Penniless Propagation with Mixtures of Truncated Exponentials*

Rafael Rumí and Antonio Salmerón

Dept. Estadística y Matemática Aplicada,
Universidad de Almería, 04120 Almería, Spain
{rrumi, Antonio.Salmeron}@ual.es

Abstract. Mixtures of truncated exponential (MTE) networks are a powerful alternative to discretisation when working with hybrid Bayesian networks. One of the features of the MTE model is that standard propagation algorithm can be used. In this paper we propose an approximate propagation algorithm for MTE networks which is based on the Penniless propagation method already known for discrete variables. The performance of the proposed method is analysed in a series of experiments with random networks.

1 Introduction

A Bayesian network is an efficient representation of a joint probability distribution over a set of variables, where the network structure encodes the independence relations among the variables. Bayesian networks are commonly used to make inferences about the probability distribution on some variables of interest, given that the values of some other variables are known. This task is usually called *probabilistic inference* or *probability propagation*.

Much attention has been paid to probability propagation in networks where the variables are discrete with a finite number of possible values. Several exact methods have been proposed in the literature for this task [8, 13, 14, 20], all of them based on *local computation*. Local computation means to calculate the marginals without actually computing the joint distribution, and is described in terms of a message passing scheme over a structure called *join tree*. Also, approximate methods have been developed with the aim of dealing with complex networks [2, 3, 4, 7, 18, 19].

In mixed Bayesian networks, where both discrete and continuous variables appear simultaneously, it is possible to apply local computation schemes similar to those for discrete variables. However, the correctness of exact inference depends on the model.

This problem was deeply studied before, but the only general solution is the discretisation of the continuous variables [5, 11] which are then treated as if they

* This work has been supported by the Spanish Ministry of Science and Technology, project Elvira II (TIC2001-2973-C05-02) and by FEDER funds.

L. Godo (Ed.): ECSQARU 2005, LNAI 3571, pp. 39–50, 2005.

were discrete, and therefore the results obtained are approximate. Exact propagation can be carried out over mixed networks when the model is a conditional Gaussian distribution [12, 17], but in this case, discrete variables are not allowed to have continuous parents. This restriction was overcome in [10] using a mixture of exponentials to represent the distribution of discrete nodes with continuous parents, but the price to pay is that propagation cannot be carried out using exact algorithms: Monte Carlo methods are used instead.

The Mixture of Truncated Exponentials (MTE) model [15] provide the advantages of the traditional methods and the added feature that discrete variables with continuous parents are allowed. Exact standard propagation algorithms can be performed over them [6], as well as approximate methods. In this work, we introduce an approximate propagation algorithm for MTEs based on the idea of Penniless propagation [2], which is actually derived from the Shenoy-Shafer [20] method.

This paper continues with a description of the MTE model in section 2. The representation based on mixed tress can be found in section 3. Section 4 contains the application of Shenoy-Shafer algorithm to MTE networks, while in section 5 the Penniless algorithm is presented, and is illustrated with some experiments reported in section 6. The paper ends with conclusions in section 7.

2 The MTE Model

Throughout this paper, random variables will be denoted by capital letters, and their values by lowercase letters. In the multi-dimensional case, boldfaced characters will be used. The domain of the variable $\mathbf{X}$ is denoted by $\Omega_{\mathbf{X}}$. The MTE model is defined by its corresponding potential and density as follows [15]:

Definition 1. (MTE potential) *Let* $\mathbf{X}$ *be a mixed* n*-dimensional random vector. Let* $\mathbf{Y} = (Y_1, \ldots, Y_d)$ *and* $\mathbf{Z} = (Z_1, \ldots, Z_c)$ *be the discrete and continuous parts of* $\mathbf{X}$*, respectively, with* $c + d = n$*. We say that a function* $f : \Omega_{\mathbf{X}} \mapsto \mathbb{R}_0^+$ *is a* Mixture of Truncated Exponentials potential (MTE potential) *if one of the next conditions holds:*

i. $\mathbf{Y} = \emptyset$ *and* f *can be written as*

$$f(\mathbf{x}) = f(\mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{\sum_{j=1}^{c} b_i^{(j)} z_j\right\} \tag{1}$$

