbner bases, and Section 6 deals with more general MaxEnt inference rules. Section 7 concludes the paper with a short summary and an outlook.

2 Basics of Knowledge Representation

We consider a propositional language $\mathcal{L} = \mathcal{L}(\mathcal{V})$ over a finite alphabet $\mathcal{V} = \{a, b, c, \ldots\}$. Roman uppercase letters $A, B, C, \ldots$ denote atoms or formulas in $\mathcal{L}$. The language $\mathcal{L}$ is equipped with the usual logical connectives $\wedge$ (*and*), $\vee$ (*or*) and $\neg$ (*negation*). To shorten mathematical formulas we write AB instead of $A \wedge B$ and $\overline{A}$ instead of $\neg A$. By introducing the conditional operator $|$, a conditional language $(\mathcal{L}|\mathcal{L}) = \{(B|A) \mid A, B \in \mathcal{L}\}$ is defined. Let Ω be the set of all possible worlds ω; here, Ω is simply a complete set of interpretations of $\mathcal{L}$. If a world ω satisfies a formula A, we write $\omega \models A$ and call ω a *model* of A, i.e., $\omega \in Mod(A)$. Usually, we identify each possible world ω with the minterm (or complete conjunction) that has exactly ω as a model.

Atoms in $\mathcal{L}$ can be understood as propositional variables, and possible worlds as elementary events. Therefore, when considering a probability distribution $\mathcal{P}$ over $\mathcal{V}$, every $A \in \mathcal{L}(\mathcal{V})$ can be assigned a probability via $\mathcal{P}(A) = \sum_{\omega \models A} \mathcal{P}(\omega)$. In this way we obtain a probabilistic interpretation of $\mathcal{L}$. Finally, we obtain a probabilistic conditional language

$$(\mathcal{L}|\mathcal{L})^{prob} = \{(B|A)[x] \mid (B|A) \in (\mathcal{L}|\mathcal{L}),\ x \in [0,1]\}$$

by adding a probability to every conditional. Elements in $(\mathcal{L}|\mathcal{L})^{prob}$ are called *(probabilistic) conditionals*. Conditionals are interpreted by distributions via conditional probabilities. If $\mathcal{P}$ is a distribution, satisfaction of a conditional by $\mathcal{P}$ is defined by

$$\mathcal{P} \models (B|A)[x] \text{ iff } \mathcal{P}(A) > 0 \text{ and } x = \mathcal{P}(B|A) = \frac{\mathcal{P}(AB)}{\mathcal{P}(A)}.$$

For a distribution $\mathcal{P}$ and a set $\mathcal{C} \subseteq (\mathcal{L}|\mathcal{L})^{prob}$ of conditionals, $\mathcal{P}$ satisfies $\mathcal{C}$, denoted by $\mathcal{P} \models \mathcal{C}$, iff $\mathcal{P} \models (B|A)[x]$ for all $(B|A)[x] \in \mathcal{C}$. A set of conditionals $\mathcal{C} \subseteq (\mathcal{L}|\mathcal{L})^{prob}$ is called *consistent* if there exists a distribution $\mathcal{P}$ with $\mathcal{P} \models \mathcal{C}$. A finite set of conditionals

$$\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\} \subseteq (\mathcal{L}|\mathcal{L})^{prob}$$

is called a *knowledge base*, and there may exist several (or none) distributions satisfying it.

In order to use inductively the information in $\mathcal{KB}$ it is very helpful to choose a "best" model of $\mathcal{KB}$. The *principle of maximum entropy* (cf. (Kern-Isberner 1998) and (Paris 1994)) provides a well-known solution to this problem. The idea is to fulfill the paradigm of *informational economy*, i.e., of least amount of assumed information (cf. (Gärdenfors 1988)). Therefore, one maximizes the entropy

$$H(\mathcal{Q}) = -\sum_{\omega \in \Omega} \mathcal{Q}(\omega) \log \mathcal{Q}(\omega)$$

of a distribution $\mathcal{Q}$ with $\mathcal{Q}$ being a model of $\mathcal{KB}$. It can be shown that for every consistent knowledge base $\mathcal{KB}$ such a distribution $\mathcal{ME}(\mathcal{KB})$ with maximal entropy exists and, in particular, $\mathcal{ME}(\mathcal{KB})$ is unique (cf. (Paris 1994)). It immediately follows that $\mathcal{KB}$ is consistent iff $\mathcal{ME}(\mathcal{KB})$ exists. Taking the conventions $\infty^0 = 1$, $\infty^{-1} = 0$ and $0^0 = 1$ into account, the distribution $\mathcal{ME}(\mathcal{KB})$ is given by

$$\mathcal{ME}(\mathcal{KB})(\omega) = \alpha_0 \prod_{\substack{1 \le i \le n \\ \omega \models A_i B_i}} \alpha_i^{1-x_i} \prod_{\substack{1 \le i \le n \\ \omega \models A_i \overline{B_i}}} \alpha_i^{-x_i} \qquad (1)$$

with a normalizing constant α_0 and *effects*

$$\alpha_i \begin{cases} = \infty, & \text{iff } x_i = 1 \\ \in (0, \infty), & \text{iff } x_i \in (0,1) \\ = 0, & \text{iff } x_i = 0 \end{cases},$$

associated with each conditional and solving

$$\begin{aligned} (1 - x_i)\, \alpha_i^{1-x_i} \sum_{\omega \models A_i B_i} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \omega \models A_j \overline{B_j}}} \alpha_j^{-x_j} \\ = x_i\, \alpha_i^{-x_i} \sum_{\omega \models A_i \overline{B_i}} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} \alpha_j^{1-x_j} \prod_{\substack{j \neq i \\ \omega \models A_j \overline{B_j}}} \alpha_j^{-x_j} \end{aligned} \qquad (2)$$

for $1 \le i \le n$ (cf. (Kern-Isberner 2001)). This logic-based representation of MaxEnt distributions will be the starting point for our application of Gröbner basis theory. If $\mathcal{ME}(\mathcal{KB})$ exists then we can compute the MaxEnt probability of any further conditional $\mathcal{C}_{\inf} = (B_{n+1}|A_{n+1})$ from $\mathcal{ME}(\mathcal{KB})$. This yields a (nonmonotonic) *MaxEnt inference relation* $\mid\!\sim_{\mathcal{ME}}$ with

$$\mathcal{KB} \mid\!\sim_{\mathcal{ME}} \mathcal{C}_{\inf}[\xi_{n+1}] \text{ iff } \mathcal{ME}(\mathcal{KB}) \models \mathcal{C}_{\inf}[\xi_{n+1}].$$

