

---

# On Relaxing Determinism in Arithmetic Circuits

---

Arthur Choi<sup>1</sup> Adnan Darwiche<sup>1</sup>

## Abstract

The past decade has seen a significant interest in learning tractable probabilistic representations. Arithmetic circuits (ACs) were among the first proposed tractable representations, with some subsequent representations being instances of ACs with weaker or stronger properties. In this paper, we provide a formal basis under which variants on ACs can be compared, and where the precise roles and semantics of their various properties can be made more transparent. This allows us to place some recent developments on ACs in a clearer perspective and to also derive new results for ACs. This includes an exponential separation between ACs with and without determinism; completeness and incompleteness results; and tractability results (or lack thereof) when computing most probable explanations (MPEs).

## 1. Introduction

Arithmetic circuits (ACs) were introduced into the AI literature more than fifteen years ago as a tractable representation of probability distributions. The original work on these circuits proposed the compilation of such circuits from Bayesian networks, while identifying and assuming three circuit properties, called determinism, decomposability and smoothness (Darwiche, 2001a;b; 2003). Since then, the literature on using arithmetic circuits for probabilistic reasoning has seen two key developments. The first is the proposal made by (Lowd & Domingos, 2008) to learn these circuits directly from data—instead of just compiling them from models—therefore creating two distinct construction modes for these circuits. The second development, reported by (Poon & Domingos, 2011), amounted to proposing a class of arithmetic circuits that does not satisfy determinism, under the name of sum-product networks (SPNs).

---

<sup>1</sup>University of California, Los Angeles, California, USA. Correspondence to: Arthur Choi <aychoi@cs.ucla.edu>, Adnan Darwiche <darwiche@cs.ucla.edu>.

An examination of the literature surrounding arithmetic circuits and their variants suggests that the implications of relaxing determinism are not very well understood, even leading to conflicting claims in some cases. The treatment of smoothness has also not been very consistent as far as its necessity for certain operations on arithmetic circuits, and the complexity of enforcing it. Our goal in this paper is to address some of these issues by providing a systematic and formal treatment of arithmetic circuits, while focusing on the precise roles and semantics of their various properties and the implications of relaxing determinism.

We make several contributions in this paper. We start by reconstructing the original definition of arithmetic circuits given in (Darwiche, 2002; 2003), while assuming that these circuits represent arbitrary factors, instead of just distributions induced by Bayesian networks (a particular class of factors). We then provide definitions for decomposability, smoothness and determinism in the context of this reconstruction, while isolating precisely the role that each one of these properties play. Some of what we report on this has already been observed in the literature, but we provide alternate or more formal proofs for the sake of a systematic and inclusive treatment. We also derive new results. The first of these is a separation theorem showing that relaxing determinism can lead to exponentially smaller arithmetic circuits, while preserving the ability of these circuits to compute marginals in linear time. This begs the question of whether anything is lost from relaxing determinism. On this front, we highlight a finding that has already been reported in the literature and introduce new ones. In particular, we provide an expanded proof for the observation that relaxing determinism deprives arithmetic circuits from the ability to compute MPE in linear time. We also add a new result showing that enforcing decomposability has the power of solving MPE, even though the MPE query is not tractable for decomposable circuits. Moreover, we show that relaxing determinism leads to a type of incompleteness that we call *parametric incompleteness*, with important implications on the compilability of circuits from models. Our final contribution is a formal correctness proof of the linear-time MPE algorithm, originally proposed by (Chan & Darwiche, 2006), but with respect to the reconstructed definition of arithmetic circuits satisfying decomposability, determinism and smoothness.

This paper is structured as follows. We reconstruct the definition of arithmetic circuits as given by (Darwiche, 2002; 2003) in Section 2, but with respect to factors instead of distributions (of Bayesian networks). We then provide a new treatment of decomposability and smoothness in Section 3, followed by a new treatment of determinism in Section 4. We finally focus on the relaxation of determinism in Section 5, where we provide a new set of results and insights. An extended version of this paper includes some proofs that have been omitted here for space limitations.<sup>1</sup>

## 2. Arithmetic Circuits

Capital letters ( $X$ ) denote variables and lower-case letters ( $x$ ) denote their values. Bold capital letters ( $\mathbf{X}$ ) denote sets of variables and bold lower-case letters ( $\mathbf{x}$ ) denote their instantiations. Value  $x$  is *compatible* with instantiation  $\mathbf{y}$  iff  $\mathbf{y}$  assigns value  $x$  to  $X$  or does not assign any value to  $X$ .

**Definition 1** A factor  $f(\mathbf{X})$  over variables  $\mathbf{X}$  maps each instantiation  $\mathbf{x}$  of variables  $\mathbf{X}$  into a non-negative number  $f(\mathbf{x})$ . The factor is a distribution when  $\sum_{\mathbf{x}} f(\mathbf{x}) = 1$ .