for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$*, where* a_i*,* $i = 0, \ldots, m$ *and* $b_i^{(j)}$*,* $i = 1, \ldots, m$*,* $j = 1, \ldots, c$ *are real numbers.*

ii. $\mathbf{Y} = \emptyset$ *and there is a partition* $D_1, \ldots, D_k$ *of* $\Omega_{\mathbf{Z}}$ *into hypercubes such that* f *is defined as*

$$f(\mathbf{x}) = f(\mathbf{z}) = f_i(\mathbf{z}) \quad \text{if} \quad \mathbf{z} \in D_i \enspace ,$$

where each f_i*,* $i = 1, \ldots, k$ *can be written in the form of (1).*

iii. $\mathbf{Y} \neq \emptyset$ *and for each fixed value* $\mathbf{y} \in \Omega_{\mathbf{Y}}$*,* $f_{\mathbf{y}}(\mathbf{z}) = f(\mathbf{y}, \mathbf{z})$ *can be defined as in* ii.

Definition 2. (MTE density) *An MTE potential f is an* MTE density *if*

$$\sum_{\mathbf{y}\in\Omega_{\mathbf{Y}}}\int_{\Omega_{\mathbf{Z}}} f(\mathbf{y},\mathbf{z})d\mathbf{z} = 1 \ .$$

In a Bayesian network, we find two types of densities:

1. For each variable X which is a root of the network, a density $f(x)$ is given.
2. For each variable X with parents $\mathbf{Y}$, a conditional density $f(x|\mathbf{y})$ is given.

A *conditional MTE density* $f(x|\mathbf{y})$ is an MTE potential $f(x,\mathbf{y})$ such that fixing $\mathbf{y}$ to each of its possible values, the resulting function is a density for X.

3 Mixed Trees

In [15] a data structure was proposed to represent MTE potentials: The so-called *mixed probability trees* or mixed trees for short. The formal definition is as follows:

Definition 3. (Mixed tree) *We say that a tree $\mathcal{T}$ is a* mixed tree *if it meets the following conditions:*

i. *Every internal node represents a random variable (either discrete or continuous).*
ii. *Every arc outgoing from a continuous variable Z is labeled with an interval of values of Z, so that the domain of Z is the union of the intervals corresponding to the arcs Z-outgoing.*
iii. *Every discrete variable has a number of outgoing arcs equal to its number of states.*
iv. *Each leaf node contains an MTE potential defined on variables in the path from the root to that leaf.*

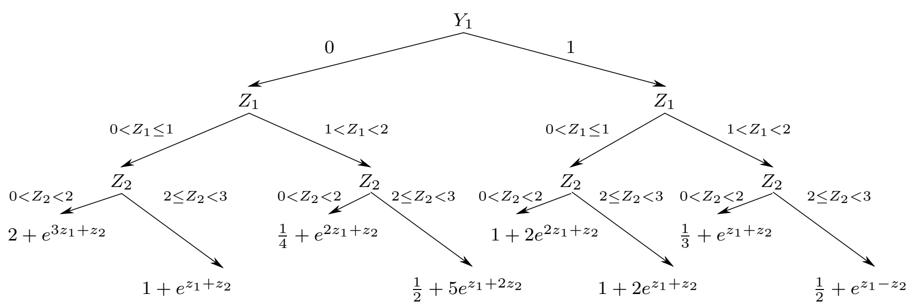

Fig. 1. A mixed probability tree representing an MTE potential

Mixed trees can represent MTE potentials defined by parts. Each entire branch in the tree determines one sub-region of the space where the potential is defined, and the function stored in the leaf of a branch is the definition of the potential in the corresponding sub-region. An example of a mixed tree is shown in Fig. 1.

The operations required for probability propagation in Bayesian networks (restriction, marginalisation and combination) can be carried out by means of algorithms very similar to those described, for instance in [11, 18].

4 Shenoy - Shafer Propagation Algorithm with MTEs

In [15] it was shown that MTE networks can be solved using Shenoy-Shafer algorithm [20]. This algorithm requires an adequate order of elimination of the variables to get the join tree, since different orders may result in join trees of distinct sizes, and the efficiency of probability propagation depends on the complexity of the join tree. This problem has been widely studied for discrete networks [1, 9], but not yet for MTE models. Here we propose a one-step lookahead strategy to determine the elimination order. We will choose the next variable to eliminate according to the size of the potential associated with the resulting clique.

