Reasoning with BKBs - Algorithms and Complexity

Tzachi Rosen and Solomon Eyal Shimony*
Department of Computer Science - Ben Gurion University
P. O. Box 653, Beer-Sheva 84105, ISRAEL
e-mail: {tzachi, shimony}@cs.bgu.ac.il

Eugene Santos Jr.†
Department of Computer Science and Engineering
University of Connecticut - Storrs, CT 06269
eugene@cse.uconn.edu

October 31, 2002

Abstract

Bayesian Knowledge Bases (BKB) are a rule-based probabilistic model that extends the well-known Bayes Networks (BN), by naturally allowing for context-specific independence and for cycles in the directed graph. We present a semantics for BKBs that facilitate handling of marginal probabilities, as well as finding most probable explanations.

Complexity of reasoning with BKBs is NP hard, as for Bayes networks, but in addition, deciding consistency is also NP-hard. In special cases that problem does not occur. Computation of marginal probabilities in BKBs is another hard problem, hence approximation algorithms are necessary - stochastic sampling being a commonly used scheme. Good performance requires importance sampling, a method that works for BKBs with cycles is developed.

1 Introduction

Managing uncertainty in complex domains continues to remain a difficult task, especially during knowledge acquisition and verification and validation. There are a wide variety of approaches from fuzzy logics to probabilistic networks [11, 26, 16, 12, 25, 5, 23, 18, 8, 1]. Bayesian Knowledge Bases (BKB - [15]) are a rule-based probabilistic model that is a generalization of the well-known Bayes Networks (BN - [12]). BKBs extend the BN model in two ways: by naturally allowing for context-specific independence through "if-then" style constructs (also called rules), and by permitting cycles in the directed graph. These generalizations to Bayes networks are necessary when we need to model populations (or sample spaces) where the causal mechanism varies across the population (see, e.g. [24] Sections 3.4.2, 7.3 p 212).

*Partially supported by the Israeli Ministry of Science, and by the Paul Ivanier Center for Robotics at Ben-Gurion University. Part of the research done while S. E. Shimony was on sabbatical leave at University of Connecticut at Storrs, 1999-2000.

†Supported by AFOSR Grant Nos. F49620-99-1-0059 and 940006
Several models in the literature permit such context-sensitive independence, using rules [13, 27, 17], and trees [2], but none of them allow for cycles between variables.

In BKBs, the different causal structures are achieved by using a separate node in the graph to denote each different instantiation to the same random variable. For example, Figure 1 shows a simple Bayes network represented as a BKB. The variable \( X \) in the figure has two possible instantiations, thus its counterpart in the BKB is two separate nodes (called I-nodes), one for each possible instantiation in \( \{T, F\} \). Each rule in a BKB (also called an S-node) corresponds to a Bayes network CPT entry - although in general one rule can stand for a large number of CPT entries with the same conditional probability. Thus, the size of the BKB is at most a constant factor larger than the BN it represents, but may actually be considerably smaller in cases where a single rule represents numerous CPT entries [19, 21]. For example, to represent an \( n \)-parent OR node, only \( 2n \) rules are needed, rather than \( 2^n \) CPT entries in an explicitly represented BN.

In a BKB, the topology of a sub-graph representing a particular part of the model reflects that specific causal structure, including context-specific independence. If such a structure were mapped back into a Bayes network, much of the (local) causal structure would be lost - which is undesirable. If the direction of some causal mechanism also depends on specific attributes of members in the population, it is no longer possible to directly map the causal structure into a Bayes network - this would result in cycles, which the model cannot handle. Collapsing all variables in the cycle into a single variable, or using an undirected model, such as Markov networks, is also undesirable - due to space complexity in the former case, and loss of causal structure description in both schemes.

Originally presented in prior work [15], a method for finding most-probable explanations, or states of affairs given evidence, was shown. However, methods for handling marginal probabilities, and marginal probabilities given evidence (also called "belief updating") were lacking - and adding them in was non-
trivial given the semantics as defined. We provide a slightly different semantics for BKBs that facilitate handling of marginal probabilities in the model. The addition of belief updating provides capabilities for predictive reasoning, as well as alternative methods for abductive reasoning - to complement the existing scheme [15] of finding most probable inferences (explanations).

The complexity of reasoning with BKBs is NP-hard, as for Bayes networks, but we show that in addition, deciding consistency is also NP-hard. We single out special cases where that problem does not occur. Since computation of marginal probabilities in BKBs is a hard problem, approximation algorithms are necessary - stochastic sampling being a commonly used scheme. Good performance requires importance sampling, but the scheme used to perform importance sampling for BNs is not directly applicable to BKBs - a method that works for BKBs is developed here.

The rest of the paper is organized as follows. Section 2 reviews the definitions of BKBs from prior work. Section 3 then presents the new semantics. Special cases of BKBs and their properties are examined in Section 4, followed by reasoning by stochastic sampling in Section 5.

2 Background

A Bayesian knowledge-base (abbrev. BKB) represents objects/world states and the relationships between them using a directed graph. The graph consists of nodes which denote various random variable instantiations while the edges represent conditional (in)dependencies between them. Definition of BKBs is in terms of graph-theoretic entities (repeated from [15]), which are equivalent to a set of probabilistic rules. We use the terms synonymously throughout, pointing out the correspondence as we proceed.

**Definition 1** A correlation-graph $G = (I \cup S, E)$ is a directed graph such that $I$ and $S$ are disjoint, and $E \subseteq (I \times S) \cup (S \times I)$. Furthermore, for all $a \in S$, $(a, b)$ and $(a, b')$ are in $E$ if and only if $b = b'$. $(I \cup S)$ are the nodes of $G$ and $E$ are the edges of $G$. A node in $I$ is called an instantiation-node (abbrev. I-node) and a node in $S$ is called a support-node (abbrev. S-node).

I-nodes represent the various instantiations of random variables (abbrev. r.v.s), that is, an assignment of a value to a random variable. S-nodes, on the other hand, explicitly embody the relationships (conditional (in)dependence) between the I-nodes. See Figure 2 - filled circles represent S-nodes, ovals represent I-nodes.

**Notation 1** Let $a$ be any node in $I \cup S$. $\text{PRED}_G(a) = \{b \mid (b, a) \in E\}$ are the immediate predecessors of $a$ in graph $G$. $\text{DESC}_G(a) = \{b \mid (a, b) \in E\}$ are the immediate descendants of $a$ in graph $G$.

