Inference in hybrid Bayesian networks with Mixtures of Truncated Basis Functions

Helge Langseth
Department of Computer and Information Science
The Norwegian University of Science and Technology
Trondheim (Norway)
helgel@idi.ntnu.no

Thomas Dyhre Nielsen
Department of Computer Science
Aalborg University
Aalborg (Denmark)
tdn@cs.aau.dk

Rafael Rumí
Department of Statistics and Applied Mathematics
University of Almería, Almería (Spain)
rrumi@ual.es

Antonio Salmerón
Department of Statistics and Applied Mathematics
University of Almería, Almería (Spain)
antonio.salmeron@ual.es

Abstract

In this paper we study the problem of exact inference in hybrid Bayesian networks using mixtures of truncated basis functions (MoTBFs). We propose a structure for handling probability potentials called Sum-Product factorized potentials, and show how these potentials facilitate efficient inference based on *i*) properties of the MoTBFs and *ii*) ideas similar to the ones underlying Lazy propagation (postponing operations and keeping factorized representations of the potentials). We report on preliminary experiments demonstrating the efficiency of the proposed method in comparison with existing algorithms.

1 Introduction

Inference in hybrid Bayesian networks has received considerable attention over the last decade. In order to perform exact propagation in hybrid domains, the main challenge is to find a representation of the joint distribution that supports an efficient implementation of the usual inference operators: marginalization, and combination.

If the joint distribution belongs to e.g. the class of *conditional Gaussian* distributions (Lauritzen, 1992; Lauritzen and Jensen, 2001; Olesen, 1993), then inference can be performed exactly. However for this model class the continuous variables are assumed to follow a linear Gaussian distribution and the discrete variables are not allowed to have continuous parents. The mixture of truncated exponentials (MTE) model (Moral et al., 2001) does not impose such restrictions, but instead allows continuous and discrete variables to be treated in a uniform fashion. Furthermore, the MTE model class supports both exact (Cobb et al., 2004) and approximate inference methods (Rumí and Salmerón, 2007). Recently, the mixtures of polynomials (MOPs) model has been proposed as an alternative to the MTE model (Shenoy and West, 2011); the MOP model shares the advantages of MTEs, but it also provides a more flexible way of handling deterministic relationships among variables (Shenoy, 2011).

A more general approach for representing hybrid Bayesian networks has been introduced by Langseth et al. (2012) in the form of the mixtures of truncated basis functions (MoTBFs) model. MoTBFs are based on general real-valued basis functions that includes exponential and polynomial functions as special cases. Langseth et al. (2012) also show that efficient algorithms can be devised for approximating arbitrary probability density functions using MoTBFs.

In this paper we explore the problem of exact inference in hybrid Bayesian networks, where the potentials are specified using MoTBFs. We propose an algorithm that exploits properties of the basis functions and makes use of the ideas behind *Lazy propagation* in discrete networks (Madsen and Jensen, 1999; Madsen, 2010), i.e., postponing operations as long as possible and keeping factorized representations of the probability potentials. Preliminary experimental results show that the proposed algorithm provides significant improvements in efficiency compared to existing inference procedures for MTEs.

2 Preliminaries

The MoTBF potential was proposed by Langseth et al. (2012) as an alternative to the MTE and the MOP models, for which the exponential and polynomial functions are replaced by the more abstract notion of a basis function.[1]

Definition 1. *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* $c + d = n$*. Let* $\Psi = \{\psi_i(\cdot)\}_{i=0}^{\infty}$ *with* $\psi_i : \mathbb{R} \to \mathbb{R}$ *define a collection of real basis functions. A function* $\hat{f} : \Omega_{\mathbf{X}} \to \mathbb{R}_0^+$ *is a* mixture of truncated basis functions *(MoTBF) potential of level* k *wrt.* Ψ *if one of the following two conditions holds:*

