1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June - 2 July 1999

A Review of Propagation Algorithms for Imprecise Probabilities

Andrés Cano, Serafín Moral
Departamento de Ciencias de la Computación e I.A.
Universidad de Granada, 18071 - Granada - Spain
e-mail: acu,smc@decsai.ugr.es

Abstract

This paper reviews algorithms for local computation with imprecise probabilities. These algorithms try to solve problems of inference (calculation of conditional or unconditional probabilities) in cases in which there are a large number of variables. There are two main types depending on the nature of assumed independence relationships in each case. In both of them the global knowledge is composed of several pieces of local information. The objective is to carry out a sound global computation but mainly using the initial local representation.

Keywords. Propagation algorithms, valuations based systems, imprecise probabilities.

1 Introduction

This paper reviews local computation algorithms to compute with imprecise probabilities. In general, it is assumed the case closed and convex sets with a finite set of extreme probability distributions.

Local computation techniques have been successfully applied to the case of classical probabilities [30, 23]. With the development of algorithms in an abstract framework [36, 6], these techniques have also been applied to other formalisms of representing uncertainty including convex sets of probabilities.

Several and very different methodologies have been designed for the computation with imprecise probabilities. In this paper, these approaches are classified according to the underlying independence relationships. Section 2 considers the fundamentals of the calculus with convex sets. The concepts of marginal, conditional and 'a posteriori' information are given. Two definitions of independence [11, 8] are also given. The first one is called unknown interaction. It is a rather weak notion and it is not a generalization of classical stochastic independence. The second one, called strong independence, generalizes stochastic independence and it is the most common type of independence that can be found in the propagation of imprecise probabilities literature. Section 3 introduces local computation techniques in an abstract way, following Shafer and Shenoy [36]. Section 4 is devoted to local computation algorithms for the case of unknown interaction (this problem is often known as propagation of probabilistic restrictions). Section 5 is devoted to the algorithms under strong independence. It is shown how this problem can be transformed in an optimization problem by adding new variables which are called transparent variables [4]. Then, some general global optimization techniques to solve this problem are considered, such as simulated annealing, gradient techniques or genetic algorithms. Finally, Section 6 is devoted to the conclusions.

2 Basic Notions of Convex Sets of Probability Distributions

Assume that we have a population Ω and a n-dimensional variable $(X_1, X_2, \ldots, X_n)$ defined on Ω and such that each X_i takes its values on a finite set U_i. For each $I \subseteq \{1, \ldots, n\}$, X_I will denote the variable $(X_i)_{i \in I}$, taking values on set $\prod_{i \in I} U_i$ (denoted by U_I).

In this paper our knowledge of a problem will be given by convex sets of probabilities or conditional probabilities. In general, a piece of information relating the variables in I will be a closed and convex set, H, of mappings: $h : U_I \Leftrightarrow \mathbb{R}$, with a finite set of extreme points.

As U_I is finite, a mapping is given by the vector of values $(h(u))_{u \in U_I}$. By this reason we shall use the word vector or point to refer to a mapping h. This point has $|U_I|$ dimensions, where $|U_I|$ is the number of elements of U_I. The convex set will be usually degenerated in this space (some linear equation satisfied). For example, in the case of probability distributions

the different values have to add one.

If h is a function from U_I onto $\mathbb{R}$, and $J \subseteq I$, then the marginal of h to U_J is the function $h^{\downarrow J}$ defined on U_J and given by, $h^{\downarrow J}(u) = \sum_{v^{\downarrow J}=u} h(v)$, where $v^{\downarrow J}$ is the element from U_J obtained by deleting the coordinates in $I \Leftrightarrow J$. If H is a convex set of functions on U_I, with extreme points, Ext $(H) = \{h_1, \ldots, h_k\}$, and $J \subseteq I$ then the marginalization of H to J is the convex set given by,

$$H^{\downarrow J} = \text{CH } \{h_1^{\downarrow J}, \ldots, h_k^{\downarrow J}\} \tag{1}$$

where CH stands for the convex hull operator (the minimum convex set containing a given set).

$H^{\downarrow J}$ is equal to the marginalization on U_J of all the functions h in H. However not all the marginal of the extreme points are extreme and it is possible that some of the functions $h_i^{\downarrow J}$ are not extreme in $H^{\downarrow J}$.

Assume that h is a function from U_I onto $\mathbb{R}$ and h' a function from U_J onto $\mathbb{R}$, then the multiplication of these two functions is a function, $h.h'$, defined on $U_{I\cup J}$ and given by, $h.h'(u) = h(u^{\downarrow I}).h'(u^{\downarrow J})$.

This operation is extended to convex sets of functions. If H is a convex set of mappings in U_I, and H' is a convex set in U_J, with $\text{Ext}(H) = \{h_1, \ldots, h_k\}, \text{Ext}(H') = \{h'_1, \ldots, h'_l\}$. Then the combination of H and H' will be a convex set of mappings in $U_{I\cup J}$, $H \otimes H'$ given by

$$H \otimes H' = \text{CH}\{h_1.h'_1, .., h_1.h'_l, \ldots, h_k.h'_1, .., h_k.h'_l\} \tag{2}$$

An important remark of this operation is that $H \otimes H'$ is not equal to the set obtained by multiplying all the functions on H and all the functions on H', because this set may be non-convex. It is the minimum convex set containing it.