3 Basics of Gröbner Basis Theory

Gröbner bases are specific generating sets of ideals in polynomial rings that allow to condense information given by algebraic specifications of problems; a recommendable reference for the material presented in this section is (Cox, Little, and O'Shea 2007). Let $\mathbb{C}[\mathcal{Y}] = \mathbb{C}[y_1, \ldots, y_n]$ be the polynomial ring in the variables collected in the set $\mathcal{Y} = \{y_1, \ldots, y_n\}$ over the algebraically closed field of complex numbers $\mathbb{C}$. $\mathbb{C}[\mathcal{Y}]$ may be understood as a vector space over $\mathbb{C}$ with basis

$$\mathcal{T} = \{y_1^{m_1} \cdots y_n^{m_n} \mid m_1, \ldots, m_n \in \mathbb{N}_0\}.$$

Note that $1 = y_1^0 \cdots y_n^0 \in \mathcal{T}$. An element $t \in \mathcal{T}$ is called a *term*. Therefore, every polynomial $f \in \mathbb{C}[\mathcal{Y}]$ can be written as a finite linear combination of terms $t \in \mathcal{T}$ over $\mathbb{C}$. The set of terms occurring in a polynomial $f \in \mathbb{C}[\mathcal{Y}]$ with non-vanishing coefficients is called *support* of f, written as $supp(f)$. Terms $t \in \mathcal{T}$ can be embedded into $\mathbb{C}[\mathcal{Y}]$ as monomials with coefficient 1. An element $(\eta_1, \ldots, \eta_n) \in \mathbb{C}^n$ is called a *root* (or a *zero*) of a polynomial $f \in \mathbb{C}[\mathcal{Y}]$ iff $f(\eta_1, \ldots, \eta_n) = 0$.

Definition 1 (Ideal). *A subset $\mathcal{I} \subseteq \mathbb{C}[\mathcal{Y}]$ is a* (polynomial) ideal, *denoted by $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$, iff $0 \in \mathcal{I}$ and, in addition, for all $f, g \in \mathcal{I}, h \in \mathbb{C}[\mathcal{Y}]$ also $f + g \in \mathcal{I}$ as well as $h\, f \in \mathcal{I}$.*

Let $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ be an ideal, and let $\mathcal{F} \subseteq \mathcal{I}$ so that for all $f \in \mathcal{I}$, there are $f_1, \ldots, f_m \in \mathcal{F}, h_1, \ldots, h_m \in \mathbb{C}[\mathcal{Y}]$ such that $f = \sum_{i=1}^m h_i\, f_i$ holds. Then $\mathcal{F}$ is called a *generating set* of $\mathcal{I}$, written as $\mathcal{I} = \langle \mathcal{F} \rangle$. Obviously, the ideal $\mathcal{I}$ only consists of polynomials that vanish in the common roots of the polynomials in $\mathcal{F}$. Therefore, we can speak of the common roots of $\mathcal{I}$, which are exactly the same as the common roots of $\mathcal{F}$. Special ideals are the *zero ideal* $\langle \{0\} \rangle$

only consisting of the polynomial being constantly zero, and $\langle\{1\}\rangle$ that is in fact the entire polynomial ring $\mathbb{C}[\mathcal{Y}]$. Since $1 \in \langle\{1\}\rangle$, there are no common roots of $\langle\{1\}\rangle$.

In order to understand the fundamental importance of Gröbner bases, imagine that the problem under consideration can be described by a set $\mathcal{F}$ of polynomials, and the solutions of the problem correspond to the roots of $\mathcal{F}$. Then, the ideal generated by $\mathcal{F}$ provides an algebraic context to condense the problem description without changing (essentially) the solutions of the problem. Gröbner bases are special generating sets of ideals $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ that condense the problem description best in a specific sense that is made precise in the following. For this, we need specific orderings on the set of terms $\mathcal{T}$.

Definition 2 (Term Ordering)**.** *Let $\preceq$ be a total ordering with related strict ordering $\prec$ on $\mathcal{T}$. Then $\preceq$ is a* term ordering *iff for all $t \in \mathcal{T}$ we have $1 \preceq t$, and for all $t, t_1, t_2 \in \mathcal{T}$, $t_1 \preceq t_2$ implies $t\, t_1 \preceq t\, t_2$.*

An important class of term orderings are the *lexicographical term orderings*. A lexicographical term ordering $\preceq_{\text{lex}}$ on the set of terms $\mathcal{T}$ is recursively defined by

$$y_1^{e_1} \cdots y_n^{e_n} \preceq_{\text{lex}} y_1^{f_1} \cdots y_n^{f_n} \text{ iff } e_1 < f_1$$
$$\text{or } e_1 = f_1 \wedge y_2^{e_2} \cdots y_n^{e_n} \preceq_{\text{lex}} y_2^{f_2} \cdots y_n^{f_n}$$

presupposing that $y_n \prec_{\text{lex}} y_{n-1} \prec_{\text{lex}} \ldots \prec_{\text{lex}} y_1$. Of course, $y_1^{e_1} \cdots y_n^{e_n} = y_1^{f_1} \cdots y_n^{f_n}$ holds iff $e_i = f_i$ for all $1 \leq i \leq n$. Permuting the order between variables generates different lexicographical term orderings; there are $n!$ different lexicographical term orderings on $\mathcal{T}$.

Definition 3 (Leading Coefficient, Leading Monomial)**.** *Let $\preceq$ be a term ordering on $\mathcal{T}$ and $f = \sum_{i=1}^{m} c_i\, t_i \in \mathbb{C}[\mathcal{Y}]$ with $c_i \in \mathbb{C} \setminus \{0\}$ and $t_i \in \mathcal{T}$ for $1 \leq i \leq m$. Further, let $t_1 \prec \cdots \prec t_m$. The* leading coefficient *of f is $lc_{\preceq}(f) = c_m$, and the* leading monomial *of f is $lm_{\preceq}(f) = c_m\, t_m$. For a subset $\mathcal{F} \subseteq \mathbb{C}[\mathcal{Y}]$ we denote the set of the leading monomials of all polynomials in $\mathcal{F}$ by $lm_{\preceq}(\mathcal{F})$.*

Now we have all formal prerequisites to recall the definition of Gröbner bases.