Let $\pi$ be a partition on $I$. Each cell in $\pi$ will denote the set of I-nodes (instantiations) which belong to a single r.v. and are mutually exclusive instantiations. For a variable $X$, we denote the set of all its possible instantiations by $D(X)$ - this set corresponds to the partition cell for variable $X$. In BKBs, we can represent random variables with discrete but multiple instantiations. In Figure 2, one cell in $\pi$ would be $\{U = 0, U = 1\}$ which are two instantiations for the r.v. $U$. A set of I-nodes that contains at most one I-node in each partition (i.e., for each rv) is called a state (w.r.t. $\pi$). A state that contains exactly one I-node for each rv in a set of variables $X$ is complete for $X$ (resp., for a partition $\pi$). The set of variables assigned in a correlation graph segment, rule, or set of rules, or I-node, are
called the span of that object (denoted \( \text{span}(\cdot) \)). For example, the sets of I-nodes \( \{U = 1, Z = 1\} \) and \( \{U = 1, Z = 1, Y = 1, T = 0, X = 0\} \) are states, the latter being complete for the variable set \( \{T, U, X, Y, Z\} \).

**Definition 2** \( G \) is said to respect \( \pi \) if

- for any S-node \( b \in S \), the predecessor I-nodes of \( b \), \( \text{PRED}_G(b) \), assigns at most one instantiation to each r.v., and

- for any two distinct S-nodes \( b_1 \) and \( b_2 \) in \( S \) such that \( \text{DESC}_G(b_1) = \text{DESC}_G(b_2) \), there exists an I-node in \( \text{PRED}_G(b_1) \) whose r.v. instantiation contradicts an I-node in \( \text{PRED}_G(b_2) \). Furthermore, the \( b_1 \) and \( b_2 \) are said to be mutually exclusive.

Intuitively, an S-node represents a direct conditional dependency between the single immediate I-node descendant of the S-node (also called the consequent) and the immediate I-node predecessors (also called the antecedent) (See Figure 2), and corresponds to a conditioning case, or a conditional-probability table (CPT) entry, in a Bayes network. For example, the value attached to the S-node \( R3 \) in the figure represents the conditional probability \( P(Y = 1|X = 0, Z = 0) = 0.6 \). Priors are denoted by S-nodes without inputs as shown in Figure 2, e.g., the S-node denoted \( R4 \). We call the subgraph consisting of an S-node \( s \), its incident edges, and its immediate neighbors, together with the attached conditional probability, a conditional probability rule (CPR). A conditional probability rule is written out as an implication \( \text{antecedent} \xrightarrow{p} \text{consequent} \), where \( 1 \geq p \geq 0 \) is the conditional probability (weight) of the
rule. A set of rules is said to be mutually exclusive (w.r.t. a partition \( \pi \)) if their correlation graph respects \( \pi \). For a rule \( R \), we denote its antecedent by \( \text{ant}(R) \), and its consequent by \( \text{cons}(R) \). We use the graph and rule notation interchangeably throughout, as convenient.

The conditions in Definition 2 guarantee that conditional (in)dependencies are meaningful (see [15]). The first condition prevents conditionals of the form \( P(A = a | \ldots, B = b, \ldots, B = u, \ldots) \) where \( b \neq u \). The second condition does not allow the model to specify conditioning events that overlap in probability space for the same I-node (e.g., the events \( \{X = 0\}, \{Y = 0\} \)) have an "overlap", or conjunction, \( \{X = 0, Y = 0\} \)), in turn preventing local inconsistency [15]. Enforcing mutual exclusivity in BKBs is a potential limitation in this representation. However, when knowledge engineering BKBs for various domains, this can be overcome by increasing the level of detail or specificity of the "rules" involved.

To complete our knowledge-graph, we define a function \( w \) from \( S \) to \([0, 1]\). This serves as the actual conditional probability value associated with each S-node. Figure 2 is an example of a fragment of a BKB.

**Definition 3** A Bayesian knowledge-base \( K \) is a 3-tuple \((G, w, \pi)\) where \( G = (I \cup S, E) \) is a correlation-graph, \( w \) is a function from \( S \) to \([0, 1]\), \( \pi \) is a partition on \( I \), and \( G \) respects \( \pi \). Furthermore, for each \( a \in S \), \( w(a) \) is the weight of \( a \).

The main differences between Bayesian knowledge-bases and Bayesian networks are: first, for BKBs, it is possible to have an I-node with no inputs. This situation is not precluded in Definition 3 above, in order to handle incompleteness in our knowledge, but such incomplete I-nodes are essentially ignored during reasoning. Other forms of incompleteness which arise from missing interactions between r.v.s can also occur and are discussed in the next section. Bayesian networks on the other hand require a complete specification of the probability distributions of its r.v.s. Another primary distinction between r.v.s and Bayesian networks is the ability of r.v.s to handle various forms of cyclic knowledge. Bayesian networks must be acyclic in its r.v. relations. In Figure 2, we have a directed cycle in the BKB fragment.

Let \( r = (I' \cup S', E') \) be some subgraph of our correlation-graph \( G = (I \cup S, E) \) where \( I' \subseteq I \), \( S' \subseteq S \), and \( E' \subseteq E \). Then, \( r \) has a weight \( \omega(r) \) defined as follows: \( \omega(r) = \prod_{s \in S'} w(s) \).

**Definition 4** An I-node \( a \in I' \) is said to be well-supported in \( r \) if there exists an edge \((b, a) \) in \( E' \). Furthermore, \( r \) is said to be well-supported if for all I-nodes \( a \) in \( I' \), \( a \) is well-supported.

Each I-node must have an incoming S-node in \( r \).

**Definition 5** An S-node \( b \in S' \) is said to be well-founded in \( r \) if for all \((a, b) \) in \( E \), \((a, b) \) in \( E' \). Furthermore, \( r \) is said to be well-founded if for all S-nodes \( b \) in \( S' \), \( b \) is well-founded.

If an S-node \( b \) is present in \( r \), then all incoming I-nodes (conditions) to \( b \) in \( G \) must also be present in \( r \).

---

1When the actual probability is irrelevant to the discussion (e.g., proofs of computational complexity), we sometimes, omit the probability and write antecedent \( \rightarrow \) consequent.

2Equivalently, we can define a BKB by a mutually exclusive set of rules \( \mathcal{R} \) (which define the correlation graph and weights - the weight of a rule is the weight of its S-node) over a set of variables \( \mathcal{X} \) (which defines the partition) - these notions are used interchangeably.
Definition 6 An S-node $b \in S'$ is said to be well-defined in $r$ if there exists an edge $(b, a) \in E'$. Furthermore, $r$ is said to be well-defined if for all S-nodes $b$ in $S'$, $b$ is well-defined.

Each S-node in $r$ must support some I-node in $r$.

Definition 7 $r$ is said to be an inference over $K$ if $r$ is well-supported, well-founded, well-defined, acyclic, and the set of I-nodes of $r$ are a state w.r.t. $\pi$. Furthermore, $r$ is said to be a complete inference over $K$ if $r$'s I-nodes are a complete state for $\pi$. $r$ is said to be maximal in $K$ if no proper superset of $r$ is an inference over $K$.