Another important operation with convex sets is the intersection. If H is a convex set of mappings in U_I, and H' is a convex set in U_J, then $H \cap H'$ is the convex set of mappings h defined on $U_{I\cup J}$ verifying that $h^{\downarrow I} \in H$ and $h^{\downarrow J} \in H'$. $H \cap H'$ is also a closed and convex set with a finite set of extreme points.

We shall consider that our 'a priori' knowledge about how a variable takes its values is represented by a closed and convex set of probability distributions, H, with a finite set of extreme points $\text{Ext}(H) = \{p_1, ..., p_k\}$. Each p_i is a probability distribution on U and $\text{Ext}(H)$ are the extreme points of H.

In the following, we give the elementary concepts to work with several variables under convex sets of probability distributions. We shall assume that (X, Y) is a pair of variables, X taking values on a finite set U, and Y on a finite set V. If H is an 'a priori' piece of information about (X, Y), that is a convex set of probability distributions on $U \times V$, then the marginal of this information for variable X, H^X, is the convex set of probabilities $H^{\downarrow U}$.

A conditional information about Y given X will be a closed and convex set, $H^{Y|X}$, of mappings, $h : U \times V \Leftrightarrow\!\!\rightarrow [0, 1]$, verifying

$$\sum_{v\in V} h(u, v) = 1, \forall u \in U$$

and with a finite set of extreme points, $\text{Ext}(H^{Y|X}) = \{h_1, \ldots, h_l\}$.

This is more general than assuming that a conditional piece of information is a convex set of probabilities for every possible value of X, that is, for every element $u \in U$ [7].

From a marginal convex set H^X and a conditional information $H^{Y|X}$, we can calculate a global information about (X, Y): $H = H^X \otimes H^{Y|X}$. However, it is very simple to prove that every global convex set of probabilities can not be decomposed this way [28].

Above, we have considered the problem related with general probabilistic information, that is, information valid for all the population under study. Now, a different aspect is considered: in front of a particular case, How to particularize the general knowledge to the observations we have carried out on it? This is called focusing conditioning [13]. The resulting information will be called 'a posteriori' information.

Here we shall only consider the most usual definition of conditioning consisting in focusing all the possible probability distributions. Other alternatives can be found in Moral and Campos [29].

Assume a convex set for variable X: $H = \text{CH}\{p_1, \ldots, p_k\}$ and that we have observed *'X belongs to A'*, then the result of conditioning is the convex set, $H|A$, generated by the points $\{p(.|A) : p \in H, p(A) \neq 0\}$.

The definition can be extended to the case in which l is a general likelihood function, $l : U \to [0, 1]$. $H|l$ is equal to the set generated by points $\{p(.|l) : p \in H, E_p[l] \neq 0\}$, where $p(.|l)$ is calculated by applying Bayes's rule and E_p is the mathematical expectation.

If we have variables X and Y taking values on U and V respectively and H is a global convex set of probability distributions for these two variables, then by $H^X|(Y \in B)$ we will denote the conditioning of H to the set $U \times B$ and the marginalization of the result to

U. That is, $H^X|(Y \in B) = (H|U \times B)^{\downarrow U}$.

The concept of conditional independence is fundamental for propagation algorithms. Here we shall consider only two types of independence: unknown interaction and strong independence. A more detailed study of independence with alternative definitions can be found in De Campos and Moral [11] and in Couso, Moral, and Walley [8], where different justifications and conditions for their application are considered.

Intuitively, X and Y are conditionally independent given Z, when they are independent under a perfect knowledge of the value of Z.

Definition 1 (Unknown Interaction) *If $H^{X,Y,Z}$ is a convex set of probability distributions for (X, Y, Z), then we say that there is unknown interaction of X and Y given Z if and only if $H^{X,Y,Z} = H^{X,Z} \cap H^{Y,Z}$.*

This is a very weak definition of conditional independence and it is not a generalization of probabilistic conditional independence. It does not imply that there is no relationship between the variables given Z. Only implies that knowing the probability about Z there is not influence between the knowledge of the probability distributions about X and Y. In that sense, it is some type of independence. In fact, it is the minimum we can ask to make the local computation possible. When there is unknown interaction, the joint set $H^{X,Y,Z}$ is the natural extension of marginal sets $H^{X,Z}$ and $H^{Y,Z}$ (the least informative set with these marginals). As this definition is very weak and is not really a generalization of stochastic independence, this case in usually referred in the literature as *not assuming independence relationships*. In some cases we will use this expression, though as we have said earlier it implies the verification of a weak independence property.

Definition 2 (Strong conditional independence) *If $H^{X,Y,Z}$ is a global convex set of probabilities for (X, Y, Z), we say that X and Y are conditionally strong independent given Z if and only if, $H^{X,Y,Z} = H^{X,Z} \otimes H^{Y|Z}$ or $H^{X,Y,Z} = H^{Y,Z} \otimes H^{X|Z}$.*

This definition is really a generalization of stochastic independence. If we have two joint sets $H_1^{X,Y,Z}$ and $H_2^{X,Y,Z}$ with the same marginals and such that unknown interaction is satisfied in $H_1^{X,Y,Z}$ and strong independence in $H_2^{X,Y,Z}$, then it is easy to show that $H_2^{X,Y,Z} \subseteq H_1^{X,Y,Z}$. So strong independence produces more informative joint sets (there are less possible probability distributions) than unknown interaction.

This is the usual condition considered in the literature when it is said that independence relationships are assumed, without specifying which type of independence relationships are being considered.