The classical, tabular representation of a factor  $f(\mathbf{X})$  is clearly exponential in the number of variables  $\mathbf{X}$ , yet it allows one to answer key probabilistic queries efficiently. The interest here is in a more compact representation of these factors, using arithmetic circuits, while preserving the ability to answer some of these queries efficiently. We focus on the following queries, all with respect to a factor  $f(\mathbf{X})$ , with its variables  $\mathbf{X}$  partitioned into sets  $\mathbf{Y}$  and  $\mathbf{Z}$ :

- Computing the *value* of factor  $f(\mathbf{X})$  at instantiation  $\mathbf{y}$ , defined as  $f(\mathbf{y}) = \sum_{\mathbf{z}} f(\mathbf{y}, \mathbf{z})$  and called a *marginal* in this paper. This corresponds to the *probability of evidence*  $\mathbf{y}$  when the factor is a distribution.
- Computing the *MPE* of factor  $f(\mathbf{X})$ , defined as  $\operatorname{argmax}_{\mathbf{x}} f(\mathbf{x})$ . This corresponds to the most likely instantiation when the factor is a distribution.
- Computing the *MAP* over variables  $\mathbf{Y}$ , defined as  $\operatorname{argmax}_{\mathbf{y}} \sum_{\mathbf{z}} f(\mathbf{y}, \mathbf{z})$ . This is the most likely state of variables  $\mathbf{Y}$  when the factor is a distribution.

For Bayesian networks (interpreted as factors), the decision variants of the MPE, marginals, and MAP problems are, respectively, NP-complete (Shimony, 1994), PP-complete (Roth, 1996), and NP<sup>PP</sup>-complete (Park & Darwiche, 2004); see also (Darwiche, 2009). Hence, computing marginals is more difficult than computing MPE—an observation that will be quite relevant later.

|              |       |                 |     |       |                   |     |     |
|--------------|-------|-----------------|-----|-------|-------------------|-----|-----|
| $A$          | $f_1$ | $A$             | $B$ | $f_2$ | $A$               | $B$ | $f$ |
| 1            | 1     | 1               | 1   | 3     | 1                 | 1   | 3   |
| 1            | 1     | 1               | 0   | 4     | 1                 | 0   | 4   |
| 0            | 2     | 0               | 1   | 5     | 0                 | 1   | 10  |
| 0            | 2     | 0               | 0   | 6     | 0                 | 0   | 12  |
| (a) $f_1(A)$ |       | (b) $f_2(A, B)$ |     |       | (c) $f = f_1 f_2$ |     |     |

Figure 1. Two factors and their product.

We also need to define the *projection* of factor  $f(\mathbf{X})$  on variables  $\mathbf{Y}$  as the factor  $g(\mathbf{Y})$  with  $g(\mathbf{y}) = \sum_{\mathbf{z}} f(\mathbf{y}, \mathbf{z})$ . This projection will be denoted by  $\sum_{\mathbf{Z}} f(\mathbf{X})$ .

We next define an arithmetic circuit over discrete variables  $\mathbf{X}$ , as utilized in (Darwiche, 2002; 2003) to represent distributions, except that we will utilize it to represent factors. A key observation here is that the circuit inputs are not variables  $\mathbf{X}$ , but *indicator variables* that are derived from the values of variables  $\mathbf{X}$  (Darwiche, 2002; 2003).

**Definition 2** An *arithmetic circuit*  $\mathcal{AC}(\mathbf{X})$  over variables  $\mathbf{X}$  is a rooted DAG whose internal nodes are labeled with  $+$  or  $*$  and whose leaf nodes are labeled with either indicator variables  $\lambda_x$  or non-negative parameters  $\theta$ . Here,  $x$  is the value of some variable  $X$  in  $\mathbf{X}$ . The *value* of the circuit at instantiation  $\mathbf{y}$ , where  $\mathbf{Y} \subseteq \mathbf{X}$ , is denoted  $\mathcal{AC}(\mathbf{y})$  and obtained by assigning indicators  $\lambda_x$  the value 1 if  $x$  is compatible with instantiation  $\mathbf{y}$  and 0 otherwise, then evaluating the circuit in the standard way (in linear time).

The following definition makes a distinction that has not been made explicit in the literature as far as we know, but is critical for a clear semantics of arithmetic circuits.

**Definition 3** The circuit  $\mathcal{AC}(\mathbf{X})$  computes factor  $f(\mathbf{X})$  iff  $\mathcal{AC}(\mathbf{x}) = f(\mathbf{x})$  for each instantiation  $\mathbf{x}$  of variables  $\mathbf{X}$ . The circuit computes the factor marginals iff  $\mathcal{AC}(\mathbf{y}) = f(\mathbf{y})$  for each instantiation  $\mathbf{y}$  of every  $\mathbf{Y} \subseteq \mathbf{X}$ .