Definition 4. (Size of an MTE potential) *The* size *of an MTE potential is defined as the number of exponentials terms, including the independent term, out of which the MTE potential is composed.*

Example 1. The potential represented in Fig. 1 has size equal to 16, because it has 8 leaves, and in each one an independent term, and one exponential term, so $8 \times (1 + 1) = 16$.

The decision on which variable to select next time, requires the knowledge about the size of the clique that would result from combining all the potentials defined for the variable. In the case of some MTE networks, it is possible to estimate it beforehand. If the MTE potentials are such that for each of them, the number of exponential terms in each leaf is the same, and the number of splits of the domain of the continuous variables also coincides, and only one variable appears in the MTE functions stored in the leaves of the mixed tree (the rest of the variables are used just to split the domain), as in [15] and [16], then there is an upper bound on the potential size:

Proposition 1. *Let $\mathcal{T}_1, \ldots, \mathcal{T}_h$ be h mixed probability trees, $\mathbf{Y}_i$, $\mathbf{Z}_i$ the discrete and continuous variables of each of them, and n_i the number of intervals into which the domain of the continuous variables of T_i is split. Let Ω_{Y_i} be the set of possible values of the discrete variable Y_i. The size of the tree $\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2 \times \ldots \times \mathcal{T}_h$ is lower than*

$$\left(\prod_{Y_i \in \bigcup_{i=1}^{h} \mathbf{Y}_i} |\Omega_{Y_i}| \right) \times \left(\prod_{j=1}^{h} n_j^{k_j} \right) \times \left(\prod_{j=1}^{h} t_j \right) ,$$

where t_j is the number of exponential terms in each leaf of $\mathcal{T}_j$, and k_j is the number of continuous variables in $\mathcal{T}_j$.

5 Penniless Propagation with MTEs

Using the algorithm cited above, it is usual in large discrete networks that the size of the potentials involved grow so much that the propagation becomes infeasible. In the case of MTE networks, the complexity is higher, since the potentials are larger in general.

To overcome this problem in the discrete case, the Penniless propagation algorithm was proposed [2]. This propagation method is based on the Shenoy-Shafer method, but modifying it so that the results are approximations of the actual marginal distributions in exchange of lower time and space requirements.

The Shenoy-Shafer algorithm operates over the join tree built from the original network using a message passing scheme between adjacent nodes. Between every pair of adjacent nodes C_i and C_j there is a mailbox for the messages from C_i to C_j and another one for the messages from C_j to C_i. Sending a message from C_i to C_j can be considered as transfering the information contained in C_i that is relevant to C_j. Messages stored in both mailboxes are potentials defined for $C_i \cap C_j$. Initially these mailboxes are empty and once a message is stored it is full. A node C_i is allowed to send a message to its neighbor C_j if and only if every mailbox for messages arriving to C_i is full except the one from C_j to C_i.

The propagation is organised in two steps: in the first one messages are sent from leaves to a previously selected root node, and in the second one the messages are sent from the root to the leaves.

The message from C_i to C_j is recursively defined as follows:

$$\phi_{C_i \to C_j} = \left\{ \phi_{C_i} \cdot \left(\prod_{C_k \in ne(C_i) \setminus \{C_j\}} \phi_{C_k \to C_i} \right) \right\}^{\downarrow C_i \cap C_j} , \qquad (2)$$

where ϕ_{C_i} is the original potential defined over C_i, $ne(C_i)$ is the set of adjacent nodes of C_i and superscript $\downarrow C_i \cap C_j$ indicates the marginal over $C_i \cap C_j$.

The main feature of the Penniless algorithm is that the messages sent are approximated, decreasing their size. This approximation [2, 4] is performed after every combination and marginalisation in (2), and also when obtaining the posterior marginals. It consists of reducing the size of the probability trees used to represent the potentials by *pruning* some of their branches (namely, those that are more similar). The same approach can be taken within the MTE framework, with the difference that this time, instead of probability trees, the potentials are represented as mixed trees. Let us consider now how the pruning operation can be carried out over mixed trees.