If we have a variable X taking values on a finite set U, then a convex set of probabilities, H, can be given by a set of linear restrictions, R. Each element in $r \in R$ is an inequality:

$$r \equiv \sum_{u \in U} \alpha_u . p(u) \leq \beta \tag{3}$$

The set of probability distributions verifying a set of restrictions R is always a convex set, which will be denoted as $\mathcal{H}(R)$. The set of all the restrictions which are verified by a convex set H will be denoted as $\mathcal{R}(H)$. It is immediate that $R \subseteq \mathcal{R}(\mathcal{H}(R))$ and that $\mathcal{H}(\mathcal{R}(H)) = H$.

If $\mathcal{H}(R) = H$ we will say that the set of restrictions R defines the convex set H. In general, given a convex set of probability distributions with a finite set of extreme points, there is a finite minimal set of restrictions defining it.

A set of restrictions R is said to be minimal if and only if for every set of restrictions $R' \subseteq R$ such that $\mathcal{H}(R) = \mathcal{H}(R')$ we have that $R = R'$. A restriction $r \in R$ is said to be redundant if and only if $\mathcal{H}(R) = \mathcal{H}(R \Leftrightarrow \{r\})$.

For a convex set, H, we can use the representation given by a finite set of points including its extreme points or the one given by a finite set of restrictions defining it. In both cases, it is preferable for the representation to be minimal.

Algorithms to calculate minimal representations and to make transformations between them are classical in the theory of Convex Sets. In concrete we can point out the following ones:

- *Convex Hull calculation.*- These algorithms are used to remove all the non-extreme points of a finite set. At the same time, they calculate a minimal set of restrictions defining the convex hull containing these points. Descriptions can be found in [14, 32].
- *Redundancy elimination.*- These algorithms remove the set of redundant constraints from a finite set. A survey can be found in [20].
- *Vertex enumeration.*- These algorithms calculate all the extreme points of the convex set defined by a set of linear constraints. A survey can be found in [26].

Depending on the operations we want to carry out, some representations are more appropriate than the

others. For the intersection the restrictions representation is more appropriate. If $\mathcal{H}(R_1) = H_1, \mathcal{H}(R_2) = H_2$, then it is immediate to show that $\mathcal{H}(R_1 \cup R_2) = H_1 \cap H_2$. A redundancy elimination algorithm can be used to obtain a minimal representation of $H_1 \cap H_2$.

For the combination the most appropriate representation is the use of the extreme points. In fact this operation is defined by means of (2) in terms of the extreme points defining the convex set. However not all the calculated points are extreme, and a convex hull algorithm should be used if we want to keep minimal the representation.

The marginalization is expressed in (1) in terms of the extreme points. So the extreme points representation is appropriate for this operation. As in the case of combination a convex hull algorithm is necessary if we want to remove the non-extreme points. If the convex set is represented by means of linear restrictions, it is not a good idea to enumerate all the extreme points and then to calculate the marginalization. Direct algorithms to carry out the marginalization of a convex set given by linear restrictions are available and much more efficient [22]. The direct algorithm calculates all the extreme points of the marginalized set and not of the original convex set.

3 An Axiomatic View of Propagation Algorithms

In this section, we briefly describe the Shafer and Shenoy axiomatic framework for local computation [34, 36]. A valuation is a primitive concept meaning the mathematical representation for a piece of information in a given uncertainty theory. In our case, a valuation will be a convex set, which could represent an 'a priori' set of probability distributions, or a set of conditional probability distributions, or an 'a posteriori' conditional information.

We shall assume that for each $I \subseteq \{1, \ldots, n\}$ there is a set $\mathcal{V}_I$ of valuations defined on the Cartesian product, U_I. If $V \in \mathcal{V}_I$ we shall say that V is defined on U_I or that U_I is the frame of V. We shall also say that V is defined on I.

$\mathcal{V}$ will be the set of all valuations $\mathcal{V} = \cup_{I \subseteq \{1,\ldots,n\}} \mathcal{V}_I$

Two basic operations are necessary:

- *Marginalization.-* If $J \subseteq I$ and $V_1 \in \mathcal{V}_I$ then the marginalization of V_1 to J is a valuation $V_1^{\downarrow J}$ in $\mathcal{V}_J$.
- *Combination.-* If $V_1 \in \mathcal{V}_I$ and $V_2 \in \mathcal{V}_J$, then their combination is a valuation $V_1 \otimes V_2$ in $\mathcal{V}_{I \cup J}$

The following axioms are assumed to be verified by these operations:

Axiom 1 $V_1 \otimes V_2 = V_2 \otimes V_1$, $\quad (V_1 \otimes V_2) \otimes V_3 = V_1 \otimes (V_2 \otimes V_3)$.

Axiom 2 If $I \subseteq J \subseteq K$, and $V \in \mathcal{V}_K$, then $(V^{\downarrow J})^{\downarrow I} = V^{\downarrow I}$.

Axiom 3 If $V_1 \in \mathcal{V}_I$, $V_2 \in \mathcal{V}_J$, then $(V_1 \otimes V_2)^{\downarrow I} = V_1 \otimes {V_2}^{\downarrow (J \cap I)}$.

