Evidence Absorption for Belief Networks

L. van der Gaag

RUU-CS-93-35
November 1993

Utrecht University

Department of Computer Science

Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

ISSN: 0924–3275

2 Preliminaries

In this section we review the basic notions involved in the belief network formalism and outline Pearl's algorithms for computing probabilities from a belief network and for processing evidence; for further details, the reader is referred to [Pearl, 1988] or [van der Gaag, 1992].

2.1 The Belief Network Formalism

A *belief network* for a given problem domain is a formal representation of the joint probability distribution Pr on the set of variables discerned in the said domain. A belief network consists of an *acyclic digraph* G and a set of associated functions. The digraph G of the network represents the variables in the domain and their probabilistic interdependencies. Each vertex in G represents a variable; the arcs in G are taken to represent the independency relation holding between the variables. Associated with the digraph is a set of *probability assessment functions* representing the numerical aspects of the joint probability distribution.

Before defining the notion of a belief network more formally, we provide some additional terminology and introduce our notational convention. In the sequel, we will assume that the variables discerned are binary, taking one of the values *true* and *false*. We will use the following notation: v_i denotes the proposition that the variable V_i takes the truth value *true*; $V_i = false$ will be denoted by $\neg v_i$. For a given set of variables V, the conjunction $C_V = \bigwedge_{V_i \in V} V_i$ of the variables from V is called the *configuration template* of V; a conjunction c_V of value assignments to the variables from V is called a *configuration* of V. The independency relation holding between the variables discerned in view of a given joint probability distribution Pr will be denoted as I_{Pr}; an independency statement $I_{Pr}(X, Y, Z)$ signifies that in the distribution Pr the sets of variables X and Z are conditionally independent given the set of variables Y.

We now define the notion of a belief network more formally.

Definition 2.1 *A* belief network *is a tuple* $B = (G, \Gamma)$ *such that*

- $G = (V(G), A(G))$ *is an acyclic digraph with vertices* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, *and*
- $\Gamma = \{\gamma_{V_i} \mid V_i \in V(G)\}$ *is a set of real-valued nonnegative functions* $\gamma_{V_i} : \{v_i, \neg v_i\} \times \{c_{\pi_G(V_i)}\} \to [0, 1]$, *called (conditional)* probability assessment functions, *such that for each configuration* $c_{\pi_G(V_i)}$ *of the set* $\pi_G(V_i)$ *of (immediate) predecessors of vertex* V_i *we have that* $\gamma_{V_i}(\neg v_i \mid c_{\pi_G(V_i)}) = 1 - \gamma_{V_i}(v_i \mid c_{\pi_G(V_i)})$, $i = 1, \ldots, n$.

Note that in the previous definition V_i is viewed as a vertex from the graph and as a probabilistic variable, alternatively.

To link the qualitative and quantitative parts of a belief network, a probabilistic meaning is assigned to the topology of the digraph of the network.

Definition 2.2 *Let* $G = (V(G), A(G))$ *be an acyclic digraph with vertices* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. *Let* s *be a chain in* G *from vertex* $V_i \in V(G)$ *to vertex* $V_j \in V(G)$; *let* V_s *be the set of vertices on* s *and let* A_s *be the set of arcs on* s. *Then, we say that* s *is* blocked *by a set* $W \subseteq V(G)$ *if one of the following conditions holds:*

- *The chain* s *contains a vertex* $X_2 \in W \cap V_s$ *and two vertices* $X_1, X_3 \in V_s$ *such that* $(X_2, X_1) \in A_s$ *and* $(X_2, X_3) \in A_s$.

- *The chain s contains a vertex $X_2 \in W \cap V_s$ and two vertices $X_1, X_3 \in V_s$ such that $(X_1, X_2) \in A_s$ and $(X_2, X_3) \in A_s$.*

- *The chain s contains vertices $X_1, X_2, X_3 \in V_s$ such that $(X_1, X_2) \in A_s$ and $(X_3, X_2) \in A_s$, and $\sigma^*(X_2) \cap W = \varnothing$, where $\sigma^*(X)$ denotes the set of vertices composed of X itself and all its descendants.*

Building on the notion of blocking we define the d-separation criterium.