The notion of “computes a factor” constrains the value of an arithmetic circuit under a strict subset of its inputs (i.e., those corresponding to complete instantiations). However, the notion of “computes marginals” constrains its value under every input. Hence, two arithmetic circuits that represent distinct functions (over indicator variables) may still compute the same factor. Consider an arithmetic circuit that computes a factor  $f(X, \dots)$ , where  $X$  has values  $x$  and  $\bar{x}$ . Replacing  $\lambda_x + \lambda_{\bar{x}}$  with 1 in this circuit preserves its ability to compute the factor since  $\lambda_x + \lambda_{\bar{x}} = 1$  for every input that is relevant to computing the factor. This replacement, however, will change the function represented by the circuit and its ability to compute the factor marginals.

**Corollary 1** An arithmetic circuit that computes the marginals of a factor also computes the factor. However,

<sup>1</sup>Available at <http://reasoning.cs.ucla.edu>.

an arithmetic circuit that computes a factor does not necessarily compute its marginals.

Consider the following arithmetic circuit which computes the factor in Figure 1(c):

$$[\lambda_a + 2\lambda_{\bar{a}}][3\lambda_a\lambda_b + 4\lambda_a\lambda_{\bar{b}} + 5\lambda_{\bar{a}}\lambda_b + 6\lambda_{\bar{a}}\lambda_{\bar{b}}].$$

This circuit *does not* compute the factor marginals. Moreover, this circuit is the product of two circuits, one computing factor  $f_1$ , the other computing factor  $f_2$ , as in Figure 1.

To get further insights into the notion of “computing marginals,” we appeal to the notion of a network polynomial (Darwiche, 2003), while lifting it to factors.

**Definition 4** The polynomial of factor  $f(\mathbf{X})$  is defined as:

$$\sum_{\mathbf{x}} f(\mathbf{x}) \prod_{x \sim \mathbf{x}} \lambda_x,$$

where  $x \sim \mathbf{x}$  means that the value  $x$  of variable  $X \in \mathbf{X}$  is compatible with instantiation  $\mathbf{x}$  of variables  $\mathbf{X}$ .

The polynomial of factor  $f(A, B)$  in Figure 1(c) is  $3\lambda_a\lambda_b + 4\lambda_a\lambda_{\bar{b}} + 10\lambda_{\bar{a}}\lambda_b + 12\lambda_{\bar{a}}\lambda_{\bar{b}}$ . The polynomial of factor  $f(\mathbf{X})$  corresponds to the simplest circuit that computes the factor marginals. It is a two-level circuit though, which has an exponential size. The interest, however, is in deeper circuits that may not be exponentially sized. We later discuss circuit properties that allows us to achieve this, sometimes.

One can construct an arithmetic circuit that computes the distribution of a Bayesian network or the partition function of a Markov network in time and space that are linear in the size of these models. Each of these models correspond to a set of factors  $f_1, \dots, f_n$ , with the model representing the product of these factors. We can construct a circuit that computes each factor as given in Definition 4, then simply combine these circuits using a multiplication node. The result will compute the factor  $f = f_1 \dots f_n$  but it will not necessarily compute its marginals. We next show that if we enforce the properties of decomposability and smoothness on such a circuit, while maintaining its ability to compute the factor  $f$ , the resulting circuit will also compute the factor marginals. Therefore, these two properties turn the circuit into a tractable representation of the factor, allowing one to compute marginals by simply evaluating the circuit as in Definition 2 (in time linear in the circuit size).

### 3. Decomposability and Smoothness

The property of *decomposability* (Darwiche, 2001b) was used for tractable probabilistic reasoning in (Darwiche, 2003) by compiling Bayesian networks into arithmetic circuits that are guaranteed to be decomposable. This property was also enforced by the algorithm proposed in (Lowd & Domingos, 2008) for learning arithmetic circuits.

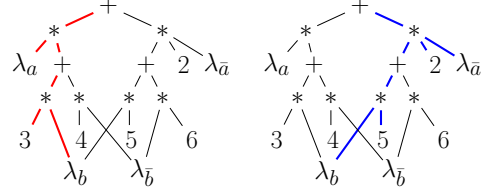


Figure 2. An AC that computes factor  $f = f_1 \cdot f_2$  from Figure 1, where an  $ab$ -subcircuit (left) and an  $a\bar{b}$ -subcircuit (right) are highlighted. This circuit is deterministic, decomposable and smooth.

**Definition 5 (Decomposability)** Let  $n$  be a node in an arithmetic circuit  $\mathcal{AC}(\mathbf{X})$ . The variables of  $n$ , denoted  $\text{vars}(n)$ , are all variables  $X \in \mathbf{X}$  with some indicator  $\lambda_x$  appearing at or under node  $n$ . An arithmetic circuit is decomposable iff every pair of children  $c_1$  and  $c_2$  of a  $*$ -node satisfies  $\text{vars}(c_1) \cap \text{vars}(c_2) = \emptyset$ .