Then, local computation algorithms are developed and expressed in terms of these operations with valuations. In general, the problem they try to solve is the following: Let $R = \{V_1, \ldots, V_m\}$ be a set of valuations where each V_i is defined on $s(V_i) = I_i$. We are interested in the projection on U_j of the combination of all the valuations in R. That is, we want to calculate[34]:

$$R_j = \left(\bigotimes R\right)^{\downarrow \{j\}} = (V_1 \otimes \ldots \otimes V_m)^{\downarrow \{j\}}$$

for a value $j \in \{1, \ldots, n\}$.

The propagation algorithms use the following basic step (deletion of k) [34]:

- Let $k \in \{1, \ldots, n\}$, $k \neq j$. Consider $K = \{V_i \in R : k \in s(V_i)\}$ and $L = s(\bigotimes K) \Leftrightarrow \{k\}$. Then R is transformed into

$$R \Leftrightarrow K \cup \{(\bigotimes_{V \in K} V)^{\downarrow L}\} \tag{4}$$

This step is repeated deleting all k different from j. In that moment, all the valuations are defined on $\{j\}$ and the desired valuation, R_j, is the combination of all the valuations R in K.

This procedure is much more efficient than combining all the valuations and marginalizing afterwards and it is known as the deletion algorithm.

More sophisticated procedures have been developed. Most of them are based on organizing these calculations in graphical structures (join trees) [34, 23, 7, 30], to improve the efficiency of the computations, but all of them can be considered variations of this basic deletion algorithm.

4 Propagation under Unknown Interaction

The problem can be stated in the following way: we have an n-dimensional variable $(X_1, \ldots, X_n)$, each one of the X_i taking values on a finite set U_i.

We have m pieces of information, each one of them a convex set of probability distributions H_j about some of the variables in the problem, X_{I_j}, i.e. a convex set of probability distributions on U_{I_j}. Our objective will be to calculate the induced information about some variables of interest, X_J, which is given by:

$$(H_1 \cap H_2 \cap \ldots \cap H_m)^{\downarrow J} \tag{5}$$

Each H_j is given by a set of linear restrictions, R_j, which may represent bounds on events; or bounds in conditional probabilities; or in the expected value of any real valued function defined on U_{I_j}. See Hansen and Jaumard [18] for a detailed description of how this representation can be effectively carried out.

The objective is to know which are the possible probabilities induced by $H_1, \ldots, H_m$ in the set of variables X_J. In general, this is more general than calculating the bounds for the probability of an event $p(a_J)$. If H_J is calculated then, the bounds for $p(a_J)$ can be easily obtained by linear programming from the linear restrictions defining H_J. Even if we want to calculate the conditional probability $p(a_I|a_J)$, this can be done by calculating $H^{\downarrow I \cup J}$ and then by solving the corresponding fractional programming problems. However, we think that when we only want to calculated bounds on events or conditional events, it is more efficient to apply linear programming techniques to the original problem $H_1 \cap \ldots \cap H_m$, instead of calculating the marginalization in a previous step. There are linear programming algorithms based on the column generation technique able to cope with problems of bounds calculations with a large number of linear restrictions [18]. The algorithms in this section should be applied when our objective is to calculate the restrictions defining $H^{\downarrow J}$. In this case, if we further want to know the bounds for an event associated to U_J, we could apply the column generation technique to the restrictions defining $H^{\downarrow J}$.

The local computation algorithm is a consequence of verification of Shafer and Shenoy axioms for the convex sets marginalization and the intersection as combination operation. More details about the algorithm can be found in Verdegay [39].

The most important aspect in the implementation of these algorithms is how operations are carried out. If the convex sets are represented by linear restrictions, then combination has no problem: we only have to make the union of sets of restrictions. If we want to keep always a minimal set of restrictions, then we should remove redundant restrictions. However, in general, we think that the gaining in simplicity does not compensate for the cost of redundancy elimination operation. Anyway, we could apply an algorithm detecting some of the redundancies (it is not complete) but running in a very short time, i.e., we only reduce the simpler redundancies. An example of this type of algorithms is given by Imbert and Van Hentenryck [19]. The marginalization operation can be also implemented so that it does not generate redundant restrictions. Therefore, we can wait and not eliminate redundancy until a marginalization is carried out.

Marginalization is more difficult in terms of linear restrictions. Hansen and Jaumard [18] claim that this operation involves the enumeration of the vertices of a convex set which is a very time consuming operation. This is true but the marginalization can be done by enumerating essentially the points of the marginalized set, $H^{\downarrow I}$, instead of the original convex set H. This makes marginalization much more efficient because the number of extreme points of the marginalized set is much more smaller. An example of this type of algorithms can be obtained by applying the quantifier elimination technique by Lassez and Lassez [22].

4.1 Constraints Propagation

In this section we describe some procedures of propagating general knowledge based on the application of local rules. In general, in all these procedures it is not possible to propagate every type of restrictions, but only some particular types, usually bounds in the conditional probabilities. The proposed rules are always sound (the results are correct) but, in most of the cases, they are not complete (there is no guarantee that we obtain the optimal bounds). The most relevant work in this direction can be found in Amarger, Dubois, and Prade [1], Dubois et al. [12], Thöne [38], Lukasiewicz [25], and Salo [33].