1. $\hat{f}$ *can be written as*

$$\hat{f}(\mathbf{x}) = \hat{f}(\mathbf{y}, \mathbf{z}) = \prod_{j=1}^{c} \sum_{i=0}^{k} a_{i,\mathbf{y}}^{(j)} \psi_i(z_j), \quad (1)$$

where $a_{i,\mathbf{y}}^{(j)}$ *are real numbers.*

2. *There is a partitioning* $\Omega_{\mathbf{X}}^1, \ldots, \Omega_{\mathbf{X}}^m$ *of* $\Omega_{\mathbf{X}}$ *for which the domain of the continuous variables,* $\Omega_{\mathbf{Z}}$*, is divided into hyper-cubes such that* $\hat{f}$ *is defined as*

$$\hat{f}(\mathbf{x}) = f^{(l)}(\mathbf{x}) \text{ if } \mathbf{x} \in \Omega_{\mathbf{X}}^l, \quad (2)$$

where each $f^{(l)}$*,* $l = 1, \ldots, m$*, can be written in the form of Equation* (1)*.*

An MoTBF potential is a density if $\sum_{\mathbf{y}\in\Omega_{\mathbf{Y}}} \int_{\Omega_{\mathbf{Z}}} \hat{f}(\mathbf{y}, \mathbf{z}) d\mathbf{z} = 1$.

[1] We give a definition wich is slightly altered compared to the orignal deinition by Langseth et al. (2012). This subtle difference is introduced to simplify the following deductions and will not have practical implications.

An MoTBF $f(\mathbf{z}_1, \mathbf{z}_2, \mathbf{y})$ defined over the continuous variables $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and the discrete variables $\mathbf{Y}$ is a *conditional MoTBF* for $\mathbf{Z}_1$ if $\int_{\mathbf{z}_1} f(\mathbf{z}_1, \mathbf{z}_2, \mathbf{y}) d\mathbf{z}_1 = 1$ for all $\mathbf{z}_2 \in \Omega_{\mathbf{Z}_2}$ and $\mathbf{y} \in \Omega_{\mathbf{Y}}$.[2] Following Langseth et al. (2012) we assume that the conditioning variables only affect the density through the hyper-cubes over which the density is defined. Thus, for variables $\mathbf{Z}_1$ and $\mathbf{Z}_2$ with $\Omega_{\mathbf{Z}_2}$ partitioned into $\Omega_{\mathbf{Z}_2}^1, \ldots, \Omega_{\mathbf{Z}_2}^m$ we define the conditional MoTBF density $f(\mathbf{z}_1|\mathbf{z}_2, \mathbf{y})$ as:

$$f(\mathbf{z}_1|\mathbf{z}_2, \mathbf{y}) = \prod_{j=1}^{c} \sum_{i=0}^{k} a_{i,\mathbf{y},l}^{(j)} \psi_i(\mathbf{z}_1^{(j)}), \quad (3)$$

for $\mathbf{z}_2 \in \Omega_{\mathbf{Z}_2}^l$. Consequently, given a partitioning of the conditioning variables $\mathbf{z}_2$, finding a conditional MoTBF $f(z_1|\mathbf{z}_2)$ for a single variable z_1 reduces to specifying a collection of univariate MoTBFs. By extension, specifying the distributions of a hybrid Bayesian network therefore involves univariate MoTBF potentials only, and the form of the MoTBF potentials simplifies to

$$\hat{f}(z_1|\mathbf{z}_2, \mathbf{y}) = \sum_{i=0}^{k} a_{i,\mathbf{y},l}^{(1)} \psi_i(z_1),$$

for $\mathbf{z}_2 \in \Omega_{\mathbf{Z}_2}^l$.

The approximation procedure described in (Langseth et al., 2012) also assumes that the basis functions Ψ are both *legal* and *orthonormal*: If $\mathcal{Q}$ is the set of all linear combinations of the members of a set of basis functions $\Psi = \{\psi_i(\cdot)\}_{i=0}^{\infty}$, then Ψ is said to be a *legal* set of basis functions if the following conditions hold:

- ψ_0 is constant in its argument.
- If $f \in \mathcal{Q}$ and $g \in \mathcal{Q}$, then $(f \cdot g) \in \mathcal{Q}$.
- For any pair of real numbers s and t, there exists a function $f \in \mathcal{Q}$ such that $f(s) \neq f(t)$.

Clearly, the sets of basis functions $\{x^i\}_{i=0}^{\infty}$ and $\{\exp(-i \cdot x), \exp(i \cdot x)\}_{i=0}^{\infty}$ are legal and correspond to examples of bases for the MOP and MTE frameworks, respectively.