Definition 2.3 *Let $G = (V(G), A(G))$ be an acyclic digraph with vertices $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. Let $X, Y, Z \subseteq V(G)$ be sets of vertices. The set Y is said to* d-separate *the sets X and Z, denoted as $\langle X \,|\, Y \,|\, Z \rangle_G^d$, if for each $V_i \in X$ and $V_j \in Z$ every chain from V_i to V_j in G is blocked by Y.*

The d-separation criterium provides for reading independency statements from a digraph, as stated in the following definition.

Definition 2.4 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let Pr be a joint probability distribution on $V(G)$ and let I_{Pr} be the independency relation of Pr. Then, the digraph G is called an* I-map *for Pr if for all $X, Y, Z \subseteq V(G)$ we have: if $\langle X \,|\, Y \,|\, Z \rangle_G^d$ then $I_{Pr}(X, Y, Z)$.*

The following proposition now states that the initial probability assessment functions of a belief network provide all information necessary for uniquely defining a joint probability distribution on the variables discerned that respects the independency relation portrayed by the graphical part of the network; henceforth, we will call this the *joint probability distribution defined by the network.*

Proposition 2.5 *Let $B = (G, \Gamma)$ be a belief network as defined in Definition 2.1, where $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. Then,*

$$Pr(C_{V(G)}) \quad = \quad \prod_{i=1,\ldots,n} \gamma_{V_i}(V_i \mid C_{\pi_G(V_i)})$$

defines a joint probability distribution Pr on the set of variables $V(G)$ such that G is an I-map for Pr.

2.2 Pearl's Algorithms

Once a belief network is constructed for a given application domain, it is used for making probabilistic statements concerning the variables discerned in the said domain. For this purpose, a set of algorithms for probabilistic inference is associated with the belief network formalism:

- an algorithm for (efficiently) computing probabilities of interest from a belief network, and

- an algorithm for processing evidence, that is, an algorithm for entering evidence into the network and subsequently (efficiently) computing the revised probability distribution given the evidence.

Several such sets of algorithms have been developed, [Pearl, 1988], [Lauritzen & Spiegelhalter, 1988], [Shachter, 1990]. In this paper, we build on Pearl's set of algorithms.

In briefly outlining the basic idea of Pearl's algorithms for probabilistic inference we will take an object-oriented point of view. The digraph of a belief network is viewed as a computational architecture by taking the vertices of the digraph as autonomous objects having a local processor capable of performing certain probabilistic computations and a local memory in which the associated probability assessment function is stored; the arcs of the digraph are viewed as bi-directional communication channels through which the objects can send messages. The vertices of the digraph send each other *parameters* providing information about the joint probability distribution and the evidence obtained sofar. Each vertex computes the (revised) probabilities of its values from the information it receives from its neighbours and its own local conditional probability assessment function. The time complexity of computing these parameters depends to a large extent on the number of neighbours a vertex has.

Initially, the belief network is in an *equilibrium state*: recomputing parameters will not result in a change in any of them. When a piece of evidence for a specific variable is entered into the belief network, this equilibrium is perturbed. The variable modifies the parameters to send to its neighbours to reflect the entered evidence. These modifications activate updating parameters throughout the network: after receiving modified parameters, each vertex in turn computes new parameters to send to its neighbours. The new information is thus diffused through the network in a single pass. This process is termed *spreading activation*. The belief network will reach a new equilibrium state once each vertex in the digraph has been visited by the process of spreading activation.

We emphasize that the computational effort involved in probabilistic inference with a belief network is largely determined by the sparsity of the digraph of the network.