The property of *smoothness* (Darwiche, 2001b) was also used for probabilistic reasoning in (Darwiche, 2003) by compiling circuits that are smooth. It was also enforced by the learning algorithm of (Lowd & Domingos, 2008). This property was later called *completeness* in the works on SPNs, initially in (Poon & Domingos, 2011).

**Definition 6 (Smoothness)** An arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  is smooth iff (1) it contains at least one indicator for each variable in  $\mathbf{X}$ , and (2) for every child  $c$  of  $+$ -node  $n$ , we have  $\text{vars}(n) = \text{vars}(c)$ .

Consider a factor  $f(X)$ , where variable  $X$  is binary and  $f(x) = f(\bar{x}) = \theta$ . A circuit that consists of the single parameter  $\theta$  will compute this factor but is not smooth. The circuit  $\theta\lambda_x + \theta\lambda_{\bar{x}}$  computes this factor and is smooth.

Consider a variable  $X$  with values  $x_1, \dots, x_m$ . Multiplying a circuit node by  $\lambda_{x_1} + \dots + \lambda_{x_m}$  preserves the circuit’s ability to compute a given factor since  $\lambda_{x_1} + \dots + \lambda_{x_m} = 1$  under any circuit input that is relevant to this computation. One can use this technique to ensure the smoothness of any circuit, while incurring only a polynomial overhead.<sup>2</sup> Hence, contrary to decomposability and determinism, enforcing smoothness is not difficult computationally, yet it is necessary for an arithmetic circuit to compute marginals as we discuss later. We also state the following observation, used extensively in inductive proofs that we utilize later.

**Lemma 1** Consider a decomposable and smooth arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  and define  $\mathbf{X}_n = \text{vars}(n)$  for each circuit node  $n$ . Each arithmetic circuit  $\mathcal{AC}(\mathbf{X}_n)$  rooted at node  $n$  is also decomposable and smooth.

<sup>2</sup>(Darwiche, 2001a) shows how to smooth an NNF circuit in  $O(mn)$  space and time, where  $n$  is the size of the circuit and  $m$  is the number of variables (the method can be adapted to ACs).

A main insight in this paper is the use of subcircuits, first introduced in (Chan & Darwiche, 2006) for a different purpose. They were also adopted in (Dennis & Ventura, 2015; Zhao et al., 2016) to motivate SPN learning algorithms.

**Definition 7 (Subcircuits)** *A complete subcircuit  $\alpha$  of an arithmetic circuit is obtained by traversing the circuit top-down, while choosing one child of each visited  $+$ -node and all children of each visited  $*$ -node. The term of  $\alpha$  is the set of values  $x$  for which indicator  $\lambda_x$  appears in  $\alpha$ . The coefficient of  $\alpha$  is the product of all parameters in  $\alpha$ .*

The circuit  $2*\lambda_x+1*\lambda_{\bar{x}}+3*\lambda_x$  computes factor  $f(X)$  with  $f(x) = 5$  and  $f(\bar{x}) = 1$ . It is decomposable, smooth and has three complete subcircuits, with  $(x, 2)$ ,  $(\bar{x}, 1)$  and  $(x, 3)$  as their term/coefficient pairs. Note that two subcircuits may have the same term but different coefficients.

The following lemma and its proof reveal the precise roles of decomposability and smoothness. Given decomposability, the term of a complete subcircuit will not contain conflicting values for any variable. Given smoothness, the term must contain a value for each variable.

**Lemma 2** *Let  $\alpha$  be a complete subcircuit of an arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  that is decomposable and smooth. The term of subcircuit  $\alpha$  must be an instantiation of variables  $\mathbf{X}$ .*

**Proof** Given smoothness, every variable  $X \in \mathbf{X}$  must have at least one indicator  $\lambda_x$  in  $\alpha$  (no variables are dropped from the circuit if we keep only a single child of a  $+$ -node). Given decomposability, each variable  $X \in \mathbf{X}$  must have at most one indicator  $\lambda_x$  in  $\alpha$ . Hence,  $\alpha$  will contain exactly one indicator for each variable  $X \in \mathbf{X}$ . The term of  $\alpha$  is therefore an instantiation  $\mathbf{x}$  of variable  $\mathbf{X}$ .  $\square$

A complete subcircuit with term  $\mathbf{x}$  will be called an  $\mathbf{x}$ -subcircuit. Figure 2 depicts an  $ab$ -subcircuit (in red) and an  $\bar{a}b$ -subcircuit (in blue).