Note that each I-node must have a single S-node input and that all the inputs of each S-node in the original correlation graph must also be present in the inference. Also, only one S-node from the set of S-node inputs to an I-node can be present. Because of mutual exclusion in the S-nodes, multiple S-node inclusion in the inference would result in also including contradictory I-nodes. Finally, we can view a correlation graph as an AND-XOR graph and build inferences from valid truth assignment subgraphs. For example Figure 2, contains several inferences, among them the subgraphs delimited by the rectangles denoted $I_i$, $J_i$, $K$ respectively. Inference $K$ is complete, it assigns values to all variables.

Definition 8 A node $v$ is said to be grounded in a correlation graph $G$ if there exists an inference $r \subseteq G$ such that $v \in r$. Likewise, a CPR is grounded in a correlation graph $G$ if its S-node is grounded in $G$.

3 Semantics of BKBs

The issue of which independence statements ("independence semantics") hold in a BKB has been partially covered in [15]. Although work still remains to be done on this topic, this paper concentrates on the probability distribution(s) sanctioned by a BKB. The existence of such a distribution is assured by requiring normalization - a scheme of local normalization was defined in [15]; but we require a somewhat stronger global normalization in order to get a single, complete, probability distribution, that makes sense intuitively. A normalized set of CPRs is one for which the extenders of all its inferences are normalized. (Henceforth, we will assume for conciseness, that we have a given BKB, and unless otherwise specified, all CPRs are taken from its correlation graphs, all variables are from the set of the BKB variables $\mathcal{X}$, etc.)

Definition 9 Let $R$ be a CPR. $R$ is called an extender of inference $I$ if $R \notin I$ and $I \cup \{R\}$ is an inference.

A set of CPRs $R$ is called complementary w.r.t. an inference $I$ and a variable $X$, if each of them extends $I$, their consequent variable is $X$, but no two of them has the same I-node as a consequent. $R$ is called complete (for $X$) if the consequences include all possible instantiations of $X$.

For example, the CPRs $R_2$ and $R_3$ in Figure 2 are complementary w.r.t. the inference $I$ and the variable $Y$. Note that, in a mutually exclusive set of CPRs, an inference $I$ has, w.r.t. any variable $X$, a unique maximal complementary set of CPRs, which we denote by $\text{com}(I, X)$. We also introduce the function $\text{ext}(I)$ to denote the set of all immediate extension inferences of $I$, as well as $\text{ext}^+$ and $\text{ext}^*$ to denote the transitive closure, and reflexive transitive closure of $\text{ext}$, respectively.
Definition 10 Let \( C \) be a complementary set of CPRs w.r.t. an inference \( I \) and a variable \( X \), and \( W(C) = \sum_{R \in C} w(R) \). 
\( C \) is called normalized w.r.t. \( I \) and \( X \) if 
\[ W(C) \leq 1, \text{ with } W(C) = 1 \text{ when } C \text{ is a complete complementary set w.r.t. } I \text{ and } X. \] A set of CPRs (or correlation graph) is called normalized if for each inference \( I \) in the graph, all complementary sets of rules are normalized.

This definition is global, in the sense that directly checking whether the definition holds requires looking at an exponential number of inferences. In the following section, we discuss cases where a polynomial number of local tests is sufficient.

Definition 11 The state of an inference \( I \) (denoted \( st(I) \)) is the set of I-nodes in its correlation graph. \( I \) is called relevant to a state \( S \) if \( st(I) \subseteq S \).

An inference is the maximal relevant inference (MRI) w.r.t. a state \( S \) if it is the (setwise) largest inference relevant to \( S \).

As an example, the state of the inference \( I \) in Figure 2 is \( \{Z = 1, U = 1\} \), and that of \( K \) is \( \{X = 0, Y = 1, Z = 0, T = 0, U = 0\} \).

\( I \) is relevant, for instance, to the states \( \{X = 0, Z = 1, U = 1\} \) and \( \{X = 0, Y = 1, Z = 1, T = 0, U = 1\} \), and \( K \) is an MRI to the complete state \( \{X = 0, Y = 1, Z = 0, T = 0, U = 0, V = 0\} \).

Proposition 1 In a mutually exclusive set of CPRs (or correlation graph), a state \( S \) has a unique maximal relevant inference (denoted \( \text{MRI}(S) \)).

Definition 12 The composite state of an inference, denoted \( C(I) \), is the set of complete states to which \( I \) is relevant. The dominated composite state of an inference, denoted \( C_D(I) \), is the set of complete states for which \( I \) is the maximal relevant inference.

Definition 13 The dominated weight of an inference \( I \) is:
\[ w_D(I) = w(I) \prod_{X \in \mathcal{X}} \left[ 1 - \sum_{R \in \text{mes}(I, X)} w(R) \right]. \]

where \( \mathcal{X} \) is the set of variables not assigned in \( I \).

We denote the set of all complete states for a set of variables \( \mathcal{X} \) by \( C_X \). The probability of a complete state \( S \in \text{cal}C_X \) is defined based on the dominated weight of the most relevant inference to \( S \) as follows:

Definition 14 Let \( K \) be a normalized BKB over variables \( X \), and \( f \) a function from \( C_X \) into \([0, 1]\). Function \( f \) is said to be consistent with \( K \) (denoted \( K \models f \)) if for each inference \( I \) in the correlation graph of \( K \),
\[ \sum_{S \in C_D(I)} f(S) = w_D(I). \]
f is called the default distribution of K if K ⪰ f and for each inference I and a pair of complete states S, S’ ∈ C_I(T), f(S) = f(S’).

If the BKB is incomplete (normalization holds with sum of rules being less than 1), there may be more than one consistent distribution - the default distribution is method for spreading out the remaining probability mass uniformly. In order to be consistent with common notation, we denote the (discrete) default distribution over K by P_K, omitting the subscript K when unambiguous, and then P(S) denotes the probability of a state S according to the default distribution over K.

**Theorem 1** Let K be a normalized BKB over X, and f a distribution consistent with K. Then f is a joint probability distribution over C_X. (In particular, P_K is a (discrete) probability distribution.)

**Proof:** (outline) - see appendix for details.

1. The set of dominated composite states of all the inferences is a partition of C_X.

2. Every inference with non-zero dominated weight has a non-empty dominated composite state.

3. The weight of an inference I is the sum of the dominated weights of all inferences J such that I ⊆ J.

4. Since the latter also holds for the empty inference, which has a weight of 1 by definition, we can show that the 0 ≤ f(S) ≤ 1 and that its sum over all states in C_X is 1.

□

The probability of an inference I is naturally defined as the probability of the event st(I), which is equal to the sum of probability of all complete states consistent with I (i.e., the complete states to which I is relevant). The following corollary follows immediately from the proof of theorem 1 (see lemma 3 in appendix):

**Corollary 1** Let K be a normalized BKB over X, f a distribution consistent with K, and I an inference in K. Then f(st(I)) = w(I).