3 Evidence Absorption and Pearl's Algorithms

In building a belief network for a given application domain, its graphical part is constructed to reflect as many of the independencies between the variables discerned as possible. There are several reasons for seeking to represent these independencies to accuracy. The most important reason is a computational one. The more independencies are represented explicitly, the sparser the digraph of the network will be, and as we have mentioned above, the computational effort involved in probabilistic inference with a belief network is largely determined by the sparsity of its digraph. In fact, Pearl's algorithms for probabilistic inference exploit the independencies portrayed by the digraph of a belief network explicitly and perform the better from a computational point of view as the digraph is sparser.

Now consider entering evidence into a belief network and processing it. Each piece of evidence provides additional information on the joint probability distribution in a given context. More in specific, new dependencies and independencies may hold in this context. It is possible to modify the topology of the digraph of the belief network dynamically so as to reflect these newly created dependencies and independencies. In the sequel, however, we will argue that it is worthwhile to modify the topology of the digraph to reflect the new *in*dependencies only. We propose extending Pearl's algorithms for probabilistic inference to this end with the method of *evidence absorption*.

In Section 3.1 we introduce the method of evidence absorption. Section 3.2 briefly reviews the algorithm for probabilistic inference introduced by R.D. Shachter, [Shachter, 1990], that incorporates the method of evidence absorption. In Section 3.3 we propose integrating the method into Pearl's algorithms.

3.1 The Method of Evidence Absorption

Informally speaking, the method of evidence absorption amounts to modifying a belief network after a piece of evidence has been entered for a specific variable: the topology of the digraph of the network is modified by deleting all arcs departing from the vertex for which the evidence has been entered, and the probability assessment functions for the (former) successors of this vertex are adjusted to reflect the evidence. The modified network is defined more formally in the following definition.

Definition 3.1 *Let $B = (G, \Gamma)$ be a belief network where $G = (V(G), A(G))$ is an acyclic digraph and $\Gamma = \{\gamma_{V_i} \mid V_i \in V(G)\}$ is a set of probability assessment functions. Let V_i be a vertex in G for which the evidence $V_i =$ true is entered. We define the tuple $B^{v_i} = (G^{v_i}, \Gamma^{v_i})$ by*

- *$G^{v_i} = (V(G^{v_i}), A(G^{v_i}))$ is an acyclic digraph such that $V(G^{v_i}) = V(G)$ and $A(G^{v_i}) = A(G) \setminus \{(V_i, V_j) \mid V_j \in \sigma_G(V_i)\}$, and*
- *$\Gamma^{v_i} = \{\gamma^{v_i}_{V_j} \mid V_j \in V(G)\}$ is the set of real-valued nonnegative functions $\gamma^{v_i}_{V_j} : \{v_j, \neg v_j\} \times \{c_{\pi_{G^{v_i}}(V_j)}\} \rightarrow [0,1]$ such that*
  - *$\gamma^{v_i}_{V_j}(V_j \mid C_{\pi_{G^{v_i}}(V_j)}) = \gamma_{V_j}(V_j \mid C_{\pi_G(V_j)\setminus\{V_i\}} \wedge v_i)$, for all vertices $V_j \in \sigma_G(V_i)$, and*
  - *$\gamma^{v_i}_{V_k}(V_k \mid C_{\pi_{G^{v_i}}(V_k)}) = \gamma_{V_k}(V_k \mid C_{\pi_G(V_k)})$, for all vertices $V_k \in V(G) \setminus \sigma_G(V_i)$.*

The tuple $B^{\neg v_i} = (G^{\neg v_i}, \Gamma^{\neg v_i})$ is defined analogously by substituting $\neg v_i$ for v_i in the above.

The method of evidence absorption is illustrated by means of an example.

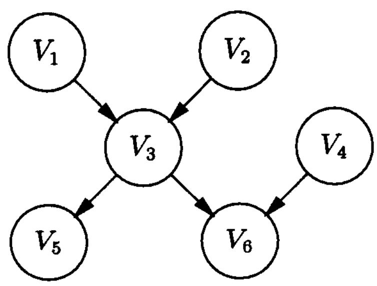

Figure 1: The Digraph G of the Belief Network B.