[2] For ease of presentation we disregard possible partitionings of $\Omega_{\mathbf{Z}_1}$.

When considering orthonormal basis functions, we focus on the space $L^2[a,b]$ of quadratically integrable real functions over the finite interval $[a,b]$. For two functions $f(x)$ and $g(x)$ defined on $[a,b]$ we define the inner product as

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx,$$

and say that two functions are orthonormal if and only if $\langle f, g\rangle = 0$ and $\langle f, f\rangle = \langle g, g\rangle = 1$. A set of non-orthonormal basis functions can easily be orthonormalized using, for instance, the Gram-Schmidt procedure.

In this paper we will often refer to MoTBF potentials defined only over continuous variables. In such cases, we understand, unless specified otherwise, that all the claims about such potentials are extensible to those potentials also containing discrete variables in their domains, simply by having the claims hold for each configuration of the discrete variables. Furthermore, in the remainder of the paper we assume a fixed set of m basis functions for each continuous variable in the network, i.e., $\Psi = \{\psi_0, \psi_1, \ldots, \psi_{m-1}\}$.[3]

3 Operations over MoTBFs

There are three operations used by exact inference algorithms: restriction, combination, and marginalization. The first operation is trivial and is basically used to incorporate evidence prior to the inference process, while the others are used throughout the inference process.

Definition 2 (Combination). *Let $f_1(\mathbf{y}_1, \mathbf{z}_1)$ and $f_2(\mathbf{y}_2, \mathbf{z}_2)$ be MoTBF potentials defined over the partitions $\mathcal{P}_1 = \{\Omega^1_{\mathbf{Z}_1}, \ldots, \Omega^{k_1}_{\mathbf{Z}_1}\}$ and $\mathcal{P}_2 = \{\Omega^1_{\mathbf{Z}_2}, \ldots, \Omega^{k_2}_{\mathbf{Z}_2}\}$ of $\Omega_{\mathbf{Z}_1}$ and $\Omega_{\mathbf{Z}_2}$, respectively:*

$$f_h(\mathbf{y}_h, \mathbf{z}_h) = \prod_{j=1}^{c_h} \sum_{i=0}^{m-1} a^{(j)}_{i,\mathbf{y},l_h}\, \psi_i(z^{(j)}_h),$$

for $\mathbf{z}_h \in \Omega^{l_h}_{\mathbf{Z}_h}$ ($\Omega^{l_h}_{\mathbf{Z}_h} \in \mathcal{P}_h$) and $h = 1, 2$. The combination *of f_1 and f_2 is a potential $f(\mathbf{y}, \mathbf{z})$ over $\mathbf{Y} = \mathbf{Y}_1 \cup \mathbf{Y}_2$ and $\mathbf{Z} = \mathbf{Z}_1 \cup \mathbf{Z}_2$:*

$$\begin{aligned} f(\mathbf{y}, \mathbf{z}) = \prod_{j=1}^{c_1}\prod_{r=1}^{c_2} &\Big(\sum_{i=0}^{m-1} a^{(j)}_{i,\mathbf{y}_1,l_1}\, \psi_i(\mathbf{z}^{(j)}_1)\Big) \\ &\Big(\sum_{i=0}^{m-1} a^{(r)}_{i,\mathbf{y}_2,l_2}\, \psi_i(\mathbf{z}^{(r)}_2)\Big) \end{aligned} \tag{4}$$

for all $\mathbf{z} \in \Omega^{l_1}_{\mathbf{Z}_1} \times \Omega^{l_2}_{\mathbf{Z}_2}$ and $\mathbf{y} \in \Omega^{l_1}_{\mathbf{Y}_1} \times \Omega^{l_2}_{\mathbf{Y}_2}$.

Observe that each factor in Equation (4) is an MoTBF. If the products in Equation (4) are not expanded any further, we call the operation *lazy combination*, pointing out the fact that no actual product is carried out; instead the factors in the original potentials are concatenated in a single list.

Definition 3 (Factorized potential). *A potential defined over hyper-cubes, and where in each hyper-cube the potential is defined as a list of factors of the form given in Equation (4) is called a* factorized potential.

Generalizing the notion of lazy combination to factorized potentials follows immediately.