# 4 BKB Characteristics and Complexity Results

This section introduces a natural classification of BKBs. One characteristic is whether the correlation graph contains cycles. Another has to do with completeness of CPR consequents and antecedents. Some complexity results are discussed.

**Definition 15** A variable-cycle is a directed path in the correlation graph that contains two or more I-nodes (not necessarily distinct) that correspond to the same rv.

For example, one variable-cycle in Fig. 3 is \((Y = 0, R9, Z = 0, R12, T = 0, R11, Y = 1)\), and another is \((Y = 0, R9, Z = 0, R12, T = 0, R10, Y = 0)\) - where the latter is also a cycle. Obviously, a cycle in any correlation graph implies a variable-cycle.
Definition 16 A set of CPRs \( R \) is called consequent-variant (C-variant) if they have the same antecedent, and the same consequent variable \( X \). Additionally, \( R \) is called complete (for \( X \)) if its consequent I-nodes cover all the domain of \( X \). The correlation graph is consequent-complete if every maximal (w.r.t. set inclusion) C-variant set is also complete.

In Fig. 3, \( \{R10, R11\} \) is a C-variant set, which is complete (for \( Y \), as there are only two possible instantiations for \( Y \). Thus, consequent-completeness (see Fig. 3) means that if there is a rule that can deduce an I-node (that assigns a value to variable \( X \)) from some antecedent state, then all other values of \( X \) may be deduced from the same antecedents (through other rules, possibly with different weights).

Definition 17 A set of CPRs \( R \) is called antecedent-variant (A-variant) if they have the same consequent I-node. Additionally, \( R \) is called a cover of its antecedent-variables \( V_A \) if all possible states for \( V_A \) are consistent with the antecedent of some rule in \( R \). If all possible states for \( V_A \) are also equal to the antecedent of some rule in \( R \), then \( R \) is antecedent-complete for \( V_A \). A correlation graph has antecedent-cover (resp. completeness) if every maximal (w.r.t. set-inclusion) antecedent-variant set of CPRs is a cover (resp. antecedent-complete).

Thus, antecedent-cover means that an I-node can be deduced by any possible state of its ancestor variables. Antecedent-completeness means that all these states are fully enumerated in the rules.

Proposition 2 A Bayes network corresponds naturally to a BKB that is acyclic, consequent-complete, and antecedent-complete.
The construction of a BKB from a Bayes network is straightforward. For each \((\text{variable}, \text{value})\) pair construct an I-node \(q\). For each conditional probability table (CPT) entry construct a rule, with antecedent being the I-nodes corresponding to the state of predecessors of \text{variable}, and consequent being the I-node \(q\). The weight of the rule is the CPT entry. By construction, and the fact that Bayes networks are acyclic, the correlation graph is variable-acyclic. Additionally, since the CPTs refer to all possible states of antecedents and consequents, the correlation graph is both consequent-complete and antecedent-complete.

In a BKB, rules that are not grounded are redundant, that is, they are not in any inference, and should be removed from the model. Unfortunately, checking whether a rule is grounded is hard:

**Theorem 2** Deciding groundedness of a rule \(R\) in a correlation graph is NP-complete. Groundedness remains NP-hard in the special case where the BKB is consequent-complete and has antecedent-cover.

**Proof:** [The problem is clearly in NP, since a certificate (presumably a subgraph - an inference that should contain \(R\)) can be checked in polynomial time.] Groundedness is shown to be NP-hard by reduction from 3-SAT, by constructing 2 rules with no antecedents for every logical variable (one for each possible state), for each clause \(c \equiv l_1 \lor l_2 \lor l_3\) rules corresponding to each literal that derive an I-node \(c\) for the clause: \(l_1 \rightarrow c\), \(l_2 \rightarrow c\), \(l_3 \rightarrow c\) (added negated terms needed for mutual exclusion); and a rule \(R\) that requires all clauses to hold to derive the single "sentence" I-node. Obviously, \(R\) is grounded in the resulting correlation graph just when the 3-SAT sentence is satisfiable. NP-completeness for the above special case follows from a somewhat more complicated reduction from 3-SAT - see appendix for the extended proof.

Since in a BKB constructed from a Bayes networks all rules are grounded by construction (the problem never arises in a Bayes network), the above result is somewhat negative - although reasoning in Bayes networks is hard, at least checking for consistency is immediate, and we seem to lose that when moving into the realm of BKBs. Thus, it is of much interest whether there exists a subset of BKBs that is still significantly more general than BNs, but where deciding consistency is tractable.

**Theorem 3** If a BKB has consequent-completeness and antecedent-completeness, then checking whether all rules are grounded can be done in polynomial time.

**Proof:** (outline) To test if an I-node \(q = (V, \text{value})\) is grounded, i.e. appears in some inference, do:

1. Convert all rules to their variable-only form by ignoring, for each assignment, the value assigned the variables in the rules.

2. Treating each variable as a literal (we now have a positive Horn theory \(H\)) determine if \(V\) (and thus \(q\)) is valid in \(H\) by using a polynomial-time algorithm, such as [7].

To check whether a rule \(R\) is grounded, check \(q = (V, \text{value})\), and remove all other rules that have variable \(V\) in the consequent. Call the above procedure to determine whether \(q\) is grounded in the resulting correlation graph - \(R\) is grounded just when that occurs. The proof is completed by showing that the Horn theory version of the problem is the same as the original problem, given consequent-completeness and antecedent-completeness.
Unfortunately, the above result, while theoretically interesting, is less than useful in practice. First, requiring antecedent-completeness (rather than the less restrictive antecedent-cover) precludes context-specific independence. Worse, another consequence of antecedent-completeness is that every cycle must contain ungrounded rules and $I$-nodes, and removing these objects leaves us with an acyclic graph - i.e., essentially a Bayes network. Thus, although the original BKB may have cycles, and thus may be more general than a Bayes network, the usable part of the BKB is a Bayes network, whenever we have antecedent and consequent completeness.

Consistency in a BKB requires groundedness and normalization - both are global properties of the BKB. While the former can be tested efficiently when consequent-completeness and antecedent-completeness hold, we can get a local (and thus polynomial time) test for normalization in the following special case:

**Theorem 4** Let $K$ be a consequent-complete BKB. If each maximal $C$-variant set of CPRs

$\mathcal{R}$ is locally normalized\(^3\), then $K$ is normalized. If, in addition, all nodes in the correlation-graph of $K$ are grounded, the above sufficient condition is also necessary for normalization.

**Proof:** (outline) To show that the condition is sufficient, note that consequent-completeness implies the following property. Let $I$ be an inference, and $\mathcal{R}$ a set of rules consistent with $I$ and with consequent variable $X$. If all rules are grounded, then $\mathcal{R}$ is a maximal consequent-variant set of rules just when it is equal to the maximal complementary set $mcs(I, X)$, and the above test for normalization reduces to the definition of normalization. Otherwise, $mcs(I, X) \subseteq \mathcal{R}$ and thus the sum of weights for $mcs(I, X)$ can only be smaller than for $\mathcal{R}$.