Definition 4 (Gröbner Basis)**.** *Let $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ be an ideal with $\mathcal{I} \neq \langle\{0\}\rangle$ and let $\preceq$ be a term ordering on $\mathcal{T}$. A subset $\mathcal{B}_{\preceq} = \{b_1, \ldots, b_m\} \subseteq \mathcal{I}$ is called a* Gröbner basis for $\mathcal{I}$ with respect to $\preceq$ *iff*

$$\langle lm_{\preceq}(\mathcal{B}_{\preceq})\rangle = \langle\{lm_{\preceq}(f) \mid f \in \mathcal{I}\}\rangle. \tag{3}$$

In particular, $\mathcal{I} = \langle\mathcal{B}_{\preceq}\rangle$ holds, i.e., $\mathcal{B}_{\preceq}$ is a generating set of $\mathcal{I}$. $\mathcal{B}_{\preceq}$ is called a minimal *Gröbner basis for $\mathcal{I}$ with respect to $\preceq$ iff, in addition, $t \notin \langle\{lm(\mathcal{B}_{\preceq} \setminus \{b_i\})\}\rangle$ as well as $lc(b_i) = 1$ hold for all $t \in supp(b_i)$ and $1 \leq i \leq m$.*

As a consequence of the Hilbert's Basis Theorem (cf. (Cox, Little, and O'Shea 2007)), every ideal $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ with $\mathcal{I} \neq \langle\{0\}\rangle$ has a *unique* minimal Gröbner basis with respect to a given term ordering $\preceq$, written as $\mathcal{GB}_{\preceq}(\mathcal{I})$. The standard method to calculate Gröbner bases is *Buchberger's algorithm* that is implemented in all current computer algebra systems such as Maple or Mathematica. To gain a first insight, again (Cox, Little, and O'Shea 2007) is recommended. We will see examples of this theory later on in Section 4.

Furthermore, to allow for useful cancellations, we need the concept of saturation. For this, we consider the extended polynomial ring $\mathbb{C}[y_1, \ldots, y_n, s_1, \ldots, s_u]$ with disjoint sets of variables $\mathcal{Y} = \{y_1, \ldots, y_n\}$ and $\mathcal{S} = \{s_1, \ldots, s_u\}$; the variables in $\mathcal{S}$ are auxiliary variables that allow to take additional information into regard.

Definition 5 (Elimination Term Ordering)**.** *A term ordering $\preceq$ on the set of terms consisting of the variables in $\mathcal{Y}$ and $\mathcal{S}$ is called an* elimination term ordering for $\mathcal{S}$ *iff for all polynomials $f \in \mathbb{C}[y_1, \ldots, y_n, s_1, \ldots, s_u]$, $lm(f) \in \mathbb{C}[\mathcal{Y}]$ implies $f \in \mathbb{C}[\mathcal{Y}]$.*

For any disjoint sets $\mathcal{Y}$ and $\mathcal{S}$ of variables such an elimination term ordering exists.

Definition 6 (Elimination Ideal)**.** *Given an ideal $\mathcal{I}$ in the extended polynomial ring $\mathbb{C}[y_1, \ldots, y_n, s_1, \ldots, s_u]$, the intersection $\mathcal{I} \cap \mathbb{C}[\mathcal{Y}]$ is still an ideal, called* elimination ideal of $\mathcal{I}$ for $\mathcal{S}$.

In order to determine $\mathcal{I} \cap \mathbb{C}[\mathcal{Y}]$, one derives a minimal Gröbner basis for $\mathcal{I}$ with respect to an elimination term ordering for $\mathcal{S}$. By deleting all polynomials with terms containing a variable of $\mathcal{S}$, one obtains a Gröbner basis for $\mathcal{I} \cap \mathbb{C}[\mathcal{Y}]$ (cf. (Cox, Little, and O'Shea 2007)), i.e., for the original solution space. Now, let $\mathcal{I} = \langle\{f_1, \ldots, f_m\}\rangle$ be an ideal in $\mathbb{C}[\mathcal{Y}]$, and $h \in \mathbb{C}[\mathcal{Y}]$.

Definition 7 (Saturation)**.** *The set of polynomials*

$$(\mathcal{I} : h^*) = \{f \in \mathbb{C}[\mathcal{Y}] \mid \exists\, r \in \mathbb{N}_0 : f\, h^r \in \mathcal{I}\} \supseteq \mathcal{I}$$

is called the saturation of $\mathcal{I}$ with respect to h. *Every saturation is an ideal.*

To calculate the saturation $(\mathcal{I} : h^*)$, we consider the ideal $\mathcal{J} = \langle\{f_1, \ldots, f_m, 1 - s_1\, h\}\rangle \trianglelefteq \mathbb{C}[y_1, \ldots, y_n, s_1]$; one can show that $(\mathcal{I} : h^*) = \mathcal{J} \cap \mathbb{C}[\mathcal{Y}]$ holds. For a finite set $\mathcal{H} = \{h_1, \ldots, h_m\} \subseteq \mathbb{C}[\mathcal{Y}]$ of polynomials, we write

$$\mathcal{SAT}_{\mathcal{H}}(\mathcal{I}) := (((\mathcal{I} : h_1^*) : \ldots) : h_m^*).$$

The order of the polynomials $h_1, \ldots, h_m$ is irrelevant. In particular, calculating the saturation $\mathcal{SAT}_{\mathcal{Y}}(\mathcal{I})$ for a given ideal $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ means cancelling terms in the polynomials in $\mathcal{I}$. The polynomials in $\mathcal{SAT}_{\mathcal{Y}}(\mathcal{I})$ have the same common roots as the polynomials in $\mathcal{I}$, except for "trivial" roots containing at least one entry that is zero.

Concluding this section we recall a fact (cf. (Cox, Little, and O'Shea 2007) for a proof) and an interesting corollary that are important to clarify whether the given problem description has a solution at all.

Theorem 1 (Weak Nullstellensatz)**.** *Let $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ be an ideal having no common roots. Then $\mathcal{I} = \mathbb{C}[\mathcal{Y}]$.*

Corollary 1. *Let $\mathcal{I} \trianglelefteq \mathbb{C}[\mathcal{Y}]$ be an ideal and $\preceq$ an arbitrary term ordering on $\mathcal{T}$. Then $\mathcal{I}$ has common roots iff $\mathcal{I} \neq \langle\{1\}\rangle$, i.e., iff $\mathcal{GB}_{\preceq}(\mathcal{I}) \neq \{1\}$.*

So, checking algebraic problem descriptions for consistency reduces to having non-trivial Gröbner bases.

4 Compiling Probabilistic Knowledge Bases to Gröbner Bases

Let

$$\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\}$$

be a probabilistic conditional knowledge base with given probabilities. For our further considerations in this paper we assume the probabilities ξ_i to be rational numbers and non-trivial, i.e., $\xi_i \in (0,1)$ for $1 \le i \le n$. Thus, each probability ξ_i is equal to the quotient of two natural numbers. We therefore introduce new variables $p_1, \ldots, p_n, q_1, \ldots, q_n$ over $\mathbb{N}$ and call

$$\widehat{\mathcal{KB}} := \{(B_1|A_1)[p_1/q_1], \ldots, (B_n|A_n)[p_n/q_n]\}$$

the *skeleton* of $\mathcal{KB}$. Indeed, conditional probability constraints in practical applications are usually rational, and cases where $\xi_i = 0$ respectively $\xi_i = 1$ for some $1 \le i \le n$ occur can be treated similarly (cf. (Kern-Isberner 2001)). For compiling $\mathcal{KB}$ to a Gröbner basis, we distinguish three compilation phases, where the first phase abstracts from the concrete probabilities in $\mathcal{KB}$ by just considering the skeleton $\widehat{\mathcal{KB}}$.