Amarger, Dubois, and Prade [1], Thöne [38], and Lukasiewicz [25] consider propositional variables $\{A_1, \ldots, A_n\}$, that is taking only two possible values: true and false, and then rules which can be of the form $A \overset{x_1,x_2}{\Leftrightarrow} B$, with the meaning that $P(A) > 0$ and $0 \leq x_1 \leq P(B|A) \leq x_2 \leq 1$. It is also possible to work with bidirectional rules $A \overset{x_1,x_2}{\underset{y_1,y_2}{\longleftrightarrow}} B$ where $(x_2 = 0 \Leftrightarrow y_2 = 0)$ and the meaning that $A \overset{x_1,x_2}{\Leftrightarrow} B$ and $B \overset{y_1,y_2}{\Leftrightarrow} A$. They provide local rules to obtain new bounds for events of this type from given bounds. Dubois et al. [12] generalize these bounds to the case of linguistic probabilities (only a linguistic value is assigned to the probability of an event).

Salo [33] gives another local rules which are not always precise but they are given for more general cases of linear restrictions. Only two variables are considered, which will be called X and Y. First we have a set of

linear constraints

$$\sum_{i=1}^{n} a_i^k p(u_i) \leq \alpha_k, \quad k = 1, \ldots, K \tag{6}$$

representing a convex set H^X, the marginal information about X.

For each probability $p(v_j|u_i)$ for a fixed v_j, we have a convex set, $H^{Y=v_j|X}$, given by a set of linear restrictions:

$$\sum_{i=1}^{n} b_i^l p(v_j|u_i) \leq \beta_l, \quad l = 1, \ldots, L \tag{7}$$

Salo provides procedures to calculate the induced convex sets for the values of $p(u_i|v_j)$.

5 Propagation Algorithms under Strong Independence

In this case, the problem is as follows: we have a convex set of probabilities H about n-dimensional variable $(X_1, \ldots, X_n)$. It is assumed the existence of strong independence relationships represented by a directed acyclic graph G [30], allowing a decomposition of H:

$$H = H_1 \otimes \ldots \otimes H_n \tag{8}$$

where each H_i is a conditional convex set about variable X_i given its parents in the graph G: $X_{pa(i)}$.

This situation is the one that is obtained when we start with a probabilistic Bayesian network in which there is one probability distribution with stochastic independence relationships given by graph G, and we do not know in a precise way the values of the conditional probability distributions. The only thing that it is known is that they belong to the given convex sets. Furthermore, it is necessary to assume that we do not have more information about these probability distributions, in particular any joint information making additional restrictions about the unknown probability distributions. For example, in the case that we have intervals for the conditional probabilities, if we know that if $P(X_i = a_i|pa(X_i))$ is in the upper limit then so is $P(X_j = a_j|pa(X_j))$, then this situation is not directly representable in this model (it can be done, but adding more variables to the representation).

This convex set H represents the general knowledge about the problem. Then we have a particular case and an evidence or set of observations on it:

$$e = \{X_{i_1} = a_{i_1}^0, \ldots, X_{i_l} = a_{i_l}^0\} \tag{9}$$

and a variable of interest X_j. Our objective is to calculate the 'a posteriori' information marginalized on X_j. This 'a posteriori' information can be calculated as $H^j|e = (H|e)^{\downarrow j}$.

Under evidence $e = \{X_{i_1} = a_{i_1}^0, \ldots, X_{i_l} = a_{i_l}^0\}$, the 'a posteriori' information $H|e$ can be expressed as $H|e \propto (H_1 \otimes \ldots \otimes H_n) \otimes (\{l_{i_1}\} \otimes \ldots \otimes \{l_{i_l}\})$, where l_{i_j} is the likelihood associated to observation $X_{i_j} = a_{i_j}^0$, [7, 4]. In the sense that each probability in $H|e$ is proportional to a point in $(H_1 \otimes \ldots \otimes H_n) \otimes (\{l_{i_1}\} \otimes \ldots \otimes \{l_{i_l}\})$. So calculating this set is enough to obtain $H|e$. The marginal 'a posteriori' information for variable X_j can be expressed analogously as:

$$H^j|e \propto (H_1 \otimes \ldots \otimes H_n \otimes \{l_{i_1}\} \otimes \ldots \otimes \{l_{i_l}\})^{\downarrow j} \tag{10}$$

And this can be calculated by means of propagation algorithms, taking into account that $\otimes$ and marginalization verify the basic propagation axioms [7].

The main problem with convex sets of probabilities propagation is that if convex H_1 is given by m_1 extreme points and valuation H_2 is given by m_2 points, then we have to do $m_1 \times m_2$ pointwise multiplications of vectors. This may produce that if $H_1, \ldots, H_n$ are the convex sets with which we have specified the problem, and each valuation H_i has m_i points, then in the calculation of the 'a posteriori' information we may have $m_1 \times \cdots \times m_n$ points.

A first method to reduce the complexity is the use of convex hull algorithms to remove the non-extreme points after each operation of combination or marginalization. However, it is not convenient to apply these algorithms in spaces with a high dimension for the two following reasons:

- The complexity of the applying a convex hull algorithm increases a lot with the dimension of the space. For example, in the case of the gift wrapping algorithm with m points on a space of dimension k is $O(n^{[k/2]} \log(n))$, where $[k/2]$ is the integer part of $k/2$, [32].
- The number of non-extreme points decrease with the dimension of the space.

So, in most of the cases we should apply approximated or Monte-Carlo algorithms.

5.1 Approximate Algorithms

Cozman [9, 10] has given approximate methods based on gradient based search, Expectation-Maximization techniques or the Lavine's bracketing algorithm [24].