5.1 Pruning a Mixed Tree

The size of an MTE potential (and consequently the size of its corresponding mixed tree) is determined by the number of leaves it has and the number of exponential terms in each leaf. Thus, a way of decreasing the size of the MTE potentials is decreasing each one of these two quantities.

But every pruning has an error associated with it. This error will be measured in terms of divergence between the mixed trees before and after the pruning.

Definition 5. (Divergence between mixed trees) *Let $\mathcal{T}$ be a mixed tree representing an MTE potential ϕ defined for $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$. Let $\mathcal{T}^*$ be a subtree of $\mathcal{T}$ with root $Z \in \mathbf{Z}$ where every child of Z is an MTE potential. Let ϕ_1 be the potential represented by $\mathcal{T}^*$. Let $\mathcal{T}_P^*$ be a tree obtained from $\mathcal{T}^*$ replacing ϕ_1 by the potential ϕ_2 for which it holds that $\int_{\Omega_\mathbf{Z}} \phi_1 d\mathbf{z} = \int_{\Omega_\mathbf{Z}} \phi_2 d\mathbf{z}$. The* divergence *between $\mathcal{T}^*$ and $\mathcal{T}_P^*$ is defined as*

$$D(\mathcal{T}^*, \mathcal{T}_P^*) = E_{\phi_1^*}[(\phi_1^* - \phi_2^*)^2] = \int_{\Omega_\mathbf{Z}} \frac{\phi_1(\mathbf{z})}{\Delta}(\frac{\phi_1(\mathbf{z})}{\Delta} - \frac{\phi_2(\mathbf{z})}{\Delta})^2 d\mathbf{z},$$

where ϕ_i^* is the normalisation of ϕ_i and Δ is the total weight of ϕ:

$$\Delta = \sum_{\mathbf{Y}} \int_{\Omega_\mathbf{Z}} \phi(\mathbf{y}, \mathbf{z}) d\mathbf{z}.$$

We have considered three different kinds of pruning that are described in the next subsections.

Removing Exponential Terms. In each leaf of the mixed tree, the exponential terms that have little impact on the density function could be removed and the resulting potential would be rather similar to the original one.

Let $f(\mathbf{z}) = k + \sum_{i=1}^{n} a_i e^{b_i \mathbf{z}}$ be the potential stored in a leaf. The goal is to detect those exponential terms $a_i e^{b_i \mathbf{z}}$ having little influence on the entire density. We define the *weight* of each term as:

$$p_i = \int_{\Omega_\mathbf{Z}} a_i e^{b_i \mathbf{z}} d\mathbf{z}.$$

We think that two sensible criteria to remove terms in an MTE potential are the following:

1. A threshold α is established and the terms whose absolute weight, $|p_i|$, is lower than α are removed.
2. A maximum potential size is fixed and then the terms with lower absolute weight are removed until the size of the potential lies below the established maximum.

Once a term has been removed, the resulting potential is updated as follows :

- The maximum value of the term is computed , $m = \max_{\mathbf{z} \in \mathbf{Z}} \{a_i e^{b_i \mathbf{z}}\}$, and added to the independent term, $k^* = k + m$.
- The potential is normalised in order to make it integrate up to the total weight of the original potential.

The reason why the maximum of the potential is added to the independent term is to avoid negative points in the resulting potential.

Joining MTE Functions. Let $\mathcal{T}$ be a mixed tree whose root node, X, is continuous, and its children are MTE functions. The domain of X is divided into intervals, I_j, and for each of those intervals, a potential $f_j(\mathbf{z}) = k^j + \sum_{i=1}^{n} a_i^j e^{b_i^j \mathbf{z}}$ is defined. It may be that these potentials are very similar in the different intervals, I_j, and therefore some of them could be joined with little loss of information. Two intervals I_{j_1} and I_{j_2} are joined by replacing the potentials $f_{j_1}(\mathbf{z})$ and $f_{j_2}(\mathbf{z})$ by another potential $f(\mathbf{z})$, defined for over $I_{j_1} \cup I_{j_2}$.