In previous sections, we introduced rule weights, and their relationship to conditional probabilities. However, the semantics of a weight of a rule, in general, is the conditional probability only in the context of the inferences in which the variable in the consequent participates. In order for a rule weight to have the “traditional” Bayes-network semantics - that being the probability of the consequent given the antecedent marginalized over all the rest of the variables, several further assumptions are necessary. Clearly, weights for ungrounded rules are meaningless. However, even for grounded rules, complications are possible. That is because if a certain context (maximal inference) does not assign a value to a variable, its probability distribution is determined by the default mechanism (i.e. uniform in that context) - even if a rule exists that assigns it a different conditional probability in another context.

For example, if we have binary-valued variables $X, Y, Z$ we may have a rule $R : x \rightarrow y$ with $w(R) = 0.9$ that is grounded in the context $z$ with the rules $z \rightarrow x$ and $\rightarrow z$. It is still possible that there are no rules for either $x$ or $y$ in the context $\rightarrow z'$, and thus under the context $z'$ we get uniform distribution over the four states $\{x, x'\} \times \{y, y'\}$, which would entail that $P(y|x) \neq 0.9$ unless $P(z') = 0$.

We thus need to assume that certain completeness conditions hold locally in the BKB, as follows.

**Definition 18** Let $R : Y_i = y_1 \land \ldots Y_k = y_k \rightarrow X = x_j$ be a rule in a BKB $K$. $K$ is called (locally) complete w.r.t. $R$ if for every complete state $S$ with non-zero probability such that $\text{ant}(R) \cup \{X = x_j\} \subseteq S$, $x_j \in D(X)$ there is an inference $I \in K$ relevant to $S$ that contains the rule ant($R$) $\rightarrow X = x_j$.

\(^3\)That is, their sum of weights is 1 when the set is complete, and otherwise $\leq 1$. 

11
Theorem 5 Let $K$ be a normalized BKB, and $R : Y_1 = y_1 \land \ldots \land Y_k = y_k \to X = x_j$ a rule such that $P(Y_1 = y_1 \land \ldots \land Y_k = y_k) > 0$. If $K$ is locally complete w.r.t. $R$, then $P(X = x_j | Y_1 = y_1 \land \ldots \land Y_k = y_k) = w(R)$.

Proof: Let $I_{X_j}$ by the set of all inferences that contain ant($R$) $\to X = x_j$, such that $X = x_j$ is the only sink node (i.e., node with zero out-degree), for some $X_j \in D(X)$, and $I_Y$ the set of all inferences in $I_{X_j}$ with the rule deducing $X = x_j$ removed. (The latter are all inferences because $X = x_j$ is a sink node).

The inferences in $I_{X_j}$ are disjoint by construction, and cover all the states where $Y_1 = y_1 \land \ldots \land Y_k = y_k \land X = x_j$ holds, due to the conditions of the theorem, and thus by corollary 1:

$$P(Y_1 = y_1 \land \ldots \land Y_k = y_k \land X = x_j) = \sum_{I \in I_{X_j}} w(I)$$

Using the same argument for the inferences in $I_Y$, we have:

$$P(Y_1 = y_1 \land \ldots \land Y_k = y_k) = \sum_{I \in I_Y} w(I)$$

Thus, whenever the conditional probability is defined, we have:

$$P(Y_1 = y_1 \land \ldots \land Y_k = y_k | X = x_j) \equiv \frac{P(Y_1 = y_1 \land \ldots \land Y_k = y_k \land X = x_j)}{P(Y_1 = y_1 \land \ldots \land Y_k = y_k)} = \frac{\sum_{I \in I_{X_j}} w(I)}{\sum_{I \in I_Y} w(I)} = w(R)$$

The last equality holds, since each inference in $I_{X_j}$ extends exactly one inference in $I_Y$, and the extender is always $R$, and we can re-write the numerator:

$$\sum_{I \in I_{X_j}} w(I) = \sum_{I \in I_Y} w(R)w(I) = w(R) \sum_{I \in I_Y} w(I)$$

Intuitively, this means that if all inferences that have non-zero dominated weight and are consistent with $\text{ant}(R)$ include a rule of the form $\text{ant}(R) \to X = x_j$, then $w(R)$ has its intended meaning of conditional probability. Note that ungrounded rules are automatically excluded here, as they do not appear in any inference in $K$.

Theorem 6 Let $K$ be a normalized, consequent-complete, grounded BKB, such that all its maximal inferences are complete. Then $K$ is locally complete w.r.t. every rule $R \in K$, and thus $P(\text{cons}(R)|\text{ant}(R)) = w(R)$ whenever $P(\text{ant}(R)) > 0$.

Proof: The theorem’s conditions imply (through the definition of dominated weight) that only maximal inferences have non-zero dominated weight, and thus states have non-zero probability only if their MRI is a complete inference. Consequent completeness also implies that if there is a rule $R : Y_1 = y_1 \land \ldots \land Y_k = y_k \to X = x_j$, then all the complementary rules of the form $R' : Y_1 = y_1 \land \ldots \land Y_k = y_k \to X = x_l$ are in $K$, and appear in some inference. Let $S$ be a complete state with $P(S) > 0$ such that $\{Y_1 = y_1, \ldots, Y_k = y_k, X = x_j\} \subseteq S$, and let $I = \text{MRI}(S)$. Let $X = x_j$ appear as a consequent of rule $R^0 \in I$. Then the antecedent of $R^0$ overlaps with $Y_1 = y_1 \land \ldots \land Y_k = y_k$ at least at the state $S$, which violates mutual exclusion (with some $R'$ above) unless $R^0 = Y_1 = y_1 \land \ldots \land Y_k = y_k \to X = x_l$. Thus, $K$ is locally complete w.r.t. $R$, and the theorem follows. \qed
It is easy to see that while the conditions of Theorem 6 also imply that $K$ has antecedent-cover, they do not imply antecedent completeness, therefore useful cycles are still allowed. Thus, this special case, while somewhat limited, is still useful and interesting. An alternate way to state the conditions is to require that all partial inferences can be extended - note that we only need that property for inferences with non-zero weight, for Theorem 6 to hold. Hence, we can relax the requirements - requiring just that non-zero weight maximal inferences be complete, rather than all maximal inferences. That, in turn, allows the knowledge engineer to omit from the BKB rules that extend infeasible (i.e., zero-probability) partial inferences or states, without jeopardizing semantics.