Definition 4 (Marginalization of factorized potentials). *Let $f_{\mathbf{Z}}$ be a factorized potential defined for variables $\mathbf{Z} = \{Z_1, \ldots, Z_c\}$ over hyper-cubes $\cup_{h=1}^k \Omega^h_{\mathbf{Z}}$. The result of* marginalizing out *a variable Z_j from $f_{\mathbf{Z}}$ is a new potential defined on $\mathbf{Z}\setminus\{Z_j\}$ over the hyper-cubes $\cup_{h=1}^{k'} \Omega^h_{\mathbf{Z}\setminus\{Z_j\}}$, where for each new hyper-cube $\Omega^h_{\mathbf{Z}\setminus\{Z_j\}}$, $h = 1, \ldots, k'$ (obtained by projecting $\Omega^h_{\mathbf{Z}}$ onto $\mathbf{Z}\setminus\{Z_j\}$) the marginalized potential is defined as*

$$\begin{aligned} &f_{\mathbf{Z}\setminus\{Z_j\}}(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_c) \\ &\qquad = \sum_{l=1}^{r} \int_{\Omega_{Z_j}} f^{(h,l)}(z_1, \ldots, z_c)dz_j, \end{aligned}$$

where $\Omega^1_{Z_j} \cup \cdots \cup \Omega^r_{Z_j}$ is the partition of the domain of Z_j in $f_{\mathbf{Z}}$ and $f^{(h,l)}$ denotes the value of potential $f_{\mathbf{Z}}$ in hyper-cube $\Omega^h_{\mathbf{Z}\setminus\{Z_j\}} \times \Omega^l_{Z_j}$.

Proposition 1. *Let $f_{\mathbf{Z}}$ be a factorized potential under the same conditions as in Definition 4. Then, for each hyper-cube $\Omega^h_{\mathbf{Z}\setminus\{Z_j\}}$,*

[3]Note that any potential defined over a subset $\Psi' \subseteq \Psi$ can be represented by setting the coefficients of the potentials in $\Psi \setminus \Psi'$ to 0.

$h = 1, \ldots, k'$,

$$f_{\mathbf{Z}\setminus\{Z_j\}}(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_c) = \sum_{l=1}^{r} \left(\prod_{i \neq j} \sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(i)} \psi_s(z_i) \right).$$

Proof. By expanding the integral in Def. 4 we get

$$\begin{aligned}
&\int f^{(h,l)}(z_1, \ldots, z_c) dz_j \\
&= \int \prod_{i=1}^{c} (\sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(i)} \psi_s(z_i)) dz_j \\
&= \prod_{i \neq j} (\sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(i)} \psi_s(z_i)) \int \sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(j)} \psi_s(z_j) dz_j \\
&= \prod_{i \neq j} (\sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(i)} \psi_s(z_i)) (\sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(j)} \int \psi_s(z_j) dz_j).
\end{aligned}$$

Since the basis functions ψ_s, $s = 0, \ldots, m-1$ are orthonormal, it holds that $\int_{\Omega_{Z_j}} \psi_{s_1}(z_j)\psi_{s_2}(z_j) dz_j = 0$ for any $s_1 \neq s_2 \in \{0, \ldots, m-1\}$. Moreover, taking into account that $\psi_0(z_j)$ is a constant, it follows that $\int_{\Omega_{Z_j}} \psi_s(z_j) dz_j = 0$ for any $s > 0$. Furthermore, if $s = 0$ we have that

$$a_{0,\cdot,(h,l)}^{(j)} \int_{\Omega_{Z_j}} \psi_0(z_j) dz_j = 1.$$

Hence,

$$\int_{\Omega_{Z_j}} f^{(h,l)}(z_1, \ldots, z_c) dz_j = \prod_{i \neq j} \left(\sum_{s=0}^{m-1} a_{s,\cdot,(h,l)}^{(i)} \psi_s(z_i) \right).$$

□

From Prop. 1 we see that the result of marginalizing out a variable from a factorized potential is not necessarily a factorized potential, but rather a sum of factorized potentials. We therefore need to extend the concept of factorized potentials in order to allow factorized representations with respect to sums and products.