We propose to compute $f(\mathbf{z})$ as follows. Let

$$p_{j_1} = \int_{\Omega_{\mathbf{Z}}} f_{j_1}(\mathbf{z}) d\mathbf{z} \quad and \quad p_{j_2} = \int_{\Omega_{\mathbf{Z}}} f_{j_2}(\mathbf{z}) d\mathbf{z}$$

be the weights of $f_{j_1}(\mathbf{z})$ and $f_{j_2}(\mathbf{z})$ respectively, the replacing function is proportional to

$$f(\mathbf{z}) = \frac{p_{j_1} f_{j_1}(\mathbf{z}) + p_{j_2} f_{j_2}(\mathbf{z})}{p_{j_1} + p_{j_2}} \ .$$

Since both functions must integrate up to the same quantity over $I_{j_1} \cup I_{j_2}$, a constant K must be found such that

$$\int_{\Omega_{\mathbf{Z}}} K f(\mathbf{z}) d\mathbf{z} = p_1 + p_2 \ ,$$

which implies that $K = \dfrac{p_1 + p_2}{\int_{\Omega_{\mathbf{Z}}} f(\mathbf{z}) d\mathbf{z}}$.

Let $\mathcal{T}$ be the tree corresponding to the original potential, and $\mathcal{T}_{\mathcal{P}}$ the one resulting from replacing $f_{j_1}(\mathbf{z})$ and $f_{j_2}(\mathbf{z})$ by $f(\mathbf{z})$, then the error $\mathcal{D}(\mathcal{T}, \mathcal{T}_{\mathcal{P}})$ is computed, and if it is lower than a fixed parameter, we replace $\mathcal{T}$ by $\mathcal{T}_{\mathcal{P}}$.

Discrete Pruning. In this particular MTE networks, the values of the discrete variables are used only when splitting the domain of the potential, so marginal potentials defined for discrete variables are equivalent to probability tables.

If Y is a discrete variable in a mixed tree node, and its children are MTE functions, then the tree can be pruned as defined in [18] (due to space limitations we do not provide the details here).

6 Experimental Evaluation of the Algorithm

In order to test the performance of the Penniless algorithm over MTE networks, we have carried out a simulation study, in which the algorithm is run over some MTE networks, using different levels of pruning.

Three different artificial networks have been created following these restrictions:

1. Given a variable, its number of parents is a Poisson distribution with mean 0.8 and its parents are chosen at random.

Table 1. Networks studied

Net	Number of nodes	Number of discrete nodes
Net1	42	3
Net2	77	8
Net3	86	11

Table 2. Probability distribution for the number of states of the discrete variables, the number of splits of the domain of continuous variables and the number of exponential terms of MTE functions

No. states	2	3	4
Probability	1/3	1/3	1/3

No. splits	1	2	3
Probability	0.2	0.4	0.4

No. exp. terms	0	1	2
Probability	0.05	0.75	0.20

2. Discrete variables:
   (a) Its number of states is simulated from the distribution showed in Table 2.
   (b) The probability value of each state is simulated from an Exponential distribution with mean 0.5.
3. Continuous variables:
   (a) The number of splits of the variable in a potential is simulated from the distribution showed in Table 2.
   (b) Every MTE potential has an independent term which is simulated from an Exponential distribution with mean 0.01 and a number of exponential terms determined by the distribution showed in Table 2.
   (c) In every exponential term, $a\exp\{bx\}$ the coefficient a is a real number following an Exponential distribution with mean 1, and the exponent b is a real number determined by a standard Normal distribution (mean 0 and standard deviation 1).

After simulating the parameters of the potentials, they are normalised in order to guarantee that the potentials are density functions.

For each network, the 30% of its variables are observed at random. The corresponding evidence is inserted in the network by restricting the potentials to the observed values.

The Penniless propagation is carried out over each of these networks, with different parameters of pruning. For discrete pruning and for joining intervals, some parameters are chosen, and the exponential terms in every potential are removed until there are only two terms remaining (i.e. the maximum number of terms per potential leaf in a mixed tree is set to 2).