In a decomposable and smooth circuit  $\mathcal{AC}(\mathbf{X})$ , every complete subcircuit is an  $\mathbf{x}$ -subcircuit for some instantiation  $\mathbf{x}$  of variables  $\mathbf{X}$ . The circuit can then be treated as a collection of  $\mathbf{x}$ -subcircuits (multiple subcircuits can have the same term). Our proofs utilize this implication heavily.

**Definition 8** *An input  $\Lambda$  for arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  assigns a value in  $\{0, 1\}$  to each circuit indicator  $\lambda_x$ . An instantiation  $\mathbf{x}$  is compatible with a circuit input  $\Lambda$ , denoted  $\mathbf{x} \sim \Lambda$ , iff the input sets  $\bar{\lambda}_x = 1$  when  $\mathbf{x}$  sets  $X = x$ .*

A circuit input can be viewed as the set of instantiations compatible with it. Consider the binary variables  $\mathbf{X} = \{A, B, C\}$  for an example. The circuit input

$$\Lambda = \{\lambda_a = 1, \lambda_{\bar{a}} = 0, \lambda_b = 1, \lambda_{\bar{b}} = 0, \lambda_c = 1, \lambda_{\bar{c}} = 0\}$$

has a single compatible instantiation  $abc$ . The input

$$\Lambda = \{\lambda_a = 0, \lambda_{\bar{a}} = 0, \lambda_b = 1, \lambda_{\bar{b}} = 0, \lambda_c = 1, \lambda_{\bar{c}} = 0\}$$

has no compatible instantiations, and the circuit input:

$$\Lambda = \{\lambda_a = 1, \lambda_{\bar{a}} = 1, \lambda_b = 1, \lambda_{\bar{b}} = 0, \lambda_c = 1, \lambda_{\bar{c}} = 0\}$$

has two compatible instantiations  $abc$  and  $\bar{a}bc$ . In this latter case, evaluating the circuit at instantiation  $bc$ , as discussed in Definition 2, leads to evaluating it under input  $\Lambda$ .

The following lemma brings us one step away from showing why decomposability and smoothness force an AC that computes a factor to also compute the factor marginals.

**Lemma 3** *Given a decomposable and smooth arithmetic circuit, let  $\theta_1, \dots, \theta_m$  be the coefficients of complete subcircuits whose terms are compatible with circuit input  $\Lambda$ . The circuit will evaluate to  $\theta_1 + \dots + \theta_m$  under input  $\Lambda$ .*

**Proof** Given Lemma 1, we will use induction on the circuit structure. The base case is a leaf circuit node (indicator or parameter). The lemma holds trivially in this case. The inductive case is when  $n$  is an internal circuit node with children  $c_1, \dots, c_k$ . Suppose the lemma hold for these children. If  $n$  is a  $*$ -node, the lemma holds for  $n$  by decomposability (the complete subcircuits of  $n$  correspond to the Cartesian product of complete subcircuits for its children). If  $n$  is a  $+$ -node, the lemma holds for  $n$  since the complete subcircuits of  $n$  correspond to the union of complete subcircuits of its children.  $\square$

**Corollary 2** *Let  $\theta_1, \dots, \theta_m$  be the coefficients of  $\mathbf{x}$ -subcircuits in a decomposable and smooth arithmetic circuit  $\mathcal{AC}(\mathbf{X})$ . We then have  $\mathcal{AC}(\mathbf{x}) = \theta_1 + \dots + \theta_m$ .*

We are now ready for the key result we are after.

**Theorem 1** *Consider an arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  that computes factor  $f(\mathbf{X})$ . If the circuit is decomposable and smooth, then it also computes the marginals of factor  $f(\mathbf{X})$ .*

**Proof** Consider an instantiation  $\mathbf{y}$  of some variables  $\mathbf{Y} \subseteq \mathbf{X}$ , let  $\mathbf{x}_1, \dots, \mathbf{x}_m$  be all instantiations of variables  $\mathbf{X}$  that are compatible with  $\mathbf{y}$ , and let  $\Lambda$  be the circuit input corresponding to these instantiations. Let  $\theta_i$  be the sum of coefficients of all  $\mathbf{x}_i$ -subcircuits. Since the circuit computes factor  $f(\mathbf{X})$ , we have  $\mathcal{AC}(\mathbf{x}_i) = f(\mathbf{x}_i)$  and, hence,  $f(\mathbf{x}_i) = \theta_i$  by Corollary 2. By Lemma 3, the circuit evaluates to  $\theta_1 + \dots + \theta_m$  under input  $\Lambda$ , which is  $f(\mathbf{x}_1) + \dots + f(\mathbf{x}_m) = f(\mathbf{y})$ . Hence, the circuit computes the factor marginals.  $\square$

This theorem justifies the standard algorithm for computing marginals on arithmetic circuits, in linear time, as proposed in (Darwiche, 2003)—that is, by simply evaluating the circuit as in Definition 2. In that work, however, the property



of determinism was also assumed (discussed in the next section). Determinism is not necessary though for computing marginals as initially observed in (Poon & Domingos, 2011).<sup>3</sup> Our proof above uses different tools than those used in (Poon & Domingos, 2011) and is set in a more general context. Moreover, these tools and associated lemmas turn out to be essential for the rest of our treatment on the role of determinism, which we discuss in the next section.