Phase 1 For applying the theory of Gröbner bases it is necessary to transform (2) into a *polynomial* nonlinear equation system. Since the probabilities ξ_i are assumed to be rational, we use the skeleton $\widehat{\mathcal{KB}}$ and apply the substitution $\alpha_i = y_i^{q_i}$. Then by multiplying both sides with $q_i\, y_i^{p_i} \prod_{j \ne i} y_j^{p_j}$, (2) leads to

$$\begin{aligned} f_i := (q_i - p_i)\, y_i^{q_i} \sum_{\omega \models A_i B_i} \prod_{\substack{j \ne i \\ \omega \models A_j B_j}} y_j^{q_j} \prod_{\substack{j \ne i \\ \omega \models \overline{A_j}}} y_j^{p_j} \\ -p_i \sum_{\omega \models A_i \overline{B_i}} \prod_{\substack{j \ne i \\ \omega \models A_j B_j}} y_j^{q_j} \prod_{\substack{j \ne i \\ \omega \models \overline{A_j}}} y_j^{p_j} = 0 \end{aligned} \tag{4}$$

for $1 \le i \le n$. Note that $\alpha_i = 0$ and therefore $y_i = 0$ iff $\xi_i = 0$ for $1 \le i \le n$. As we concentrate on knowledge bases with non-trivial probabilities $\xi_i \in (0,1)$, this transformation of (2) does not mean any loss of information, and we will ignore the trivial roots of (4). We call (4) the *abstract polynomial system associated with* $\mathcal{KB}$ and

$$F_1(\mathcal{KB}) := \{f_1, \ldots, f_n\}$$

the *set of abstract polynomials associated with* $\mathcal{KB}$.

Example 1 (Compilation 1)**.** *Let a, b, c be propositional variables and $\mathcal{KB}_1 = \{(b|a)[3/4], (c|a)[2/3], (c|ab)[2/3]\}$. Then $F_1(\mathcal{KB}_1) = \{f_1, f_2, f_3\}$ with*

$$\begin{aligned} f_1 &= (q_1 - p_1)\, y_1^{q_1}\, (y_2^{q_2}\, y_3^{q_3} + 1) - p_1\, (y_2^{q_2}\, y_3^{p_3} + y_3^{p_3}), \\ f_2 &= (q_2 - p_2)\, y_2^{q_2}\, (y_1^{q_1}\, y_3^{q_3} + y_3^{p_3}) - p_2\, (y_1^{q_1} + y_3^{p_3}), \\ f_3 &= (q_3 - p_3)\, y_3^{q_3}\, (y_1^{q_1}\, y_2^{q_2}) - p_3\, y_1^{q_1}. \end{aligned}$$

Phase 2 In the second compilation phase, concrete probabilities in $\mathcal{KB}$ are substituted into $F_1(\mathcal{KB})$. For each ξ_i in $\mathcal{KB}$, let $\tilde{p}_i, \tilde{q}_i \in \mathbb{N}$ be the unique numbers such that $\xi_i = \tilde{p}_i/\tilde{q}_i$ and $\tilde{p}_i, \tilde{q}_i$ are relatively prime. Then the result of the second compilation phase is

$$F_2(\mathcal{KB}) := \sigma(F_1(\mathcal{KB}))$$

where σ is the substitution replacing p_i by $\tilde{p}_i$ and q_i by $\tilde{q}_i$. Since (positive and real) common roots of the polynomials in $F_2(\mathcal{KB}) \subseteq \mathbb{C}[\mathcal{Y}]$ characterize the distribution $\mathcal{ME}(\mathcal{KB})$, so do the common roots of the ideal $\langle F_2(\mathcal{KB}) \rangle \trianglelefteq \mathbb{C}[\mathcal{Y}]$, too. Distinguishing these two phases allows for reusing the result $F_1(\mathcal{KB})$ of the first compilation phase that depends only on $\widehat{\mathcal{KB}}$, e.g. when modifying $\mathcal{KB}$ just by changing the probability of one or more conditionals, but keeping the structures of the conditionals; this is a very common situation for a knowledge engineer when developing a probabilistic knowledge base.

Example 2 (Compilation 2)**.** *For $\mathcal{KB}_1$ from Ex. 2, $F_2(\mathcal{KB}_1)$ consists of the following polynomials:*

$$\begin{aligned} & y_1^4\, (y_2^3\, y_3^3 + 1) - 3\, (y_2^3\, y_3^3 + y_3^2), \\ & y_2^3\, (y_1^4\, y_3^3 + y_3^2) - 2\, (y_1^4 + y_3^2), \\ & y_3^3\, (y_1^4\, y_2^3) - 2\, y_1^4. \end{aligned}$$

Phase 3 As already mentioned, we concentrate on knowledge bases with non-trivial probabilities, and therefore, it is allowed to cancel variables $y_1, \ldots, y_n$ in the regarding polynomial system (4). To meet this fact we apply saturation to ideals associated with knowledge bases and obtain

$$F_3(\mathcal{KB}, \preceq) := \mathcal{GB}_{\preceq}(\mathcal{SAT}_{\mathcal{Y}}(\langle F_2(\mathcal{KB}) \rangle))$$

where $\preceq$ is a given term ordering on $\mathcal{T}$.

Example 3 (Compilation 3)**.** *We present a Gröbner basis representation of our knowledge base $\mathcal{KB}_1$. Therefore, we choose the lexicographical term ordering $\preceq_{\mathrm{lex}}$ on $\mathcal{T}$ with $y_3 \prec_{\mathrm{lex}} y_2 \prec_{\mathrm{lex}} y_1$. To illustrate its meaning, we reorder the polynomials in $F_2(\mathcal{KB}_1)$ as well as the terms they consist of and underline their leading monomials:*

$$\begin{aligned} & \underline{y_1^4\, y_2^3\, y_3^3} + y_1^4 - 3\, y_2^3\, y_3^2 - 3\, y_3^2, \\ & \underline{y_1^4\, y_2^3\, y_3^3} - 2\, y_1^4 + y_2^3\, y_3^2 - 2\, y_3^2, \\ & \underline{y_1^4\, y_2^3\, y_3^3} - 2\, y_1^4. \end{aligned}$$

All polynomials have leading coefficient $lc = 1$. The resulting minimal Gröbner basis is

$$F_3(\mathcal{KB}_1, \preceq_{\mathrm{lex}}) = \{y_1^4 - 3\, y_3^2, y_2^3 - 2, y_3^3 - 1\}.$$

As a first result illustrating the useful connection between probabilistic knowledge bases $\mathcal{KB}$ and their ideals, we formulate a condition that can be used for testing knowledge bases for consistency with the help of Gröbner bases.