Definition 5 (SP factorized potential)**.** *Let $f_{\mathbf{Z}}$ be a potential defined for variables $\mathbf{Z} = \{Z_1, \ldots, Z_c\}$ over hyper-cubes $\cup_{i=1}^{k} \Omega_{\mathbf{Z}}^{i}$. We say that $f_{\mathbf{Z}}$ is a* Sum-Product (SP) factorized potential *if it can be written as*

$$f_{\mathbf{Z}}(\mathbf{z}) = \sum_{j=1}^{t} f_{\mathbf{Z}}^{(j)}(\mathbf{z}), \tag{5}$$

where $t > 0$ and $f^{(j)}$, $j = 1, \ldots, t$, are factorized potentials according to Definition 3.

Corollary 1. *The result of marginalizing out a variable from a factorized potential is an SP factorized potential.*

Proof. It follows directly from the proof of Proposition 1. □

Definition 6 (Combination of SP factorized potentials)**.** *Let*

$$f_{\mathbf{X}_i}(\mathbf{x}_i) = \sum_{l=1}^{r_i} f_{\mathbf{X}_i}^{(l)}(\mathbf{x}_i) \quad \textit{for } i = 1, 2$$

be two SP factorized potentials over variables $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. The combination of $f_{\mathbf{X}_1}$ and $f_{\mathbf{X}_2}$ is a new potential over variables $\mathbf{X}_{1,2} = \mathbf{X}_1 \cup \mathbf{X}_2$ defined as

$$f_{\mathbf{X}_{1,2}}(\mathbf{x}_{1,2}) = \sum_{l=1}^{r_1} \sum_{m=1}^{r_2} f_{\mathbf{X}_1}^{(l)}(\mathbf{x}_1) f_{\mathbf{X}_2}^{(m)}(\mathbf{x}_2). \tag{6}$$

Proposition 2. *The combination of two SP factorized potentials is another SP factorized potential. That is, the class of SP factorized potentials is closed under combination.*

Proof. Notice that each summand in Equation (6) is a product of two factorized potentials, which is itself a factorized potential. Therefore, the result of the combination is a sum of factorized potentials, and that is, by definition, an SP factorized potential. □

Definition 7 (Marginalisation of SP factorized potentials)**.** *Let*

$$f_{\mathbf{Z}}(\mathbf{z}) = \sum_{l=1}^{r} f_{\mathbf{Z}}^{(l)}(\mathbf{z})$$

be an SP factorized potential defined for variables $\mathbf{Z} = \{Z_1,\ldots,Z_c\}$ *over hyper-cubes* $\cup_{h=1}^{k}\Omega_{\mathbf{Z}}^{h}$. *The result of* marginalizing out *a variable* Z_j *from* $f_{\mathbf{Z}}$ *is a new potential defined on* $\mathbf{Z}\setminus\{Z_j\}$ *over the hyper-cubes* $\cup_{h=1}^{k'}\Omega_{\mathbf{Z}\setminus\{Z_j\}}^{h}$:

$$f(z_1,\ldots,z_{j-1},z_{j+1},\ldots,z_c) = \sum_{l=1}^{r} f_{\mathbf{Z}\setminus\{Z_j\}}^{(l)}(z_1,\ldots,z_{j-1},z_{j+1},\ldots,z_n),$$

where $f_{\mathbf{Z}\setminus\{Z_j\}}^{(l)}$, $l=1,\ldots,r$, *are computed according to Definition 4.*

Proposition 3. *The class of SP factorized potentials is closed under marginalization.*

Proof. As an SP factorized potential is a sum of factorized potentials, marginalizing out one variable consists of marginalizing it out in each of the factorized potentials where the variable appears. From Prop. 1 we know that the result of marginalizing out one variable from a factorized potential is an SP factorized potential. □

4 Inference in BNs with MoTBFs

Consider an MoTBF model with discrete variables $\mathbf{Y} = \{Y_1,\ldots,Y_d\}$ and continuous variables $\mathbf{Z} = \{Z_1,\ldots,Z_c\}$. Probabilistic inference in such a BN can be carried out using standard propagation algorithms that rely on sum and product operations, as the SP factorized MoTBF potentials are closed under product and marginalization.