As for the necessity of smoothness, consider the circuits  $\mathcal{AC}_1(A, B) = \lambda_a \lambda_b + \lambda_{\bar{a}}$  and  $\mathcal{AC}_2(A, B) = \lambda_a \lambda_b + \lambda_{\bar{a}}(\lambda_b + \lambda_{\bar{b}})$ . Both circuits are decomposable and compute the same factor:  $f(a, b) = 1$ ,  $f(a, \bar{b}) = 0$ ,  $f(\bar{a}, b) = 1$  and  $f(\bar{a}, \bar{b}) = 1$ . However, circuit  $\mathcal{AC}_1$  is not smooth while  $\mathcal{AC}_2$  is smooth. Only  $\mathcal{AC}_2$  is guaranteed to compute factor marginals by Theorem 1. For example, evaluating  $\mathcal{AC}_1$  at instantiation  $\bar{a}$  gives  $\mathcal{AC}_1(\bar{a}) = 0 \cdot 1 + 1 = 1 \neq f(\bar{a}) = 2$  according to Definition 2, while  $\mathcal{AC}_2(\bar{a}) = f(\bar{a})$ .<sup>4</sup>

Before we discuss determinism, we note that decomposability and determinism were exploited recently in tractable, propositional reasoning within a semi-ring setting; initially in (Kimmig et al., 2012; 2016), then followed by (Friesen & Domingos, 2016).

## 4. Determinism

The property of *determinism* (Darwiche, 2001a) was employed for probabilistic reasoning in (Darwiche, 2003) by compiling Bayesian networks into arithmetic circuits that are deterministic. It was also enforced by the algorithm of (Lowd & Domingos, 2008) for learning arithmetic circuits. The property was later called *selectivity* in the works on SPNs, initially in (Peharz et al., 2014).

Using the terminology of our current formulation, the original definition of determinism would amount to this: An arithmetic circuit is deterministic iff the terms of each two of its complete subcircuits are conflicting. We will adopt a weaker definition, which allows conflicting subcircuits as long as at most one of them has a non-zero coefficient.

**Definition 9 (Determinism)** *An arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  is deterministic iff each  $+$ -node has at most one non-zero input when the circuit is evaluated under any instantiation  $\mathbf{x}$  of variables  $\mathbf{X}$ .*

<sup>3</sup>(Poon & Domingos, 2011) introduced sum-product networks (SPNs), which are equivalent to decomposable and smooth ACs. More precisely, each can be converted into the other in linear time and space (Rooshenas & Lowd, 2014). The conversion is straightforward and amounts to adjusting for graphical notation.

<sup>4</sup>Theorem 1 of (Friesen & Domingos, 2016) implies that factor marginals can be computed in time linear in the size of an arithmetic circuit, when the circuit is decomposable but not smooth. This complexity is not justified (but assumed) in the proof of the theorem. In fact, we are unaware of any justified algorithm that attains this complexity without smoothness; see also Footnote 2.

As mentioned earlier, the original proposal for using arithmetic circuits as a tractable representation of probability distributions (Darwiche, 2003) ensured that these circuits are deterministic, in addition to being decomposable and smooth. Moreover, several methods were proposed in (Darwiche, 2003) for compiling Bayesian networks into ACs with these properties. One of these methods ensures that the size of the AC is proportional to the size of a jointree for the network. Another method yields circuits that can sometimes be exponentially smaller, and is implemented in the publicly available ace system (Chavira & Darwiche, 2008); see also (Darwiche et al., 2008) for an empirical evaluation of this system in one of the UAI inference evaluations.

While determinism is not needed to compute factor marginals, it is needed for the correctness of the linear-time MPE algorithm of (Chan & Darwiche, 2006). This was missed in some earlier works (Poon & Domingos, 2011), which used this algorithm on non-deterministic ACs (i.e., SPNs) without realizing that it is no longer correct. This oversight was noticed in later works (Peharz et al., 2016; Mauá & de Campos, 2017).<sup>5</sup> We next reveal the reason why computing MPE without determinism is hard. Later in the section, we reveal the reason why the MPE algorithm of (Chan & Darwiche, 2006) fails without determinism.