Theorem 2. *Let $\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\}$ be a consistent knowledge base with non-trivial probabilities and let $\preceq$ be a term ordering on $\mathcal{T}$. Then*

$$F_3(\mathcal{KB}, \preceq) \ne \{1\}. \tag{5}$$

Proof. If $\mathcal{KB}$ is consistent, then the MaxEnt distribution exists and therefore at least one solution $(\eta_1, \ldots, \eta_n) \in \mathbb{R}^n$ of the respective polynomial system. For this solution, $\eta_i \neq 0$ may be assumed since $\xi_i \neq 0$ for $1 \leq i \leq n$. Hence, this solution is a common root of $\langle F_2(\mathcal{KB})\rangle$. Since $\eta_i \neq 0$ for $1 \leq i \leq n$, $(\eta_1, \ldots, \eta_n) \in \mathbb{R}^n$ is also a common root of the polynomials in $\mathcal{SAT}_{\mathcal{Y}}(\langle F_2(\mathcal{KB})\rangle)$. Finally, (5) holds because of Corollary 1. □

The previous result shall be illustrated by some examples.

Example 4 (Consistency)**.** *The knowledge base $\mathcal{KB}_1$ from Ex. 1 is consistent since*

$$F_3(\mathcal{KB}_1, \preceq_{\text{lex}}) = \{y_1^4 - 3\, y_3^2, y_2^3 - 2, y_3^3 - 1\} \neq \{1\}$$

and $(y_1, y_2, y_3) = (\sqrt[4]{3}, \sqrt[3]{2}, 1)$ is a valuable common root of $F_3(\mathcal{KB}_1, \preceq_{\text{lex}})$.

Example 5 (Consistency and Inconsistency)**.** *Obviously, $\mathcal{KB}_2 = \{(b|a)[\xi_1], (\bar{b}|a)[\xi_2]\}$ is consistent iff $\xi_1 + \xi_2 = 1$. As expected, e.g. for $\xi_1 = 1/3$ and $\xi_2 = 2/3$, i.e.,*

$$\mathcal{KB}_{2a} = \{(b|a)[1/3], (\bar{b}|a)[2/3]\},$$

we obtain $F_2(\mathcal{KB}_{2a}) = \{2\, y_1^3 - y_2^3, y_2^3 - 2\, y_1^3\}$. Hence, with the lexicographical term ordering $\preceq_{\text{lex}}$ determined by $y_2 \prec_{\text{lex}} y_1$, we obtain

$$F_3(\mathcal{KB}_{2a}, \preceq_{\text{lex}}) = \{y_1^3 - 1/2\, y_2^3\} \neq \{1\}.$$

On the other hand,

$$\mathcal{KB}_{2b} = \{(b|a)[1/3], (\bar{b}|a)[1/3]\}$$

leads to $F_2(\mathcal{KB}_{2b}) = \{2\, y_1^3 - y_2^3, 2\, y_2^3 - y_1^3\}$, and therefore, we obtain

$$F_3(\mathcal{KB}_{2b}, \preceq_{\text{lex}}) = \{1\}.$$

5 MaxEnt Reasoning for Answering Queries

An important question is what further inferences can be drawn from $\mathcal{KB}$ under the MaxEnt methodology. So, let $\mathcal{C}_{\text{inf}} = (B_{n+1}|A_{n+1})$ be an additional conditional. We investigate for which probability $\xi_{n+1} \in (0, 1)$ we have $\mathcal{KB} \mid\!\sim_{\mathcal{ME}} (B_{n+1}|A_{n+1})[\xi_{n+1}]$. For this, ξ_{n+1} must satisfy

$$\xi_{n+1} = \frac{\mathcal{ME}(\mathcal{KB})(A_{n+1}B_{n+1})}{\mathcal{ME}(\mathcal{KB})(A_{n+1})}. \tag{6}$$

So, if the MaxEnt distribution $\mathcal{ME}(\mathcal{KB})$ is known, i.e., a valuable common root $(\eta_1, \ldots, \eta_n) \in \mathbb{R}^n$ of the respective polynomial system (4), it is possible to derive ξ_{n+1} easily from (6). However, $\mathcal{ME}(\mathcal{KB})$ need not be given before deriving ξ_{n+1}. Instead, one can apply $\mathcal{ME}(\mathcal{KB})$ and calculate ξ_{n+1} in one step, as shown in the following.

We consider the polynomial ring $\mathbb{C}[y_1, \ldots, y_n, x_{n+1}]$. Making use of (1) and the substitutions $\xi_i = \tilde{p}_i/\tilde{q}_i$ as well as $\alpha_i = y_i^{\tilde{q}_i}$ for $1 \leq i \leq n$, (6) leads to the new polynomial equation

$$\begin{aligned}\widehat{f}_{n+1} := x_{n+1} \sum_{\omega \models A_{n+1}} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} y_i^{\tilde{q}_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models \overline{A_i}}} y_i^{\tilde{p}_i} \\ - \sum_{\omega \models A_{n+1} B_{n+1}} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} y_i^{\tilde{q}_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models \overline{A_i}}} y_i^{\tilde{p}_i} = 0.\end{aligned} \tag{7}$$

The next theorem derives a necessary condition for the MaxEnt probability ξ_{n+1} by exploiting

$$F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) := \mathcal{GB}_{\preceq}(\mathcal{SAT}_{\mathcal{Y}}(\langle F_2(\mathcal{KB}) \cup \{\widehat{f}_{n+1}\}\rangle)).$$

Note that F_{inf} might also be defined using the Gröbner basis $F_3(\mathcal{KB}, \preceq')$ instead of the original set of polynomials $F_2(\mathcal{KB})$, i.e., the set of polynomials $F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq)$ is equal to $\mathcal{GB}_{\preceq}(\mathcal{SAT}_{\mathcal{Y}}(\langle F_3(\mathcal{KB}, \preceq') \cup \{\widehat{f}_{n+1}\}\rangle))$.