For ease of presentation and analysis of the results, we formulate the inference process for MoTBFs using the classical variable elimination algorithm. This algorithm is designed to compute the posterior distribution over a target variable $W \in \mathbf{X}$. It is based on sequentially eliminating the variables in $(\mathbf{X}\setminus\mathbf{E})\setminus\{W\}$ (according to the minimum size heuristic with one step look ahead) from the potentials containing them. The elimination of a variable from a set of potentials is carried out by *i*) combining the potentials containing the variable and *ii*) marginalizing out that variable from the result of the combination. If not stated otherwise, we shall assume that if two variables X_1 and X_2 have the same parent Z, then Ω_Z is partitioned identically in the specification of the two MoTBF potentials for X_1 and X_2.

The complexity of the inference process is determined by the size of the potentials constructed during inference. The next proposition gives the size (the number of factors) resulting from combining a set of potentials, according to the combination operation described in Def. 6.

Proposition 4. *Let* $f^{(1)},\ldots,f^{(h)}$ *be h SP factorized potentials, and let* $\mathbf{Y}_i$, $\mathbf{Z}_i$, $i=1,\ldots,h$, *be the discrete and continuous variables of each of them. Let* $\mathbf{Z} = \cup_{i=1}^{h}\mathbf{Z}_i = \{Z_1,\ldots,Z_j\}$, *and let* n_l, $l=1,\ldots,j$, *be the number of intervals into which the domain of* Z_l *is split. If* Ω_{Y_i} *is the set of possible values of the discrete variable* Y_i, *then it holds that*

$$\textit{size}(f^{(1)}\cdots f^{(h)}) \le \prod_{Y_i\in\overset{h}{\underset{i=1}{\cup}}\mathbf{Y}_i} |\Omega_{Y_i}| \prod_{l=1}^{j} n_l \prod_{l=1}^{h} s_l \sum_{l=1}^{h} t_l, \tag{7}$$

where s_l *and* t_l, $l=1,\ldots,h$ *are, respectively, the maximum number of summands and the maximum number of factors in each summand, in potential* $f^{(l)}$ *(see Definition 5).*

Proof. The first factor is justified by the fact that for each possible configuration of the discrete variables, there is an MoTBFs parameterization corresponding to the continuous variables. The second factor comes from the definition of combination in Definition 2, and the fact that the set of possible split points is fixed for each variable. Finally, the third and fourth factors also follow directly from Definition 6. □

Proposition 4 gives an upper bound on the size of the SP factorized potentials resulting from a combination of factors. We see a significant reduction in size with respect to the bound given in (Rumí and Salmerón, 2007, Proposition 6) for the particular case of MTEs:

$$\text{size}(f^{(1)}\cdots f^{(h)}) \le \prod_{Y_i\in\overset{h}{\underset{i=1}{\cup}}\mathbf{Y}_i} |\Omega_{Y_i}| \prod_{l=1}^{j} n_l^{k_l} \prod_{l=1}^{h} t_l, \tag{8}$$

where k_l, $l = 1, \ldots, h$, is the number of continuous variables in $f^{(l)}$ and in this case, t_l, $l = 1, \ldots, h$ is the number of exponential terms of the potential in each hyper-cube of $f^{(l)}$. The differences between Equations (7) and (8) lie in the last three and two terms respectively. The difference in the second terms implies a significant reduction in size, and is based on the fact that we assume that the points in which the domain of each variable is split is selected from a fixed set of points. Under such an assumption, the number of intervals involved in the domain of a variable that appears in several potentials being combined, never increases after carrying out a combination. Note that if a potential represents a conditional MoTBF, only the domain of the conditioning variables is split and thus the corresponding term in the product would be equal to 1.

Regarding the third and forth terms, in Equation (7), they give the number of factors or univariate MoTBFs stored in each resulting hyper-cube, while in Equation (8), the third factor gives the number of exponential terms in each resulting hyper-cube. Therefore, as we are assuming that the number of summands, m, is the same for every variable, the third factor in Equation (8) could be replaced by m^h, while the number of summands reported by Equation (7) would be $(\prod_{l=1}^{h} s_l)(\sum_{l=1}^{h} t_l)m$, which grows more slowly than m^h, except for trivial cases. For instance, combining two univariate MoTBFs, would result in a potential with $2m$ summands, while the same operation with traditional combination of MTEs would yield a potential with m^2 exponential terms.