We do not at present know how hard it is to test the conditions of Theorem 6 - however, in the special case where there are no variable-cycles - this problem (for bounded in-degree) is clearly in P; as we can test consequent completeness and antecedent cover, followed by groundedness, (these implying completeness of all maximal inferences), and then use the special case for testing normalization from Theorem 4.

5 Approximate Inference in BKBs

Exact and approximate inference on BKBs is NP-hard, as BNs are BKBs and exact inferencing in BNs, as well as approximate,

is NP-hard [3, 4, 20]. Nevertheless, numerous algorithms, both exact and approximate, exist for performing belief revision (finding the most probable complete state given the evidence) and belief updating (finding marginal probabilities given evidence). Prior work discussed finding most-likely-state given evidence for BKBs, but short of the naive scheme of looking at all possible inferences, no scheme for belief updating exists for BKBs. As exact belief updating schemes for BNs do not appear to be directly applicable to BKBs, we opted for stochastic approximation algorithms.

Stochastic sampling is a scheme commonly in use for Bayes networks. The idea is simply to generate a sample of (usually complete) states for the model, and then simply accumulate the relative probability mass of samples where an event of interest occurs (possibly weighted for sampling bias). While uniform sampling is simple in BKBs, work on approximate reasoning on BNs has shown that importance sampling [9] improves convergence, both theoretically and practically - and we expect this to be true for BKBs as well. Most sampling schemes for BNs proceed according to a topological sort of the network (an exception to that is Markov blanket schemes), but these are not directly applicable to BKBs. An importance sampling scheme (Algorithm 1) that works for grounded, normalized BKBs in the presence of cycles is shown below.

Algorithm 1 below produces an importance sample $S \subseteq E$ of a complete state over $X$ in accordance with $P$, the default distribution over $K$. That is, for null evidence, the probability of generating a sample $S$, denoted $P_S(S)$ is equal to $P(S)$. In order to converge to the correct probabilities, a sample weight $w_S$ must be used, equal to the probability of the sample $S$ divided by the probability of producing the sample, and here $w_S = 1$.

When we wish to use sampling in order to compute posterior probabilities given some evidence, we should only consider states consistent with the evidence. It is obviously advantageous, for efficiency reasons, if we can generate only samples consistent with the evidence, especially if the prior probability of the evidence is low. By inspecting the algorithm, it is clear that only samples consistent with the evidence
are generated. However, this process biases the sampling distribution - the probability of generating the sample is no longer equal to the prior probability of the sample, and must be compensated - a scheme known as (evidence) likelihood weighting. In a Bayes network, this factor is simply the product of the conditional probabilities for the evidence nodes. Similarly, in Algorithm 1, it is the product of rule weights for evidence nodes (step III.i), except when we cannot generate the evidence instantiation with a rule, and need to resort to a default (step III.2). As for likelihood weighting in Bayes networks, in both cases the probabilities used in the product are the probability of instantiating the evidence node as given in \( E \) - had the algorithm treated evidence nodes as non-evidence nodes.
Algorithm 1 (for producing a single sample) : \((S, w) = \text{importance-sample}(K, E)\)

**Input:** A BKB \(K = \{R_1^k, \ldots, R_n^k, R_1^{k_2}, \ldots, R_n^{k_2}, \ldots, R_1^{k_n}, \ldots, R_n^{k_n}\}\) over a set of variables \(\mathcal{X} = \{X_1, \ldots, X_n\}\), where for each \(1 \leq i \leq n\) and \(1 \leq j \leq k_i\), \(\text{span}(\text{cons}(R_j^i)) = X_i\), and a (possibly empty) state \(E\) over \(\mathcal{X}\), also called the evidence (or observations).

**Output:** A complete state \(S \supseteq E\) over \(\mathcal{X}\) sampled according to the distribution of \(K\), and a sample weight \(w\).

**I. Initialization**

1. Let \(\mathcal{I} \leftarrow \emptyset\), \(w \leftarrow 1\).
2. For \(i = 1..n\) do
   a. If \(X_i \notin \text{span}(E)\) then let \(P_i \in [0, 1]\) chosen randomly with uniform probability.
      Else let \(P_i = 1 + \epsilon\). (Arbitrary \(\epsilon > 0\))
   b. Let \(\mathcal{R}_i \leftarrow \emptyset\) \(R_i\) denotes the complementary set of CPRs w.r.t. \(\mathcal{I}\) and \(X_i\) already considered by the algorithm.
   c. Let \(N_i \leftarrow 0\) \((N_i\) is an accumulator of the weights of the CPRs in \(\mathcal{R}_i\)).
   d. Let \(D_i \leftarrow D(X_i)\) \((D_i\) keeps states of \(X_i\) open to default selection).

**II. Repeat** (* construct maximal inference for random state \(S\) *)

For \(i = 1..n\) do
   If \(X_i \notin \text{span}(\mathcal{I})\) then
      i. For \(j = 1..k_i\), while \(N_i < P_i\), do
         If \(R_j^i \in K/\mathcal{R}_i\) and \(R_j^i\) extends \(\mathcal{I}\) w.r.t. \(X_i\) then
            \(\mathcal{R}_i \leftarrow \mathcal{R}_i \cup \{R_j^i\}\), \(N_i \leftarrow N_i + w(R_j^i)\), \(D_i \leftarrow D_i - \text{cons}(R_j^i)\).
      ii. If \(X_i \in \text{span}(E)\) then
          If exists \(R \in \mathcal{R}_i\) that extends \(\mathcal{I}\) w.r.t. \(X_i\) s.t. \(\text{cons}(R) \in E\) then
              \(\mathcal{I} \leftarrow \mathcal{I} \cup \{R\}\), \(w \leftarrow w \times w(R)\).
          Else, if \(N_i \geq P_i\), then \(\mathcal{I} \leftarrow \mathcal{I} \cup \{R_j^i\}\).
   Until \(\mathcal{I}\) no longer changes.

**III. Completion** (* randomly assign remaining variables *)

1. Let \(S \leftarrow \text{st}(\mathcal{I}) \cup E\).
2. For each \(X_i\) not in \(\text{span}(\mathcal{I})\) do:
   If \(X_i \in \text{span}(E)\) then let \(w \leftarrow \frac{w}{P_i}\).
   Else choose randomly, with uniform probability, a value \(d \in D_i\) and let \(S \leftarrow S \cup (X_i, d)\).

As written, the worst-case time complexity of Algorithm 1 is \(O(|\mathcal{X}|^2)\), assuming bounded in-degree. When performing belief updating, however, the reasoning algorithm needs to generate a large number of samples (see below), and it thus makes sense to cache lists of extenders for partial inferences - this requires linear space (assuming, again, bounded in-degree), which can be either ahead of time, or during the run. Worst-case time complexity is reduced to \(O(|\mathcal{X}|^3)\) in a naive optimization, but it may be possible to reduce the time per sample to \(O(|\mathcal{X}| \log |\mathcal{X}|)\).