To explain them, first consider a transformation of our problem on an equivalent one [4, 5]. For each variable X_i, originally we give a conditional convex set H_i with extreme points $\{h_1, \ldots, h_l\}$. Then, we add a new node, T_i, with l cases $\{c_1, \ldots, c_l\}$ and that will be called the transparent node associated to variable X_i. This node is made a parent of variable X_i. On this node we consider that all the probability distributions are possible (that is the valuation for this node is a convex set with l extreme points, each one of them degenerated in one of the possible cases of T_i). If $pa(i)$ are the original parents of node X_i then the conditional probability of X_i given $pa(i) \cup T_i$ is determined in the following way: Given $T_i = c_k$ then the conditional probability of X_i given $pa(i)$ is h_k, i.e. $P(X_i = u_i | T_i = c_k, X_{pa(i)} = u_{p(i)}) = h_k(u_i, u_{pa(i)})$.

With this transformation, the structure of the problem does not change. The only thing that has been done is that our lack of knowledge about the conditional probabilities has been made explicit with the help of an additional node expressing all the the possible conditional probability distributions. Nothing is known about this node.

The main point about nodes T_i is that when we determine a value for each transparent node, then we have a completely specified probabilistic system and then a probabilistic propagation can be carried out.

Let θ_{ik} be the unknown probability $P(T_i = c_k)$. Call Θ the vector of all θ_{ik}. Fixed a value of Θ, then we have a completely specified probabilistic directed acyclic graph defining a global probability distribution P_Θ. Let X_j the variable we want to calculate the 'a posteriori' information and $u \in U_j$. Cozman [9, 10] expresses the problem of calculating the upper value for the probability of $X_j = u$ as an optimization problem:

$$\overline{P}(X_j = u|e) =$$

$$\max \{P_\theta(X_j = u|e) \ : \ \Theta \text{ is a vector of probabilities } \}$$

Cozman says that this problem is similar to the problem of estimating the vector of parameters Θ given evidence e. To apply gradient based techniques, he defines the log-likelihood for Θ as,

$$L(\Theta) = \log P_\Theta(X_j = u|e) =$$

$$\log P_\Theta(X_j = u, e) \Leftrightarrow \log P_\Theta(e)$$

The gradient of $L(\Theta)$ is obtained by computing,

$$\frac{\partial L(\Theta)}{\partial \theta_{ik}} = \frac{P_\Theta(T_i = c_k|X_j = u, e)}{\theta_{ik}} \Leftrightarrow \frac{P_\Theta(T_i = c_k|e)}{\theta_{ik}}$$

These values can be obtained by standard probabilistic propagation algorithms, and then a conjugate gradient descent can be constructed by selecting an initial value of Θ and, at each step, normalizing the values of Θ to ensure a proper distribution.

5.2 Simulation Algorithms

Here we briefly describe Monte-Carlo algorithms in Cano, Cano and Moral [3, 4], and Genetic Algorithms in Cano and Moral [5] to calculate the 'a posteriori' information.

All the different Monte-Carlo algorithms are based in selecting randomly one value for each transparent node, T_i, then making a probabilistic propagation, obtaining an 'a posteriori' vector for the interest variable of the problem. If we repeat this random selection and probabilistic propagation we obtain an approximation of the convex set of 'a posteriori' vectors. Cano, Cano, and Moral [3] consider several of them, carrying out an experimental evaluation.

Simulated annealing is an optimization technique to solve combinatorial optimization problems [31].

Our problem is to select a configuration of transparent nodes given rise to a minimum value of probability for a case of a given variable. The calculation of the maximum is completely analogous. If given $T_i = c_i, i = 1, \ldots, n$, and $C = (c_1, \ldots, c_n)$ this determines a probability P_C, and we want to calculate the upper (lower) interval for $X_j = u$, then we want to maximize (minimize) $P_C(X_j = u|e)$.

Cano, Cano and Moral [4] use the cooling procedure introduced by Kirkpatrick, Gelatt and Vecchi [21]. With this procedure, on each step the temperature is decreased according to the formula: $t_{i+1} = \alpha.t_i$, where α is a given constant.

To solve this problem, a triangulation of the original graph with the added transparent nodes is considered. Then a sequence of transparent nodes is chosen: $T_{r_1}, \ldots, T_{r_m}$, in such a way that two consecutive nodes of the pair given by the first and the last are never in two non-connected cliques: that is, they are in the same clique or in two connected ones. The sequence should also contain all the transparent nodes (some of them repeated). Let $K(i) = i + 1$ if $i \neq m$ and $K(m) = 1$.

In these conditions, we define the simulated annealing algorithm, according to the following terms: a configuration is given by a selection of nodes for the transparent variables, $T_j = c_{i_j}, j = 1, \ldots, n$, and a position of the transparent nodes sequence l. A neighboring configuration is given by an assignation in which the value of $T_{r_{K(l)}}$ can be modified and the current place

is $K(l)$. The resulting algorithm is based on a local computation of the optimum function from a neighboring node to another.

On a tree of cliques we have a double system of messages, according to the Shafer and Shenoy architecture [36, 35] allowing to calculate the optimum function in a local way: one system is to compute $P_C(X_j = u, e)$ and the other to compute $P_C(e)$.

In Cano, Cano, and Moral [4] some experimentation results with this algorithm are reported.

5.3 Genetic Algorithms

Cano and Moral [5] describe the use of a genetic algorithm to obtain points from the non-normalized 'a posteriori' convex set for variable X_j using an evolution program [27].