Example 3.2 Consider the belief network $B = (G, \Gamma)$ where G is the singly connected digraph shown in Figure 1, and Γ consists of the six conditional probability assessment functions $\gamma_{V_1}, \ldots, \gamma_{V_6}$:

$$\gamma_{V_1}(V_1)$$
$$\gamma_{V_2}(V_2)$$
$$\gamma_{V_3}(V_3 \mid V_1 \wedge V_2)$$
$$\gamma_{V_4}(V_4)$$
$$\gamma_{V_5}(V_5 \mid V_3)$$
$$\gamma_{V_6}(V_6 \mid V_3 \wedge V_4)$$

Now suppose that the evidence $V_3 = true$ is obtained for the variable V_3. The belief network B then is modified to $B^{v_3} = (G^{v_3}, \Gamma^{v_3})$. The digraph G^{v_3} is obtained from G by deleting all arcs departing from vertex V_3, and is shown in Figure 2; the evidence is represented by drawing vertex V_3 with shading.

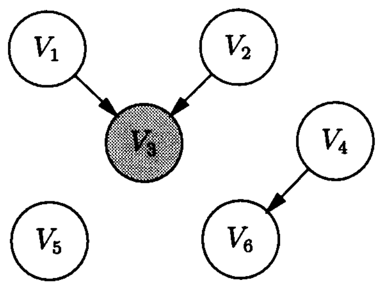

Figure 2: The Digraph G^{v_3} of the Belief Network B^{v_3}.

The set Γ^{v_3} consists of the six functions $\gamma_{V_1}^{v_3}, \ldots, \gamma_{V_6}^{v_3}$ that are obtained from the probability assessment functions of the original belief network B:

$$\begin{array}{lcl}
\gamma_{V_1}^{v_3}(V_1) & = & \gamma_{V_1}(V_1) \\
\gamma_{V_2}^{v_3}(V_2) & = & \gamma_{V_2}(V_2) \\
\gamma_{V_3}^{v_3}(V_3 \mid V_1 \wedge V_2) & = & \gamma_{V_3}(V_3 \mid V_1 \wedge V_2) \\
\gamma_{V_4}^{v_3}(V_4) & = & \gamma_{V_4}(V_4) \\
\gamma_{V_5}^{v_3}(V_5) & = & \gamma_{V_5}(V_5 \mid v_3) \\
\gamma_{V_6}^{v_3}(V_6 \mid V_4) & = & \gamma_{V_6}(V_6 \mid v_3 \wedge V_4)
\end{array}$$

□

The modified network resulting after evidence absorption once more is a belief network; this is stated more formally in the following lemma.

Lemma 3.3 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ and let the tuple B^{v_i} be defined as in Definition 3.1. Then, B^{v_i} is a belief network.*

Proof. The property stated in the lemma follows directly from Definition 2.1 and Definition 3.1. □

A similar property holds for the modified network $B^{\neg v_i}$.

3.2 Evidence Absorption and Arc Reversal

The method of evidence absorption discussed in the previous subsection has been introduced by R.D. Shachter and is part of an algorithm for processing evidence in a belief network, [Shachter, 1986], [Shachter, 1990]. The basic idea of this algorithm is to *eliminate* a vertex from the digraph of a belief network as soon as it is instantiated. The algorithm is composed of two phases. When a piece of evidence is entered for a specific variable, the method of evidence absorption is applied. Subsequently, the evidence is spread throughout the network by a method called *evidence propagation* which basically consists of repeated application of an arc modifying operation called *arc reversal*. In these two phases, the topology of the digraph of the network is modified dynamically to reflect the newly created dependencies and independencies. In doing so, new arcs may be inserted into the digraph to portray the newly created dependencies among the remaining variables and for these arcs accompanying conditional probabilities have to be calculated.