5 Experimental evaluation

In order to evaluate the proposed MoTBF-based inference procedure, we used the Variable Elimination algorithm for doing inference in a set of randomly generated networks where the number of continuous variables ranges from 4 to 24 (with increments of 2). The goal is to measure the increase in efficiency obtained by the MoTBF approach in comparison to inference using classical combination and marginalization of MTEs (Moral et al., 2001). That is, we want to measure the impact, in terms of efficiency, of using SP factorized potentials and their corresponding operations instead of traditional potentials and their operations. Note that all the methods considered here carry out exact inference, i.e. they obtain exactly the same marginals, and therefore the only differences are in terms of efficiency. The results can be extrapolated to other formalisms compatible with the MoTBF framework, like MOPs (Shenoy and West, 2011). One advantage of choosing the MTE model as strawman framework in the experiments is the availability of the Elvira software (Elvira Consortium, 2002) that implements the MTE model. As the novelties of the approach proposed in this paper concentrate on the handling of continuous variables, we have not included discrete variables in the networks.

The networks have been constructed by randomly assigning a number of parents to each variable, following a Poisson distribution with mean 2. In order to guarantee an increase in network complexity as the number of variables grows, each new network is built by adding variables to the previous one. For instance, the network with 6 variables is obtained by randomly adding 2 variables to the already generated network with 4 variables, and so on. We assumed that in each hypercube of every conditional distribution the domain of the child variable is not split (see Eq. (3)). Also, all the univariate potentials contained in the networks correspond to 1-piece and 7-terms MoTBF approximations of the standard Gaussian density (within the range $[-2.5, 2.5]$) found using the procedure described in (Langseth et al., 2012) based on exponential basis functions. The domains of the variables were split in two parts, using the mid point of the interval and keeping the split point of the variable fixed for all distributions in which they appear.[4] This also ensures that no new hypercubes are constructed by the combination operator, thereby reducing complexity during inference.

[4]For the MoTBF framework this only pertains to the conditioning variables, since the domain of a head variable is not partitioned.

#vars	M-vars	A-vars	VE[MoTBF]	VE[MTE]
4	4	2.7	0.0084	2.730
6	5	3.1	0.0094	300.909
8	5	3.2	0.0048	342.770
10	5	3.3	0.0050	363.877
12	6	3.7	0.0218	—
14	9	4.1	0.0482	—
16	11	4.7	0.2129	—
18	12	5.0	0.5462	—
20	12	5.0	0.9508	—
22	14	5.6	3.4464	—
24	14	5.7	9.1388	—

Table 1: Average run time per variable (in seconds) of the variable elimination algorithm using the MoTBF approach (VE[MoTBF]) and the same algorithm using classical combination and marginalization (VE[MTE]). #vars indicates the number of variables in the network, M-vars is the maximum number of variables in the potentials used during the variable elimination algorithm, and A-vars is the average number of variables in the above mentioned potentials.

The results are shown in Table 1, where the average time used to run the variable elimination algorithm for each variable in the network is displayed. Empty cells in the table indicate that the algorithm ran out of memory during inference in the corresponding network (2.5GB allocated). The label VE[MoTBF] refers to the proposed variable elimination algorithm using the MoTBF approach (based on SP factorized potentials), while the label VE[MTE] indicates the same algorithm, but with traditional combination and marginalization. In all the tested cases, the proposed algorithm provides a significant improvement in efficiency, which is more evident as the number of variables increases. In fact, VE[MTE] is not able to obtain any results for networks with more than 10 variables.

With the experiment described above, we have illustrated the increase in efficiency when doing MoTBF-based inference as compared to the MTE approach. However, the use of SP-factorized potentials also carry over to the traditional MTE framework; this entails lifting the restriction of having a fixed set of possible split points for the variables, allowing split points in the domains of child variables in conditional distributions, and not being able to exploit the properties of the basis functions. In order to analyze this modified inference algorithm, we designed a second experiment in which the split points for the variables in each potential were chosen at random. Also, instead of using MTE approximations of the Gaussian density, we used randomly generated MTEs with 10 exponential terms. The results of this experiment are displayed in Table 2, where we see a significant increase in efficiency for all the network sizes, supporting the use of SP factorized potentials even for the classical MTE framework.