Theorem 3. *Let $\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\}$ be a consistent knowledge base with non-trivial probabilities, $\mathcal{C}_{\text{inf}} = (B_{n+1}|A_{n+1})$ a further conditional and $\mathcal{KB} \mid\!\sim_{\mathcal{ME}} (B_{n+1}|A_{n+1})[\xi_{n+1}]$. Further, let $\preceq$ be an arbitrary elimination term ordering for $\mathcal{Y}$. Then ξ_{n+1} is a common root of*

$$F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) \cap \mathbb{C}[x_{n+1}].$$

Proof. As $\mathcal{KB}$ is consistent with non-trivial probabilities, there is a common root $(\eta_1, \ldots, \eta_n) \in \mathbb{R}^n$ of $F_2(\mathcal{KB})$ with $\eta_i \neq 0$ for $1 \leq i \leq n$ and $\mathcal{ME}(\mathcal{KB})$ exists. The probability ξ_{n+1} is given then by (6) and $(\eta_1, \ldots, \eta_n, \xi_{n+1}) \in \mathbb{R}^{n+1}$ is a common root of the polynomials $f_1, \ldots, f_n$ as well as $\widehat{f}_{n+1}$. As a consequence, it is also a common root of the ideal $\langle F_2(\mathcal{KB}) \cup \{\widehat{f}_{n+1}\}\rangle$. Again because the probabilities are non-trivial, $(\eta_1, \ldots, \eta_n, \xi_{n+1}) \in \mathbb{R}^{n+1}$ is a common root of $\mathcal{SAT}_{\mathcal{Y}}(\langle F_2(\mathcal{KB}) \cup \{\widehat{f}_{n+1}\}\rangle)$ and hence of $F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq)$. Since $F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) \cap \mathbb{C}[x_{n+1}]$ is isomorphic to the appropriate subset of $F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq)$ that does not mention any of $y_1, \ldots, y_n$, the probability ξ_{n+1} is a common root of $F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) \cap \mathbb{C}[x_{n+1}]$. □

Example 6 (Answering Queries)**.** *Since we know that the knowledge base $\mathcal{KB}_1$ is consistent from Ex. 4, so is*

$$\mathcal{KB}_3 = \{(b|a)[3/4], (c|a)[2/3]\} \subseteq \mathcal{KB}_1.$$

Given the knowledge base $\mathcal{KB}_3$, we want to infer the MaxEnt probability $\xi_3 \in (0, 1)$ of the conditional $\mathcal{C}^3_{\text{inf}} = (c|ab)$. Equation (7) leads to

$$\widehat{f}_3 = x_3\, (y_1^4\, y_2^3 + y_1^4) - y_1^4\, y_2^3.$$

Together with

$$F_2(\mathcal{KB}_3) = \{y_1^4\, (y_2^3{+}1){-}3\, (y_2^3{+}1), y_2^3\, (y_1^4{+}1){-}2\, (y_1^4{+}1),$$

the probability ξ_3 can be computed by evaluating the common root of

$$F_{\text{inf}}(\mathcal{KB}_3, \mathcal{C}^3_{\text{inf}}, \preceq_{\text{lex}}) \cap \mathbb{C}[x_3] = \{x_3 - 2/3\}$$

where $\preceq_{\text{lex}}$ is the lexicographical term ordering given by $x_3 \prec_{\text{lex}} y_2 \prec_{\text{lex}} y_1$. It obviously yields $\xi_3 = 2/3$, i.e., $\mathcal{KB}_3 \mid\!\sim_{\mathcal{ME}} (c|ab)[2/3]$.

Theorem 3 provides a symbolic representation of the inferred MaxEnt probability to which numerical methods can be applied to return the queried value or candidates thereof.

6 Generic MaxEnt Inference with Gröbner Bases

If the probabilities of the conditionals in a knowledge base $\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\}$ are not known, it is difficult to apply Gröbner bases techniques to the polynomial equation system (4). However, under some circumstances terms that may cause problems, i.e., terms with generic probabilities in its exponents, cancel out. If this happens, general statements can be made. In such cases, a coarser version of (2) resp. (4) is helpful which can be immediately tackled with Gröbner bases, even if the given probabilities in $\mathcal{KB}$ are not rational.

Let $\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\}$ be a consistent knowledge base with non-trivial probabilities, and for a further conditional $\mathcal{C}_{\text{inf}} = (B_{n+1}|A_{n+1})$ we would like to know its MaxEnt probability, i.e., the unique ξ_{n+1} such that $\mathcal{KB} \mathrel{|\!\sim}_{\mathcal{ME}} (B_{n+1}|A_{n+1})[\xi_{n+1}]$ holds. The basic idea for the coarser polynomial system is to define new variables $\beta_i = \alpha_i^{-\xi_i}$ and treat them as independent of α_i for all $1 \leq i \leq n$. We denote the set of these generic variables with $\mathcal{Y}_{\alpha\beta} = \{\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n\}$ and consider the polynomial ring $\mathbb{C}[\alpha_1, \ldots, \alpha_n, \beta_i, \ldots, \beta_n, x_1, \ldots, x_{n+1}]$. Then (2) becomes a polynomial equation system that leads to a *set of generalized polynomials* $\mathcal{G}(\mathcal{KB}) := \{g_1, \ldots, g_n\}$ *associated with* $\mathcal{KB}$:

$$\begin{aligned} g_i := (1 - x_i)\, \alpha_i \sum_{\omega \models A_i B_i} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} \alpha_j\, \beta_j \prod_{\substack{j \neq i \\ \omega \models A_j \overline{B_j}}} \beta_j \\ - x_i \sum_{\omega \models A_i \overline{B_i}} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} \alpha_j\, \beta_j \prod_{\substack{j \neq i \\ \omega \models A_j \overline{B_j}}} \beta_j = 0, \end{aligned} \tag{8}$$

for $1 \leq i \leq n$. Equation (6) leads to

$$\begin{aligned} \widehat{g}_{n+1} := x_{n+1} \sum_{\omega \models A_{n+1}} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} \alpha_i\, \beta_i \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i \overline{B_i}}} \beta_i \\ - \sum_{\omega \models A_{n+1} B_{n+1}} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} \alpha_i\, \beta_i \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i \overline{B_i}}} \beta_i = 0. \end{aligned} \tag{9}$$

Note that we also use variables x_i for the (known) probabilistic parameters ξ_i, $1 \leq i \leq n$, since we want to explore necessary relationships between the new probability ξ_{n+1} that we associate to the variable x_{n+1} and the given probabilities. This is done in the following theorem by exploiting

$$G_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) := \mathcal{GB}_{\preceq}(\mathcal{SAT}_{\mathcal{Y}_{\alpha\beta}}(\langle \mathcal{G}(\mathcal{KB}) \cup \{\widehat{g}_{n+1}\} \rangle))$$

in analogy to $F_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq)$.