Shachter's algorithm for processing evidence has several drawbacks, as has already been noted by J. Pearl, [Pearl, 1988 (pp. 144 – 145)]. Related to the computational effort involved, we note that eliminating a vertex from a belief network is *computationally expensive*: the algorithm has an exponential worst-case time complexity. In addition, the algorithm involves computations that are not local to the vertices of the digraph. So, as opposed to Pearl's algorithms for probabilistic inference, Shachter's algorithm requires a *global supervisor*. These drawbacks cannot be alleviated if the aim is to eliminate an instantiated variable from a belief network.

Upon close examination of Shachter's algorithm for processing evidence and the drawbacks mentioned above, it becomes apparent that these drawbacks are attributed entirely to the arc reversal operation employed during evidence propagation: it is this method that accounts for the exponential time complexity and for the need of a global supervisor. As opposed to evidence propagation, evidence absorption can be performed in linear time; in addition, all computations involved are local to a vertex and its successors, and can be effectuated without supervision.

3.3 Incorporating Evidence Absorption into Pearl's Algorithms

We reconsider the method of evidence absorption as introduced in Section 3.1. It will be evident that for a given piece of evidence this method takes care of modifying a belief network to reflect the new *independencies* holding in the context of this evidence, only. Now recall that Pearl's algorithms for probabilistic inference exploit the independencies reflected by the digraph of a belief network directly, and that the algorithms perform the better from a computational point of view as the digraph is sparser. Since the method of evidence absorption tends to delete arcs from the digraph of a belief network, it is worthwhile to integrate this method into Pearl's algorithms to cut down on the computational expense in future probabilistic inference. The basic idea is as follows. When a piece of evidence for a specific variable is entered into a belief network, the method of evidence absorption is applied. Subsequently, Pearl's algorithms are called upon to propagate the evidence. It is emphasized that the instantiated vertex is *not* eliminated from the digraph: as the method of evidence absorption strives to incorporate new independencies only, an instantiated vertex has to remain in the digraph to reflect the newly created dependencies.

The correctness of the enhanced algorithms for probabilistic inference as outlined above derives from the observation that after propagation of the evidence the modified belief network and the original belief network model the same updated joint probability distribution. We prove two lemmas to support this observation.

Lemma 3.4 *Let $B = (G, \Gamma)$ be a belief network and let Pr be the joint probability distribution defined by B. Let V_i be a vertex in G for which the evidence $V_i = true$ is propagated throughout the network; let Pr^{v_i} be the (updated) joint probability distribution. Now, let the modified network $B^{v_i} = (G^{v_i}, \Gamma^{v_i})$ be defined as in Definition 3.1 and let $\mathbb{P}$ be the joint probability distribution defined by B^{v_i}. Furthermore, let $\mathbb{P}^{v_i}$ be the (updated) joint probability distribution after propagation of the evidence $V_i = true$. Then, $Pr^{v_i} = \mathbb{P}^{v_i}$.*

Proof. We consider the belief network $B = (G, \Gamma)$ and the joint probability distribution Pr defined by B, and the belief network $B^{v_i} = (G^{v_i}, \Gamma^{v_i})$ and its joint probability distribution $\mathbb{P}$. To prove $Pr^{v_i} = \mathbb{P}^{v_i}$, we show that

- $Pr(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n) = \mathbb{P}(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n)$, and
- $Pr(v_i) = \mathbb{P}(v_i)$.

The main result then follows from the definition of conditional probability.

From Proposition 2.5, we have that the joint probability distribution Pr defined by the belief network B can be expressed as

$$\begin{aligned} Pr(V_1 \wedge \cdots \wedge V_n) &= \prod_{j=1,\ldots,n} Pr(V_j \mid V_{j-1} \wedge \cdots \wedge V_1) = \\ &= \prod_{j=1,\ldots,n} \gamma_{V_j}(V_j \mid C_{\pi_G(V_j)}) \end{aligned}$$

From this expression, we derive an expression for the marginal distribution $Pr(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n)$ by substituting the value v_i for the variable V_i.