In the evolution program an individual or chromosome is a configuration of cases of transparent variables, $T_j = c_{i_j}, j = 1, \ldots, n$. Each individual determines a probability $P_{c_{i_1},\ldots,c_{i_n}}$.

The objective to maximize is $\frac{P_{c_{i_1},\ldots,c_{i_n}}(e, X_j = u_j)}{P_{c_{i_1},\ldots,c_{i_n}}(e)}$. This can be calculated by making one probabilistic propagation to variable X_j under probability distribution $P_{c_{i_1},\ldots,c_{i_n}}$ and evidence e.

Two genetic operators are used: crossover and mutation. Crossover takes two chromosomes $(c_{i_1,\ldots,c_{i_n}})$ and $(d_{i_1,\ldots,d_{i_n}})$, it chooses a random crossover point *pos* and then do the crossover generating chromosomes $(c_1, \ldots c_{pos}, d_{pos+1} \ldots d_n)$ and $(d_1, \ldots d_{pos}, c_{pos+1} \ldots c_n)$.

In the mutation stage for each chromosome and for each position in the chromosome we generate a random number r. If r is less than the probability of mutation then this position is mutated, selecting a different case for the corresponding transparent node.

A more detailed exposition of the algorithms and experimentation results can be found in [5].

5.4 The Case of Interval Probabilities

An important particular case that has received some attention in the literature is when we know an interval probability for each variable X_i given a configuration of its parents [2, 17, 37, 15]. The situation here is different of interval restrictions propagation (see Section 4.1) because here strong independence relationships represented by a directed acyclic graph are assumed.

For each variable X_i and each configuration of its parents $X_{pa(i)} = v$ we have a pair of elementary probability intervals $(\alpha(.|v), \beta(.|v))$. Above procedures can not be directly applied to this case. The problem is the following: assume that for each $v \in U_{pa(i)}$ we can find n_v extreme probabilities for variable X_i. Let us call the associated convex set $H^{X_{pa(i)}=v}$. Then, the set of possible conditional probability distributions for X_i is the convex set $H^{i|pa(i)}$ given by all the conditional probability distributions p about X_i given $X_{pa(i)}$ such that for each $v \in U_{pa(i)}$, we have that $p(.|v) \in H^{X_{pa(i)}=v}$. In other words, the possible global conditional probability distributions on X_i are all the conditional tables such that the row corresponding to v_j is a probability in the convex set $H^{X_{pa(i)}=v_j}$. This makes a number of global conditional extreme probabilities equal to $\prod_{v \in U_{pa(i)}} n_v$. But the number of elements of $U_{pa(i)}$ is $|U_{pa(i)}| = \prod_{k \in pa(i)} |U_k|$. This produces a combinatorial explosion in the extreme points of $H^{i|pa(i)}$ making infeasible the definition of a transparent variable T_i, and therefore the application of above procedures.

None of the approaches in the literature has considered this problem in all its generality. Tessem [37] has considered the cases of directed acyclic graphs with no loops (undirected cycles). On the other hand, Fertig and Brease [16, 2, 17] only consider the lower bounds $\alpha(.|v)$. Fagiuoli and Zaffalon [15] have developed exact algorithms for graphs without loops with binary variables which are linear in the size of the network. However, we think that further work is necessary to apply approximate algorithms to this problem in all its generality.

6 Conclusions

In this paper, we have given a complete overview about propagation algorithms for imprecise probabilities. The main contribution is the classification of the algorithms according to the underlying independence assumptions. This may help to clarify a field in which, at a first glance, it looks as if very different approaches have been designed for the same problem. What it is shown in this paper is that we have a very rich system of different problems (much more than in the single probabilistic case) and for each one of them, different techniques are used.

The situation from our point of view is the following: there are effective exact solutions for some particular cases of problems and approximate algorithms which have been developed for the general case and which can give good solutions in some complex problems. However we feel that we are far from having general tools able of providing good solutions for most of practical situations. In this sense, it would be convenient the definition of a set of benchmark problems that could be shared by the scientific community to

test new procedures and algorithms, and that could be considered as a common objective for forthcoming years.

References

[1] S. Amarger, D. Dubois, and H. Prade. Constraint propagation with imprecise conditional probabilities. In B.D. Ambrosio, Ph. Smets, and P.P. Bonissone, editors, *Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence*, pages 26–34. Morgan & Kaufmann, 1991.

[2] J.S. Breese and K.W. Fertig. Decision making with interval influence diagrams. In L.N. Kanal P.P. Bonissone, M. Henrion, editor, *Uncertainty in Artificial Intelligence, 6*, pages 467–478. Elsevier, 1991.

[3] A. Cano, J.E. Cano, and S. Moral. Simulation algorithms for convex sets of probabilities. Technical Report TR-93-2, DECSAI, Universidad de Granada, 1993.

[4] A. Cano, J.E. Cano, and S. Moral. Convex sets of probabilities propagation by simulated annealing. In *Proceedings of the Fith International Conference IPMU'94*, pages 4–8, Paris, 1994.

[5] A. Cano and S. Moral. A genetic algorithm to approximate convex sets of probabilities. In *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference (IPMU' 96) Vol. 2*, pages 859–864, 1996.

[6] J.E. Cano, M. Delgado, and S. Moral. An axiomatic system for the propagation of uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8:253–280, 1993.

[7] J.E. Cano, S. Moral, and J.F. Verdegay-López. Propagation of convex sets of probabilities in directed acyclic networks. In B. Bouchon-Meunier et al., editors, *Uncertainty in Intelligent Systems*, pages 15–26. Elsevier, 1993.