The key observation is this. Consider variables  $\mathbf{X}$  which are partitioned into  $\mathbf{Y}$  and  $\mathbf{Z}$ . Given a decomposable and smooth arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  that computes factor  $f(\mathbf{X})$ , one can obtain in linear time a decomposable and smooth  $\mathcal{AC}(\mathbf{Y})$  that computes the projection  $\sum_{\mathbf{Z}} f(\mathbf{X})$ . This is achieved by simply setting all indicators  $\lambda_z$  to 1; see (Darwiche, 2001b) for the root of this observation. Moreover, an MPE for the projection  $\sum_{\mathbf{Z}} f(\mathbf{X})$  is a MAP for the original factor  $f(\mathbf{X})$ . Hence, a polytime MPE algorithm implies a polytime MAP algorithm on decomposable and smooth ACs. We know, however, that Naive Bayes networks have linear-size decomposable and smooth ACs, while MAP is hard on these networks (de Campos, 2011). Therefore, the existence of a polytime MPE algorithm on such circuits will contradict standard complexity assumptions. These observations can be abstracted into the following lemma, which succinctly and intuitively explains why MPE is not tractable on decomposable and smooth circuits.

**Lemma 4** *A circuit representation that supports projection and MPE in polytime also supports MAP in polytime.*

Note that deterministic, decomposable and smooth ACs do not support projection in polytime so the above argument

<sup>5</sup>(Peharz et al., 2016) proposed a polytime algorithm that converts an SPN into one that is deterministic and smooth (called an *augmented SPN*), but this new SPN computes a different factor than the one computed by the original SPN. Hence, its MPEs cannot be generally converted into MPEs of the original SPN.

does not apply in this case (setting indicators  $\lambda_z$  to 1 will generally destroy determinism).

More formally, let  $\mathcal{AC}(\mathbf{X})$  be a decomposable and smooth arithmetic circuit that computes a factor  $f(\mathbf{X})$ . For a given value  $k$ , consider the decision problems:

- **D-MPE-AC:** Is there an instantiation  $\mathbf{x}$  where  $\mathcal{AC}(\mathbf{x}) > k$ ?
- **D-MAP-AC:** Is there an instantiation  $\mathbf{y}$  where  $\sum_{\mathbf{z}} \mathcal{AC}(\mathbf{y}, \mathbf{z}) > k$ ? ( $\mathbf{X}$  is partitioned into  $\mathbf{Y}$  and  $\mathbf{Z}$ ).

We now have the following result, whose proof expands the one given in (Peharz et al., 2016) for SPNs based on the above observations; see also (Mauá & de Campos, 2017) for an in-depth discussion of MPE hardness on SPNs.

**Theorem 2** *The problem D-MPE-AC is NP-complete.*

**Proof** Given instantiation  $\mathbf{x}$  and value  $k$ , we can test whether  $f(\mathbf{x}) > k$  by evaluating the circuit  $\mathcal{AC}$  in time linear in the size of the circuit. Hence, the problem is in NP. To show NP-hardness, we reduce the (decision) problem of computing (partial) MAP in a naive Bayes network, which is NP-complete (de Campos, 2011), to MPE in a decomposable and smooth arithmetic circuit. Suppose we have a naive Bayes network with a root node  $X_0$  and leaf nodes  $\mathbf{X}$ , and inducing a distribution  $Pr(X_0, \mathbf{X})$ . We can compile this network into a polysize decomposable, deterministic and smooth arithmetic circuit  $\mathcal{AC}_0(X_0, \mathbf{X})$  that computes  $Pr(X_0, \mathbf{X})$ , e.g., as in (Chavira & Darwiche, 2008). We can sum-out variable  $X_0$  in the circuit  $\mathcal{AC}_0(X_0, \mathbf{X})$  by setting the indicators of  $X_0$  to one. The resulting circuit  $\mathcal{AC}_1(\mathbf{X})$  is decomposable and smooth, and computes the (marginals of) factor  $\sum_{X_0} Pr(X_0, \mathbf{X})$ . For a given value  $k$ , there exists an instantiation  $\mathbf{x}$  where  $\mathcal{AC}_1(\mathbf{x}) > k$  iff there exists an instantiation  $\mathbf{x}$  where  $\sum_{x_0} Pr(x_0, \mathbf{x}) > k$ , which is an NP-complete problem (de Campos, 2011).  $\square$

**Corollary 3** *The problem D-MAP-AC is NP-complete.*

The following lemma reveals the precise role of determinism, which stands behind the correctness of the linear-time MPE algorithm of (Chan & Darwiche, 2006). It basically shows a one-to-one correspondence between the non-zero rows of the factor computed by a circuit and the complete subcircuits with non-zero coefficients.