The joint probability distribution $\mathbb{P}$ defined by the belief network B^{v_i} can be expressed as

$$\begin{aligned} \mathbb{P}(V_1 \wedge \cdots \wedge V_n) &= \prod_{j=1,\ldots,n} \mathbb{P}(V_j \mid V_{j-1} \wedge \cdots \wedge V_1) = \\ &= \prod_{j=1,\ldots,n} \gamma^{v_i}_{V_j}(V_j \mid C_{\pi_{G^{v_i}}(V_j)}) \end{aligned}$$

From this expression, we derive an expression for the marginal distribution $\mathbb{P}(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n)$ by substituting the value v_i for the variable V_i.

To show that $Pr(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n) = \mathbb{P}(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n)$, it suffices to show that the corresponding terms in the expressions for the marginal distributions are the same. We distinguish between several different cases:

- for the assessment functions γ_{V_i} and $\gamma^{v_i}_{V_i}$ for the variable V_i, we have that

  $$\gamma_{V_i}(v_i \mid C_{\pi_G(V_i)}) = \gamma^{v_i}_{V_i}(v_i \mid C_{\pi_{G^{v_i}}(V_i)})$$

- for the assessment functions γ_{V_j} and $\gamma^{v_i}_{V_j}$ for a variable V_j such that $V_j \in \sigma_G(V_i)$, we have that

  $$\gamma_{V_j}(V_j \mid C_{\pi_G(V_j) \setminus \{V_i\}} \wedge v_i) = \gamma^{v_i}_{V_j}(V_j \mid C_{\pi_{G^{v_i}}(V_j)})$$

  by definition;

- none of the other assessment functions involves the variable V_i; for the functions γ_{V_k} and $\gamma^{v_i}_{V_k}$ for a variable V_k such that $V_k \in V(G) \setminus (\sigma_G(V_i) \cup \{V_i\})$, we therefore have that

  $$\gamma_{V_k}(V_k \mid C_{\pi_G(V_k)}) = \gamma^{v_i}_{V_k}(V_k \mid C_{\pi_{G^{v_i}}(V_k)})$$

We conclude that

$$Pr(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n) = \mathbb{P}(V_1 \wedge \cdots \wedge V_{i-1} \wedge v_i \wedge V_{i+1} \wedge \cdots \wedge V_n)$$

The property $Pr(v_i) = \mathbb{P}(v_i)$ now follows by further marginalization. □

A similar property can be proven with respect to the piece of evidence $V_i = false$.

In the previous lemma, it has been shown that after propagation of a piece of evidence the modified network and the original belief network represent the same updated joint probability distribution. In addition, the two networks represent the same independencies given the evidence. Note that this property is required in addition to the previous one to guarantee correct probabilistic statements being the result of applying Pearl's algorithm to the modified network.

Lemma 3.5 *Let $B = (G, \Gamma)$ be a belief network with $G = (V(G), A(G))$. Let V_i be a vertex in G for which the evidence $V_i = true$ is entered, and let the modified network $B^{v_i} = (G^{v_i}, \Gamma^{v_i})$ be as in Definition 3.1. Then, $\langle X \mid Y \mid Z \rangle^d_G$ if and only if $\langle X \mid Y \mid Z \rangle^d_{G^{v_i}}$, for all $X, Y, Z \subseteq V(G)$ such that $V_i \in Y$.*

Proof. We consider the digraphs G and G^{v_i} of the belief networks B and B^{v_i}, respectively. The property stated in the lemma will be proven by contradiction.

- Suppose that there exist vertices V_j and V_k in $V(G)$ and a set $Y \subseteq V(G)$ with $V_i \in Y$ such that $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_G$ holds and $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_{G^{v_i}}$ does not hold.

  Since the independency statement $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_{G^{v_i}}$ does not hold for the digraph G^{v_i}, it follows that there exists a chain s in G^{v_i} from V_j to V_k that is not blocked by Y. Now, we distinguish between two cases:

  - Suppose that vertex V_i is not on the chain s from V_j to V_k. Then, s is a chain in G also. From the independency statement $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_G$ we have that the chain s is blocked by the set Y in G. But then, s is blocked in G^{v_i} by the set Y as well.
  - Suppose that vertex V_i occurs on the chain s. From the topology of G^{v_i} we have that vertex V_i has no outgoing arcs. Therefore, the two arcs incident on V_i on the chain s are directed towards V_i. Then, s is a chain in G also. From the independency statement $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_G$ we have that the chain s is blocked by Y in G. But then, s is blocked in G^{v_i} by the set Y as well.