Theorem 4. *Let* $\mathcal{KB} = \{(B_1|A_1)[\xi_1], \ldots, (B_n|A_n)[\xi_n]\}$ *be a consistent knowledge base with non-trivial, not necessarily rational probabilities,* $\mathcal{C}_{\text{inf}} = (B_{n+1}|A_{n+1})$ *a further conditional and* $\mathcal{KB} \mathrel{|\!\sim}_{\mathcal{ME}} (B_{n+1}|A_{n+1})[\xi_{n+1}]$. *Further, let* $\preceq$ *be an arbitrary elimination term ordering for* $\mathcal{Y}_{\alpha\beta}$. *Then* $(\xi_1, \ldots, \xi_{n+1}) \in \mathbb{R}^{n+1}$ *is a common root of*

$$G_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) \cap \mathbb{C}[x_1, \ldots, x_{n+1}].$$

Proof. As $\mathcal{KB}$ is consistent with non-trivial probabilities, the MaxEnt distribution $\mathcal{ME}(\mathcal{KB})$ exists and there is a common root $(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_n, \tilde{\beta}_1, \ldots, \tilde{\beta}_n, \xi_1, \ldots, \xi_n) \in \mathbb{R}^{3n}$ of the set of generalized polynomials $\mathcal{G}(\mathcal{KB})$ with $\tilde{\alpha}_i, \tilde{\beta}_i, \xi_i \neq 0$ for $1 \leq i \leq n$. The probability ξ_{n+1} is given then by (8) and $rt := (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_n, \tilde{\beta}_1, \ldots, \tilde{\beta}_n, \xi_1, \ldots, \xi_n, \xi_{n+1}) \in \mathbb{R}^{3n+1}$ is a common root of the polynomials in $\mathcal{G}(\mathcal{KB}) \cup \{\widehat{g}_{n+1}\}$. As a consequence, $rt \in \mathbb{R}^{3n+1}$ is also a common root of the ideal $\langle \mathcal{G}(\mathcal{KB}) \cup \{\widehat{g}_{n+1}\} \rangle$. Again because all of the occurring probabilities are non-trivial, $rt \in \mathbb{R}^{3n+1}$ is a common root of the saturation ideal $\mathcal{SAT}_{\mathcal{Y}_{\alpha\beta}}(\langle \mathcal{G}(\mathcal{KB}) \cup \{\widehat{g}_{n+1}\} \rangle)$ and hence of the Gröbner basis $G_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq)$. Since the elimination ideal $G_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) \cap \mathbb{C}[x_1, \ldots, x_{n+1}]$ is isomorphic to the appropriate subset of $G_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq)$ that does not mention any of $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$, the probability vector $(\xi_1, \ldots, \xi_{n+1}) \in \mathbb{R}^{n+1}$ is a common root of $G_{\text{inf}}(\mathcal{KB}, \mathcal{C}_{\text{inf}}, \preceq) \cap \mathbb{C}[x_1, \ldots, x_{n+1}]$. □

As Theorem 4 holds for arbitrary non-trivial probabilities $\xi_1, \ldots, \xi_n \in (0, 1)$ that lead to a consistent knowledge base, the relations between $x_1, \ldots, x_{n+1}$ hold in full generality.

In the following, we illustrate Theorem 4 by considering the MaxEnt inference rules of cautious monotonicity and cautios cut in general. Both have been proved before in (Kern-Isberner 2001) but the proofs now presented with Gröbner bases are new. Furthermore, an additional inference rule following the same ideas is presented, too. In all three cases the lexicographical term ordering $\preceq_{\text{lex}}$ is used that is based on

$$\begin{aligned} &x_{n+1} \prec_{\text{lex}} \cdots \prec_{\text{lex}} x_1 \prec_{\text{lex}} \\ &\beta_n \quad\; \prec_{\text{lex}} \cdots \prec_{\text{lex}} \beta_1 \prec_{\text{lex}} \\ &\alpha_n \quad\; \prec_{\text{lex}} \cdots \prec_{\text{lex}} \alpha_1. \end{aligned}$$

Proposition 1 (Cautious Monotonicity)**.** *Let* $\mathcal{V} = \{a, b, c\}$ *be propositional variables,* $\mathcal{KB}_4 = \{(b|a)[\xi_1], (c|a)[\xi_2]\}$ *a knowledge base,* $\mathcal{C}^4_{\text{inf}} = (c|ab)$ *a further conditional and* $\xi_1, \xi_2 \in (0, 1)$. *Then*

$$\mathcal{KB}_4 \mathrel{|\!\sim}_{\mathcal{ME}} (c|ab)[\xi_3] \textit{ iff } \xi_3 = \xi_2.$$

Proof. The generalized polynomials associated to $\mathcal{KB}_4$ according to (8) are

$$\begin{aligned} g_1 &= (1 - x_1)\, \alpha_1\, (\alpha_2\, \beta_2 + \beta_2) - x_1\, (\alpha_2\, \beta_2 + \beta_2), \\ g_2 &= (1 - x_2)\, \alpha_2\, (\alpha_1\, \beta_1 + \beta_1) - x_2\, (\alpha_1\, \beta_1 + \beta_1). \end{aligned}$$

In addition (9) leads to

$$\widehat{g}_3 = x_3\, (\alpha_1\, \beta_1\, \alpha_2\, \beta_2 + \alpha_1\, \beta_1\, \beta_2) - \alpha_1\, \beta_1\, \alpha_2\, \beta_2.$$

We find

$$G_{\text{inf}}(\mathcal{KB}_4, \mathcal{C}^4_{\text{inf}}, \preceq_{\text{lex}}) \cap \mathbb{C}[x_1, x_2, x_3] = \{x_2 - x_3\}$$

from which we obtain the stated inference rule of cautious monotonicity in full generality. □

Proposition 2 (Cautious Cut)**.** *Let* $\mathcal{V} = \{a, b, c\}$ *be propositional variables,* $\mathcal{KB}_5 = \{(c|ab)[\xi_1], (b|a)[\xi_2]\}$ *a knowledge base,* $\mathcal{C}^5_{\text{inf}} = (c|a)$ *a further conditional and* $\xi_1, \xi_2 \in (0, 1)$. *Then*

$$\mathcal{KB}_5 \mathrel{|\!\sim}_{\mathcal{ME}} (c|a)[\xi_3] \textit{ iff } \xi_3 = {}^1\!/_2\, (2\, \xi_1\, \xi_2 - \xi_2 + 1).$$

Proof. The generalized polynomials associated to $\mathcal{KB}_5$ are

$$g_1 = (1 - x_1)\,\alpha_1\,\alpha_2\,\beta_2 - x_1\,\alpha_2\,\beta_2,$$
$$g_2 = (1 - x_2)\,\alpha_2\,(\alpha_1\,\beta_1 + \beta_1) - x_2\,(1 + 1).$$

In addition (9) leads to

$$\widehat{g}_3 = x_3\,(\alpha_1\,\beta_1\,\alpha_2\,\beta_2 + \alpha_2\,\beta_2\,\beta_1 + \beta_2 + \beta_2) - (\alpha_1\,\beta_1\,\alpha_2\,\beta_2 + \beta_2).$$

From

$$G_{\mathrm{inf}}(\mathcal{KB}_5, \mathcal{C}^5_{\mathrm{inf}}, \preceq_{\mathrm{lex}}) \cap \mathbb{C}[x_1, x_2, x_3]) = \{x_1\,x_2 - {}^1\!/_2\,x_2 - x_3 + {}^1\!/_2\}$$

it follows that $\xi_3 = {}^1\!/_2\,(2\,\xi_1\,\xi_2 - \xi_2 + 1)$ must hold. □

The MaxEnt inference rule presented in the next proposition illustrates in a special case how complex interactions between several rules are dealt with in a transparent way.