#vars	M-Vars	A-Vars	VE[MTE]_L	VE[MTE]
4	3	2.5	0.0408	25.538
6	5	2.8	0.0209	431.836
8	5	2.6	0.0274	329.008
10	6	3.1	0.3432	—
20	9	4.0	1.7602	—

Table 2: Average run time per variable (in seconds) of the lazy variable elimination algorithm (VE[MTE]_L) and the same algorithm using classical combination and marginalization (VE[MTE]) with randomly generated MTE distributions and split points. #vars, M-vars and A-vars have the same meaning as in Table 1.

In both experiments, the benefits of using the new framework are significant. The gain in efficiency, however, is not only caused by postponing the operations or by keeping the potentials in a factorized form. An important contribution to the improvement is due to the fact that basis functions are defined only over one variable. In previous MTE inference algorithms exponential functions could be defined over several variables as the products were actually carried out. This involved dealing with linear functions in the exponents (represented as lists of variables and factors in the implementation) and consume a relatively large part of the run time of the inference algorithm.

6 Conclusions

In this paper we have developed the necessary tools for carrying out inference with MoTBFs.

The efficiency of the MoTBF-based inference procedure comes from the use of SP factorized potentials, the properties of the basis functions, and the fact that the domains of the child variables in the conditional distributions are not split. We have shown how the basic operations necessary for inference are compatible with SP factorized potentials.

The gain in efficiency with respect to the classical MTE approach has been illustrated through an experimental analysis based on a version of the variable elimination algorithm operating with MoTBFs. We have also tested the use of SP factorized potentials within the classical MTE approach. In both cases, the results of the experiments indicate that the gain in efficiency is significant.

We have only reported on experiments over networks formed by continuous variables. The impact of incorporating discrete variables is exactly the same in all the algorithms considered in this paper. We leave for future work a more extensive experimentation incorporating discrete variables.

Acknowledgments

Work supported by a Senior Grant in the frame of the CALL UCM-EEA-ABEL-02-2009 of the Abel Extraordinary Chair (NILS Project), by the Spanish Ministry of Economy and Competitivity, projects TIN2010-20900-C04-02,03 and by ERDF (FEDER) funds.

References

Cobb, B., Shenoy, P., and Rumí, R. (2004). Approximating probability density functions with mixtures of truncated exponentials. In *Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-2004)*, Perugia, Italy. In press.

Elvira Consortium (2002). Elvira: An environment for creating and using probabilistic graphical models. In Gámez, J. and Salmerón, A., editors, *Proceedings of the First European Workshop on Probabilistic Graphical Models*, pages 222–230.

Langseth, H., Nielsen, T., Rumí, R., and Salmerón, A. (2012). Mixtures of truncated basis functions. *International Journal of Approximate Reasoning*, 53:212–227.

Lauritzen, S. (1992). Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87:1098–1108.

Lauritzen, S. and Jensen, F. (2001). Stable local computation with conditional Gaussian distributions. *Statistics and Computing*, 11:191–203.

Madsen, A. (2010). Improvements to message computation in lazy propagation. *International Journal of Approximate Reasoning*, 51:499–514.

Madsen, A. and Jensen, F. (1999). Lazy propagation: a junction tree inference algorithm based on lazy evaluation. *Artificial Intelligence*, 113:203–245.

Moral, S., Rumí, R., and Salmerón, A. (2001). Mixtures of truncated exponentials in hybrid Bayesian networks. In *ECSQARU'01. Lecture Notes in Artificial Intelligence*, volume 2143, pages 135–143.

Olesen, K. (1993). Causal probabilistic networks with both discrete and continuous variables. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:275–279.

Rumí, R. and Salmerón, A. (2007). Approximate probability propagation with mixtures of truncated exponentials. *International Journal of Approximate Reasoning*, 45:191–210.

Shenoy, P. (2011). A re-definition of mixtures of polynomials for inference in hybrid Bayesian networks. *ECSQARU'11. Lecture Notes in Artificial Intelligence*, 6717:98–109.

Shenoy, P. and West, J. (2011). Inference in hybrid Bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52:641–657.