Since the MTE framework is mainly an alternative to discretisation, the results of the propagation are compared with the results of applying Shenoy-Shafer propagation to the discretisation obtained by replacing every MTE function $f(\mathbf{z}) = k + \sum_{i=1}^{n} a_i e^{b_i \mathbf{z}}$ by a constant function $f^*(\mathbf{z}) = k^*$ so that

$$\int_{\Omega_{\mathbf{Z}}} f(\mathbf{z})d\mathbf{z} = \int_{\Omega_{\mathbf{Z}}} f^*(\mathbf{z})d\mathbf{z} \ .$$

After each propagation, the following quantites are computed:

1. The maximum size of the potential needed to compute the marginal distribution. It is achieved after combining all the messages sent to the clique that contains the variable in the join tree.
2. The error attached to it, according to definition 5.

For each network, the mean of these quantities is computed for all the variables that do not appear in the evidence. The summary of the obtained results are shown in Figs. 2 to 4, where the notation for the pruning parameters is shown in Table 3. The "Join parameter" is the maximum error allowed for joining two intervals, while the "Discrete parameter" indicates that discrete distributions that differ less than the value of the parameter with respect to a uniform distribution, in terms of entropy, are pruned. The foundations of this discrete parameter are explained in [18].

The results of the experiments show that the use of MTEs instead of discretisations provides more accurate results. It is not surprising, since the discretisation is just a particular case of the MTE framework (a discretised density is an

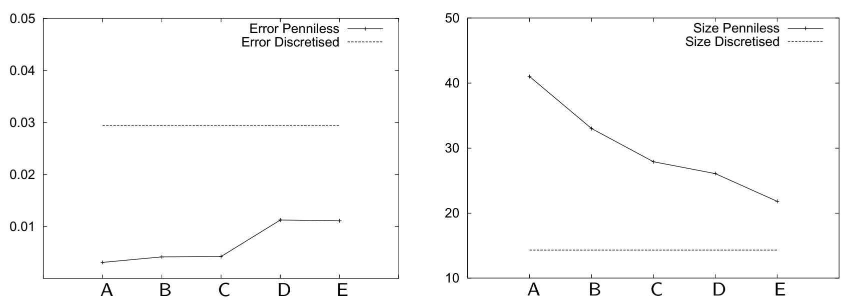

Fig. 2. Errors and sizes for Net1

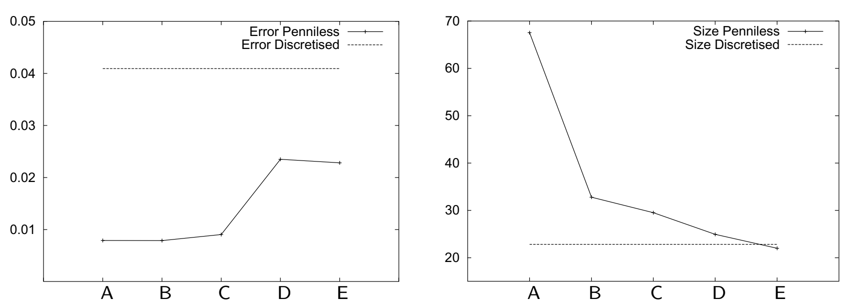

Fig. 3. Errors and sizes for Net2

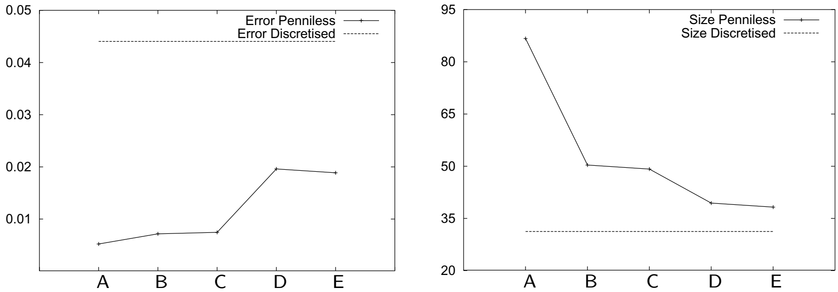

Fig. 4. Errors and sizes for Net3

Table 3. Different pruning parameters evaluated

Prune	Join parameter	Discrete parameter
A	0	0
B	0.005	0
C	0.005	0.01
D	0.05	0
E	0.05	0.01

MTE density with one independent term an zero exponential terms). However, it is important to point out that the increase in space required by the MTEs is significantly lower than the gain in accuracy, which means that the tradeoff space/accuracy, according to the evidence provided by the experiments reported here, is favourable to the MTE.