Proposition 3. *Let* $\mathcal{V} = \{a, b, c\}$ *be propositional variables,* $\mathcal{KB}_6 = \{(c|a)[\xi_1], (c|b)[\xi_2], (a|c)[\xi_3], (b|c)[\xi_4]\}$ *a knowledge base,* $\mathcal{C}^6_{\mathrm{inf}} = (ab|c)$ *a further conditional and* $\xi_i \in (0, 1)$, $1 \le i \le 4$. *Then*

$$\mathcal{KB}_6 \mathrel{|\!\sim}_{\mathcal{ME}} (ab|c)[\xi_5] \text{ iff } \xi_5 = \xi_3\,\xi_4.$$

Proof. The generalized polynomials associated to $\mathcal{KB}_6$ are

$$g_1 = (1 - x_1)\,\alpha_1\,(\alpha_2\,\beta_2\,\alpha_3\,\beta_3\,\alpha_4\,\beta_4 + \alpha_3\,\beta_3\,\beta_4) - x_1\,(\beta_2 + 1),$$
$$g_2 = (1 - x_2)\,\alpha_2\,(\alpha_1\,\beta_1\,\alpha_3\,\beta_3\,\alpha_4\,\beta_4 + \alpha_1\,\beta_1\,\beta_4) - x_2\,(\beta_1 + 1),$$
$$g_3 = (1 - x_3)\,\alpha_3\,(\alpha_1\,\beta_1\,\alpha_2\,\beta_2\,\alpha_4\,\beta_4 + \alpha_1\,\beta_1\,\beta_4) - x_3\,(\alpha_2\,\beta_2\,\alpha_4\,\beta_4 + \beta_4),$$
$$g_4 = (1 - x_4)\,\alpha_4\,(\alpha_1\,\beta_1\,\alpha_2\,\beta_2\,\alpha_3\,\beta_3 + \alpha_2\,\beta_2\,\beta_3) - x_4\,(\alpha_1\,\beta_1\,\alpha_3\,\beta_3 + \beta_3).$$

In addition (9) leads to

$$\widehat{g}_5 = x_5\,(\alpha_1\,\beta_1\,\alpha_2\,\beta_2\,\alpha_3\,\beta_3\,\alpha_4\,\beta_4 + \alpha_1\,\beta_1\,\alpha_3\,\beta_3\,\beta_4 + \alpha_2\,\beta_2\,\alpha_4\,\beta_4\,\beta_3 + \beta_3\,\beta_4) - \alpha_1\,\beta_1\,\alpha_2\,\beta_2\,\alpha_3\,\beta_3\,\alpha_4\,\beta_4.$$

We obtain

$$G_{\mathrm{inf}}(\mathcal{KB}_6, \mathcal{C}^6_{\mathrm{inf}}, \preceq_{\mathrm{lex}}) \cap \mathbb{C}[x_1, \ldots, x_5]) = \{x_5 - x_3\,x_4\}.$$

Hence, $\xi_5 = \xi_3\,\xi_4$ proves to be a necessary (and also sufficient) condition. □

7 Conclusion

We presented a novel application of Gröbner bases theory to probabilistic knowledge representation by showing how the algebraic foundations of the principle of maximum entropy fit well to the idea of Gröbner bases. Indeed, the combination of both methodologies offers numerous new ways of computing MaxEnt probabilities, as first results concerning the consistency of knowledge bases and the problem of answering queries prove. From a coarser, even more algebraic point of view general inference rules can be obtained.

We will continue this novel line of research by strengthening the connection between MaxEnt and Gröbner bases with the aim of developing new efficient methods for MaxEnt reasoning that are not based on probabilistic networks, like those in (Rödder and Meyer 1996). From a theoretical point of view, an interesting question which we pursue in our ongoing work is whether some of the necessary conditions for MaxEnt probabilities that we presented in this paper can be modified to necessary and sufficient conditions. This would allow us to comupte precise numerical probabilities to answer MaxEnt inference queries.

References

Buchberger, B. 2006. An algorithm for finding the basis elements in the residue class ring modulo a zero dimensional polynomial ideal. *Journal of Symbolic Computation* 41(3-4):475–511.

Cox, D. A.; Little, J.; and O'Shea, D. 2007. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, 3/e (Undergraduate Texts in Mathematics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Dukkipati, A. 2009. Embedding maximum entropy models in algebraic varieties by gröbner bases methods. In *ISIT*, 1904–1908.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: MIT Press.

Jaynes, E. 1983. *Papers on Probability, Statistics and Statistical Physics*. Dordrecht, Holland: D. Reidel Publishing Company.

Kern-Isberner, G., and Thimm, M. 2012. A ranking semantics for first-order conditionals. In De Raedt, L.; Bessiere, C.; Dubois, D.; Doherty, P.; Frasconi, P.; Heintz, F.; and Lucas, P., eds., *Proceedings 20th European Conference on Artificial Intelligence, ECAI-2012*, number 242 in Frontiers in Artificial Intelligence and Applications, 456–461. IOS Press.

Kern-Isberner, G. 1998. A note on conditional logics and entropy. *International Journal of Approximate Reasoning* 19:231–246.

Kern-Isberner, G. 2001. *Conditionals in nonmonotonic reasoning and belief revision*. Springer, Lecture Notes in Artificial Intelligence LNAI 2087.

Paris, J., and Vencovská, A. 1990. A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning* 14:183–223.

Paris, J. 1994. *The uncertain reasoner's companion – A mathematical perspective*. Cambridge University Press.

Paris, J. 1999. Common sense and maximum entropy. *Synthese* 117:75–93.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Ca.: Morgan Kaufmann.

Rödder, W., and Meyer, C.-H. 1996. Coherent Knowledge Processing at Maximum Entropy by SPIRIT. In *Proceedings UAI 1996*, 470–476.

Rödder, W.; Reucher, E.; and Kulmann, F. 2006. Features of the expert-system-shell SPIRIT. *Logic Journal of the IGPL* 14(3):483–500.