[8] I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. In *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications (ISIPTA'99)*, 1999.

[9] F. Cozman. Robust analysis of bayesian networks with finitely generated convex sets of distributions. Technical Report CMU-RI-TR96-41, Carnegie Mellon University, 1996.

[10] F. Cozman. Robustness analysis of bayesian networks with local convex sets of distributions. In D. Geiger and P.P. Shenoy, editors, *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 108–115. Morgan & Kaufmann, San Mateo, 1997.

[11] L.M. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In Ph. Besnard and S. Hanks, editors, *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 108–115. Morgan & Kaufmann, 1995.

[12] D. Dubois, L. Godo, R. López de Mántaras, and H. Prade. Qualitative reasoning with imprecise probabilities. *Journal of Intelligent Information Systems*, pages 319–363, 1993.

[13] D. Dubois and H. Prade. A survey of belief revision and updating rules in various uncertainty models. *International Journal of Intelligent Systems*, 9:61–100, 1994.

[14] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*. Springer Verlag, Berlin, 1987.

[15] E. Fagiuoli and M. Zaffalon. 2u: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106:77–107, 1998.

[16] K.W. Fertig and J.S. Breese. Interval influence diagrams. In M. Henrion, R.D. Shacter, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence, 5*, pages 149–161. North-Holland, Amsterdam, 1990.

[17] K.W. Fertig and J.S. Breese. Probability intervals over influence diagrams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:280–286, 1993.

[18] P. Hansen and B. Jaumard. Probabilistic satisfiability. To appear in: J. Kohlas y S. Moral, editors, *Handbook on Algorithms for Uncertain and Defeasible Management Systems*. Kluwer Academic Publishers, Dordrecht, 1999.

[19] J.L. Imbert and P. Van Hentenryck. Redundancy elimination with a lexicographic solved form. *Annals of Mathematics and Artificial Intelligence*, 17:85–106, 1996.

[20] H.H. Karwan, V. Lofti, J. Telgen, and S. Zionts. *Redundancy in Mathematical Programming: a State-of-the-Art Survey*. Lecture Notes in Economics and Mathematical Systems N. 206. Springer Verlag, 1983.

[21] S. Kirpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[22] C. Lassez and J.L. Lassez. Quantifier elimination for conjunctions of linear constraints via a convex hull algorithm. In B.R. Donald et al., editors, *Symbolic and Numerical Computation for Artificial Intelligence*, pages 103–119. Academic Press, London, 1992.

[23] S.L. Lauritzen and D.J. Spiegelhalter. Local computation with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Ser. B*, 50:157–224, 1988.

[24] M. Lavine. Sensitivity in bayesian statistics, the prior and the likelihood. *Journal of the American Statistical Association*, 86:396–399, 1991.

[25] Th. Lukasiewicz. *Precision of Probabilistic Deduction under Taxonomic Knowledge*. PhD thesis, Universität Tüningen, 1996.

[26] T.H. Matheiss and D.S. Rubin. A survey and comparison of methods for finding all vertices of convex polyedral sets. *Mathematics of Operational Research*, 5:167–185, 1980.

[27] Z. Michalewicz. *Artificial Intelligence. Genetic Algorithms + Data Structures=Evolution Programs*. Springer Verlag, 1992.

[28] S. Moral. Algorithms for imprecise probabilities. To appear in: J. Kohlas and S. Moral, editors, *Handbook of Algorithms for Defeasible and Uncertain Management System*. Kluwer Academic Publishers, Dordrecht, 1999.

[29] S. Moral and L.M. de Campos. Updating uncertain information. In B. Bouchon-Meunier, R.R. Yager, and L.A. Zadeh, editors, *Uncertainty in Knowledge Bases*, pages 58–67. Springer Verlag, Berlin, 1991.

[30] J. Pearl. *Probabilistic Reasoning with Intelligent Systems*. Morgan & Kaufman, San Mateo, 1988.

[31] P.J.M. Van Laarhoven P.J.M. and E.H.L. Aarts. *Simulated Annealing*. Reidel Publishing Company, Drodrecht, 1988.

[32] F.P. Preparata and M.I. Shamos. *Computational Geometry. An Introduction*. Springer Verlag, New York, 1985.

[33] A.A. Salo. Tighter estimates for the posteriors of imprecise prior and conditional probabilities. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 26:820–825, 1996.

[34] G. Shafer and P.P. Shenoy. Local computation in hypertrees. Working Paper N. 201, School of Business, University of Kansas, Lawrence, 1988.

[35] P.P. Shenoy. A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3:383–411, 1989.

[36] P.P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In Shachter et al., editors, *Uncertainty in Artificial Intelligence, 4*, pages 169–198. Elsevier, Amsterdam, 1990.

[37] B. Tessem. Interval probability propagation. *International Journal of Approximate Reasoning*, 7:95–120, 1992.

[38] H. Thöne, U. Güntzer, and W. Kießling. Towards precision of probabilistic bounds propagation. In D. Dubois, M.P. Wellman, B. D'Ambrosio, and Ph. Smets, editors, *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence*, pages 315–322, San Mateo, 1992. Morgan & Kaufmann.

[39] J.F. Verdegay-López. *Representaicón y Combinación de la Información con Incertidumbre mediante Convexos de Probabilidades*. PhD thesis, Universidad de Granada, 1997.