7 Conclusions

Some propagation methods have been successfully applied to MTE networks, as for example Shenoy-Shafer propagation [6], but so far they were not able to overcome the problem of the exponential increase of the sizes of the potentials involved in the propagation, specially when evidence is entered. In this paper we have presented a method to apply Penniless propagation to MTE networks, so that the sizes of the potentials are reduced because of the pruning operation.

The performance of the method has been tested on three artificial networks. The results of the experiments suggest that the Penniless algorithm is appropriate for MTE models, since the tradeoff between space requirements and accuracy is better than the one obtained with the discretisation.

The ideas contained in this paper can be extended to other propagation methods, specially the Lazy propagation and the class of Importance Sampling propagation algorithms, since these methods can take advantage of the reduction of the sizes of the potentials after pruning.

References

1. A. Cano and S. Moral. Heuristic algorithms for the triangulation of graphs. In B. Bouchon-Meunier, R.R. Yager, and L. Zadeh, editors, *Advances in Intelligent Computing*, pages 98–107. Springer Verlag, 1995.
2. A. Cano, S. Moral, and A. Salmerón. Penniless propagation in join trees. *International Journal of Intelligent Systems*, 15:1027–1059, 2000.
3. A. Cano, S. Moral, and A. Salmerón. Lazy evaluation in Penniless propagation over join trees. *Networks*, 39:175–185, 2002.
4. A. Cano, S. Moral, and A. Salmerón. Novel strategies to approximate probability trees in Penniless propagation. *International Journal of Intelligent Systems*, 18:193–203, 2003.
5. A. Christofides, B. Tanyi, D. Whobrey, and N. Christofides. The optimal discretization of probability density functions. *Computational Statistics and Data Analysis*, 31:475 – 486, 1999.
6. B. Cobb, P. Shenoy, and R. Rumí. Approximating probability density functions with mixtures of truncated expoenntials. In *Proceedings of the Tenth International Conference IPMU'04*, Perugia (Italy), 2004.
7. F. Jensen and S.K. Andersen. Approximations in Bayesian belief universes for knowledge-based systems. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, pages 162–169, 1990.
8. F.V. Jensen, S.L. Lauritzen, and K.G. Olesen. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, 4:269–282, 1990.
9. U. Kjærulff. Optimal decomposition of probabilistic networks by simulated annealing. *Statistics and Computing*, 2:1–21, 1992.
10. D. Koller, U. Lerner, and D. Anguelov. A general algorithm for approximate inference and its application to hybrid Bayes nets. In K.B. Laskey and H. Prade, editors, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pages 324–333. Morgan & Kaufmann, 1999.
11. D. Kozlov and D. Koller. Nonuniform dynamic discretization in hybrid networks. In D. Geiger and P.P. Shenoy, editors, *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 302–313. Morgan & Kaufmann, 1997.
12. S.L. Lauritzen. Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87:1098–1108, 1992.
13. S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50:157–224, 1988.
14. A.L. Madsen and F.V. Jensen. Lazy propagation: a junction tree inference algorithm based on lazy evaluation. *Artificial Intelligence*, 113:203–245, 1999.
15. S. Moral, R. Rumí, and A. Salmerón. Mixtures of truncated exponentials in hybrid Bayesian networks. In *Lecture Notes in Artificial Intelligence*, volume 2143, pages 135–143, 2001.
16. S. Moral, R. Rumí, and A. Salmerón. Estimating mixtures of truncated exponentials from data. In *Proceedings of the First European Workshop on Probabilistic Graphical Models*, pages 156–167, 2002.
17. K.G. Olesen. Causal probabilistic networks with both discrete and continuous variables. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:275–279, 1993.

18. A. Salmerón, A. Cano, and S. Moral. Importance sampling in Bayesian networks using probability trees. *Computational Statistics and Data Analysis*, 34:387–413, 2000.
19. E. Santos, S.E. Shimony, and E. Williams. Hybrid algorithms for approximate belief updating in Bayes nets. *International Journal of Approximate Reasoning*, 17:191–216, 1997.
20. P.P. Shenoy and G. Shafer. Axioms for probability and belief function propagation. In R.D. Shachter, T.S. Levitt, J.F. Lemmer, and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, pages 169–198. North Holland, Amsterdam, 1990.