Finally, in order do perform belief updating, one can employ Algorithm 1, \text{importance-sample()}, as follows (this method is very commonly used, see, e.g., [9]). In order to compute posterior probabilities \(P(Q | E)\) for a query variable \(Q\), do:
1. Initialize $A_i \leftarrow 0$, for $i = 1..|D(Q)|$ (One accumulators $A_i$ for each possible state of $Q$)

2. Repeat as long as desired:
   
   (a) Let $(S, w) \leftarrow \text{importance-sample}(K, E)$

   (b) For $(Q, q_i) \in S$ (the choice is unique), let $A_i \leftarrow A_i + w$.

3. Now, or at any earlier time (after at least one sample with non-zero weight), compute the estimate of interest, for any $q_i \in D(Q)$:

   $\hat{P}(Q = q_i | E) = \frac{A_i}{\sum_{j=1}^{|D(Q)|} A_j}$

Naturally, this can be done in parallel for other query variables, with additional accumulators. The statement “as long as desired” is a common term for \textit{anytime} algorithms, meaning: repeat as many times as necessary, or as time permits, or using any other stopping criterion.

\textbf{Theorem 7} When $P(E) > 0$, the estimate $\hat{P}(Q = q_i | E)$ converges to $P(Q = q_i | E)$ in the limit, as the number of samples goes to infinity.

\textbf{Proof:} It is sufficient to show that every state $S$ such that $P(S | E) > 0$ has a non-zero probability of being generated, and that $w_{q_i}$, the number added to accumulator $A_i$ at each step, has an expected value equal to $P(E \land Q = q_i)$. We have:

$$E[w_{q_i}] = \sum_{S \in \mathcal{X}} P_{S | E}(S)w_{q_i}(S) = \sum_{(Q, q_i) \in S \in \mathcal{X}} P_{S | E}(S)w_{q_i}(S)$$

where $w_{q_i}(S)$ denotes the weight returned for $A_i$ by the algorithm due to sample $S$ (which is 0 if $(Q, q_i) \notin S$) and $P_{S | E}(S)$ is the probability of returning sample $S$ in the presence of evidence $E$. By construction, we have:

$$E[w_{q_i}] = \sum_{(Q, q_i) \in S \in \mathcal{X}} P_{S | E}(S)w_{q_i}(S) = \sum_{((Q, q_i) \cup E) \in S \in \mathcal{X}} P_{S}(S) = \sum_{((Q, q_i) \cup E) \in S \in \mathcal{X}} P(S) = P(E \land Q = q_i)$$

where the second equality follows from the fact that we generate no samples inconsistent with the evidence, and from the following property. For each sample $S$, the weight $w$ is equal to the probability of generating the evidence - given that the state of the non-evidence nodes is consistent with $S$. The third equality follows from the fact that for null evidence, we generate states according to default probability, and the last equality is due to the fact that we are summing probability over disjoint states that cover the event $\{E \land Q = q_i\}$. \qed

6 Discussion

Bayesian Knowledge Bases are a useful generalization of Bayes networks, that allows for a more concise representation of local independence (context-specific independence), as well as for cycles in the dependency graph. These extensions are made without sacrificing proper Bayesian probabilistic semantics.

In this paper, we presented such a semantics. As BKBs (optionally) also permit incompleteness, we have a set of consistent distributions, one of which is called the \textit{default} distribution, that spreads out
the unassigned probability mass equally between all states where that mass is otherwise undetermined. When the BKB is complete, correct rule semantics are observed in the sense that rule weights are the conditional probabilities under the probabilistic model.

While we have shown that checking a BKB for consistency (in the form of groundedness and normalization) is NP-hard in the general case, which is not the case for Bayes networks, we note that such tests are usually an offline task. Additionally, complexity of probabilistic reasoning (usually an online task) for both BKBs and Bayes networks are NP-hard anyway, and thus the somewhat negative results for consistency checking are not as much of a liability as they might seem at first. For some interesting special cases BKB consistency checking is easier, in particular if we already know that the BKB is grounded, or if there are no cycles. These latter cases are useful in that they still permit a very compact representation of special types of dependencies, such as OR relationship, or the "k out of n" and its generalization, the "noisy sum" relationship [10]. We do not know of any efficient way (i.e. size polynomial in n and k) to represent these in a Bayes network, even when using existing context-specific independence schemes.

One problem not addressed here is the issue of independence semantics, i.e. the question of "what statistical independence statements are sanctioned by the BKB?". While clearly the question can be answered by looking at the distribution, one might wish for a topological criterion, similar to the d-separation criterion used for Bayes networks. While d-separation is meaningful in a BKB, it does not give correct answers for determining independence, but it is still possible to use d-separation on graphs constructed from the original BKB to produce useful answers - an issue we address elsewhere [22].

Another area for useful future work is on other algorithms for reasoning with BKBs, as well as interesting special cases where both consistency-checking and reasoning can be made efficient. An obvious special case where one might take advantage of specific topology is a BKB with small strongly-connected components, but there may also be other interesting special cases. We are also looking into using the BKB representation as a concise probabilistic representation model for knowledge discovery [6].

Appendix

**Theorem 1** Let K be a normalized BKB over X, and f a distribution consistent with K. Then f is a joint probability distribution over X.

**Proof:** The proof for the general case appears in [14]. Here, we prove the theorem for the interesting special case where the BKB is "complete" in the following sense: we require that all inferences with non-zero dominated weight be maximal. (This special case supersedes the case where all inferences with non-empty dominated states are maximal, as well as BKBs with consequent completeness.)

**Lemma 1** All maximal inferences over a mutually exclusive set of CPRs (in particular, a BKB) are disjoint.

**Proof:** Let I,J be maximal inferences that have some complete state S in common. Then both I,J are relevant to S, and both are maximal relevant inferences for S. However, MRI(S) is unique, and thus I = J. ⊓⊔
Lemma 2 Every inference with non-zero dominated weight has a non-empty dominated composite state.

Proof: Let $I$ be an inference with non-zero dominated weight. From the definition of dominated weight, it must be the case that for each variable $X$ not assigned in $I$, the sum of weights of the rules in $mcs(I, X)$ is strictly less than 1, and by normalization $mcs(I, X)$ is incomplete - i.e. there is a value $v(X)$ that does not appear as a consequence in any rule in $mcs(I, X)$. Define the state:

$$S = st(I) \cup \{(X, v(X)) \mid X \in X' - \text{span}(I)\}$$

Clearly, $st(I) \subseteq S$, and for any inference $J \in ext^+(I)$, we have $st(J) \not\subseteq S$. Thus $I = MRI(S)$. □

Lemma 3 The weight of an inference $I$ is the sum of the dominated weights of all inferences $J$ such that $I \subseteq J$, i.e.

$$w(I) = \sum_{J \in ext^+(I)} w_D(J)$$

Proof: If $w(I) = 0$ the lemma holds trivially. We thus assume without loss of generality that $w(I) > 0$. We show the lemma by (structural) induction.