- Now, suppose that there exist vertices V_j and V_k in $V(G)$ and a set $Y \subseteq V(G)$ with $V_i \in Y$ such that $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_{G^{v_i}}$ holds and $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_G$ does not hold.

  Since the independency statement $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_G$ does not hold in the digraph G, we have that there exists a chain s in G from V_j to V_k that is not blocked by the set Y. Now, we distinguish between two cases:

  - Suppose that vertex V_i is not on the chain s. Then, s is a chain from V_j to V_k in G^{v_i} also. From the independency statement $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_{G^{v_i}}$ it follows that the chain s is blocked by the set Y in G^{v_i}. But then, s is blocked in G as well.
  - Suppose that vertex V_i occurs on s. Then, we have that the two arcs incident on V_i on the chain s are directed towards V_i; otherwise, the chain s is blocked by the set Y in G through $V_i \in Y$. The chain s therefore occurs in G^{v_i} also. From the independency statement $\langle \{V_j\} \mid Y \mid \{V_k\} \rangle^d_{G^{v_i}}$ we have that s is blocked by Y in G^{v_i}. Note that since $V_i \in Y$, the set Y must contain at least one other vertex, that is responsible for blocking s. But then, s is blocked by Y in G as well.

proposed incorporating the method of *evidence absorption* into Pearl's algorithms for probabilistic inference. This method amounts to dynamically modifying a belief network as evidence becomes available by local computations only. The ability of the method to improve on the average-case performance of probabilistic inference derives to a large extent from the method's property of explicitly incorporating the new independencies created by the observation of the evidence into the digraph of the network: the method tends to make a digraph fall apart into separate components. The incorporation of the method into Pearl's algorithms then allows for lessening the computational expense involved in further probabilistic computations as these may be restricted to one component of the digraph only.

Here, we have defined the method of evidence absorption and have proven that the enhanced algorithms provide for exact inference. The ability of the method to save on the computational effort spent on the average-case performance of these algorithms, however, still remains to be demonstrated. At present, we are conducting several experiments on differing classes of randomly generated belief networks to gain insight into the impact of applying the method; the results from these experiments are presented in a forthcoming paper, [van der Gaag, 1993].

Although the incorporation of the method of evidence propagation into Pearl's algorithms is likely to yield major savings on probabilistic computation in problem solving, we expect still further optimisation by combining the method with a lazy evaluation approach to evidence propagation. Future research therefore will be aimed at the development of such a lazy evaluation approach for exact probabilistic inference in the context of the method of evidence absorption.

References

[Andreassen et al., 1987] S. Andreassen, M. Woldbye, B. Falck, S.K. Andersen (1987). MUNIN - A causal probabilistic network for interpretation of electromyographic findings, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 366 – 372.

[Bruza & van der Gaag, 1992] P.D. Bruza, L.C. van der Gaag (1992). *Index Expression Belief Networks for Information Disclosure*, Technical Report RUU-CS-92-21, Utrecht University.

[Heckerman *et al.*, 1992] D.E. Heckerman, E.J. Horvitz, B.N. Nathwani (1992). Toward normative expert systems. Part 1: The Pathfinder project, *Methods of Information in Medicine*, vol. 31, pp. 90 – 105.

[Lauritzen & Spiegelhalter, 1988] S.L. Lauritzen, D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society, Series B*, vol. 50, pp. 157 – 224.

[Pearl, 1988] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Morgan Kaufmann, Palo Alto.

[Shachter, 1986] R.D. Shachter (1986). Evaluating influence diagrams, *Operations Research*, vol. 34, pp. 871 – 882.