Base case - maximal inferences: Since maximal inferences have no extenders, we have $w_D(I) = w(I)$ whenever $I$ is a maximal inference - this follows immediately from the definition of dominated weight.

Inductive step: Assume that the lemma holds for every inference $J \in ext^+(I)$, all inferences that extend $I$.

Now, consider a non-maximal $I$ and let $X$ be a variable such that $\sum_{R \in mcs(I, X)} w(R) = 1$. Such a variable exists, as otherwise, since the product terms in the definition of dominated weight are non-negative (due to normalization) we get $w_D(I) > 0$, which would violate the assumption. For the above $X$, we can re-write $w(I)$ as:

$$w(I) = w(I) \sum_{R \in mcs(I, X)} w(R) = \sum_{R \in mcs(I, X)} w(I \cup \{R\})$$

Since each of the inferences $I \cup \{R\}$ is in $ext^+(I)$, we can assume from the induction hypothesis that the lemma holds for these $I \cup \{R\}$. Thus:

$$w(I) = \sum_{R \in mcs(I, X)} w(I \cup \{R\}) = \sum_{R \in mcs(I, X)} \sum_{L \in ext^+(I \cup \{R\})} w_D(L)$$

Since by construction the inferences $I \cup \{R\}$ are disjoint, we have:

$$w(I) = \sum_{L \in ext^+(I \cup \{R\})} w_D(L)$$

□

Proving the theorem now is straightforward. Every complete state $S$ has a unique maximal relevant inference $I$, which has a dominated weight $w_D(I)$ in $[0,1]$, and thus $f(S) \in [0,1]$ for every $f$ consistent with $K$. For the empty inference we have, by Lemma 3:

$$w(\{\}) = \sum_{I \in ext^+(\{\}) \text{ maximal}} w_D(I) = \sum_{S \in G\times} w_D(I) = \sum_{S \in G\times} f(S)$$

18
the latter equality following from the definition of $f$ and Lemmas 1 and 2. Since $w(\{\}) = 1$ by definition, the theorem follows.

**Theorem 2** Deciding groundedness of a rule $R$ in a correlation graph is NP-complete. Groundedness remains NP-hard in the special case where the BKB is consequent-complete and has antecedent-cover.

We have already shown the first part of the theorem. As the constructed BKB in the proof was not consequent-complete and did not have antecedent cover, the extended proof is a construction based on the proof for the general case, supplemented so as to meet these additional requirements. For the original proof we did not need to use cycles in the graph, but with the additional restrictions the cycles become necessary (without variable-cycles the groundedness problem is in P - and quite trivial).

The idea is to introduce a two cycles for each literal, such that exactly one of them is used in any inference, depending on the value of the literal.

**Proof:** By reduction from 3-SAT, as follows. For each propositional variable $V_i$, add the rules:

$$ \rightarrow x_i, \quad \rightarrow \neg x_i $$

and construct the following (cyclic) correlation graph fragment:

$$ x_i \rightarrow y_{i+}, \quad z_{i+} \land \neg x_i \rightarrow y_{i+}, \quad \neg z_{i+} \land \neg x_i \rightarrow y_{i+} $$

$$ x_i \rightarrow z_{i+}, \quad y_{i+} \land \neg x_i \rightarrow z_{i+}, \quad \neg y_{i+} \land \neg x_i \rightarrow z_{i+} $$

$$ y_{i+} \rightarrow w_{i+}, \quad \neg y_{i+} \rightarrow w_{i+} $$

To enforce consequent completeness, repeat the above rules with the negated form appearing in the consequents (i.e. with $\neg y_{i+}, \neg z_{i+}, \neg w_{i+}$ in the consequent, instead of $y_{i+}, z_{i+}, w_{i+}$, respectively). Likewise, construct the rules below, (as well as their negated form):

$$ \neg x_i \rightarrow y_{i-}, \quad z_{i-} \land x_i \rightarrow y_{i-}, \quad \neg z_{i-} \land x_i \rightarrow y_{i-} $$

$$ \neg x_i \rightarrow z_{i-}, \quad y_{i-} \land x_i \rightarrow z_{i-}, \quad \neg y_{i-} \land x_i \rightarrow z_{i-} $$

$$ y_{i-} \rightarrow w_{i-}, \quad \neg y_{i-} \rightarrow w_{i-} $$

The “$-$” subscript stands for “use for the literal where $V_i$ appears negated”, and likewise for the “$+$” subscript w.r.t. the un-negated form of the variable. Note that by construction, an inference will contain $w_{i+}$ or its negated form only if it contains $x_i$ (signifying a “true” assignment to $V_i$). Likewise, an inference will contain $w_{i-}$ or its negated form only if it contains $\neg x_i$.

For each clause $C_j$, do the following construction. Without loss of generality, let $C_j \equiv x_1 \vee x_2 \vee x_3$ (if a negated literal appears, use $w_{i-}$ below instead of $w_{i+}$). Introduce, for mutual exclusion, a 3-valued variable $a_j$ with values $a_j, a_j', a_j''$. The rules for $a_j$ are:

$$ \rightarrow a_j, \quad \rightarrow a_j', \quad \rightarrow a_j'' $$

And the clause determiners are:

$$ w_{i+} \land a_j \rightarrow C_j, \quad \neg w_{i+} \land a_j \rightarrow C_j $$

\footnote{While we can shorten the proof by introducing unary-valued nodes, and thus skip the necessity for the negated form of the rules, we did not wish to limit generality in this manner.}
as well as a negated version of these rules with $\neg C_j$ in the consequent instead of $C_j$. Clearly, any inference that has $C_j$ or $\neg C_j$ must have at least one of $\{x_1, x_2, x_3\}$. Finally, introduce a “sentence” node $s_n$ that “AND”s all the $C_j$, with $1 \leq j \leq n$, while taking care that the construction is polynomial, by using a “ladder” of auxiliary nodes $s_j$ as follows:

$$C_1 \rightarrow s_1, \quad \neg C_1 \rightarrow s_1$$

$$C_j \land s_{j-1} \rightarrow s_j, \quad \neg C_j \land s_{j-1} \rightarrow s_j, \quad (2 \leq j \leq n)$$

as well as the negated form with $\neg s_j$ in the consequent.

By construction, any inference that contains $s_n$ must contain all of $C_j$ or their negations, which in turn enforce the $x_i$ of the correct polarity to be in the inference. Thus, any inference containing $s_n$ has a corresponding truth assignment that satisfies the propositional sentence. Conversely, any satisfying truth assignment to the propositional sentence determines a (non-unique) inference that has $s_n$. Consequently, $s_n$ is grounded just when the propositional sentence is satisfied, and thus deciding groundedness is NP-complete. \hfill \square

References