**Lemma 5** *Consider an arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  that computes factor  $f(\mathbf{X})$  and is deterministic, decomposable and smooth (hence, can be viewed as a collection of  $\mathbf{x}$ -subcircuits). For each instantiation  $\mathbf{x}$ , we have:*

- (a) *If the circuit has two distinct  $\mathbf{x}$ -subcircuits, one of them must have a zero coefficient.*

- (b) *If  $f(\mathbf{x}) > 0$ , the circuit contains a unique  $\mathbf{x}$ -subcircuit with coefficient  $f(\mathbf{x})$ .*

**Proof** To prove (a), suppose the circuit contains two distinct  $\mathbf{x}$ -subcircuits  $\alpha_1$  and  $\alpha_2$  that have non-zero coefficients. We will now establish a contradiction. Since  $\alpha_1$  and  $\alpha_2$  are distinct, each  $\alpha_i$  must include a distinct child  $c_i$  of some  $+$ -node in the circuit. If we evaluate the circuit at instantiation  $\mathbf{x}$ , both  $c_1$  and  $c_2$  will have non-zero values. Hence, the circuit cannot be deterministic, which is a contradiction. To prove (b), suppose  $f(\mathbf{x}) > 0$  and let  $\alpha_1, \dots, \alpha_m$  be all  $\mathbf{x}$ -subcircuits. At most one  $\alpha_i$  can have a non-zero coefficient by (a). Since the circuit computes the factor, it must evaluate to  $f(\mathbf{x})$  under instantiation  $\mathbf{x}$ . Hence, exactly one  $\alpha_i$  has  $f(\mathbf{x})$  as its coefficient.  $\square$

Lemma 5 allows us to prove the correctness of the MPE algorithm given by (Chan & Darwiche, 2006) under the more general setting we have in this paper. This original algorithm is based on converting a deterministic, decomposable and smooth AC that computes a distribution  $Pr(\mathbf{X})$  into a *maximizer circuit*. Evaluating this circuit under evidence  $\mathbf{y}$ , as in Definition 2, gives the MPE value  $\arg\max_{\mathbf{x} \sim \mathbf{y}} Pr(\mathbf{x})$ .

An arithmetic circuit  $\mathcal{AC}(\mathbf{X})$  is converted into a maximizer circuit, denoted  $\mathcal{AC}_{\max}(\mathbf{X})$ , by replacing every  $+$ -node with a  $\max$ -node. The complete subcircuits of  $\mathcal{AC}_{\max}$  are defined as in Definition 7, but where exactly one child of each visited  $\max$ -node is selected.

**Theorem 3** *Let  $\mathcal{AC}(\mathbf{X})$  be a deterministic, decomposable and smooth arithmetic circuit that computes a factor  $f(\mathbf{X})$  and let  $\mathcal{AC}_{\max}(\mathbf{X})$  be its maximizer circuit. Then  $\mathcal{AC}_{\max}(\mathbf{y}) = \max_{\mathbf{x} \sim \mathbf{y}} f(\mathbf{x})$  for  $\mathbf{Y} \subseteq \mathbf{X}$ .*

**Proof** By Lemma 5, there is a one-to-one correspondence between the non-zero rows of factor  $f(\mathbf{X})$  and  $\mathbf{x}$ -subcircuits with non-zero coefficients. Let  $\theta_1, \dots, \theta_m$  be the coefficients of  $\mathbf{x}$ -subcircuits, where  $\mathbf{x}$  is compatible with  $\mathbf{y}$ . Hence,  $\max\{\theta_1, \dots, \theta_m\} = \max_{\mathbf{x} \sim \mathbf{y}} f(\mathbf{x})$ . That is, the MPE value is a coefficient of some  $\mathbf{x}$ -subcircuit—call it an MPE-subcircuit. We will think of the algorithm as composing an MPE-subcircuit in addition to computing its coefficient and show that  $\mathcal{AC}_{\max}(\mathbf{y}) = \max\{\theta_1, \dots, \theta_m\}$  by induction on the circuit structure (see Lemma 1). The base case trivially holds for leaf circuit nodes (indicators and parameters). Assume  $n$  is an internal circuit node and the above equality holds for its children  $c_1, \dots, c_k$  having MPE-subcircuits  $\alpha_i$  and coefficient  $\eta_i$ . If  $n$  is a  $*$ -node, then by decomposability, an MPE-subcircuit for  $n$  can be found by joining  $\alpha_1, \dots, \alpha_k$  with  $\eta_1 * \dots * \eta_k$  as its coefficient. If  $n$  is a  $\max$ -node, then by determinism, an MPE-subcircuit for  $n$  can be found from the  $\alpha_i$  with the largest  $\eta_i$  with  $\max_{i=1}^k \eta_i$  as its coefficient.  $\square$

Once a maximizer circuit is evaluated to  $\theta$ , one can identify

an  $\mathbf{x}$ -subcircuit that has  $\theta$  as its coefficient, with  $\mathbf{x}$  being an MPE instantiation; see (Chan & Darwiche, 2006).<sup>6</sup>

Without determinism, a circuit may have multiple  $\mathbf{x}$ -subcircuits for a given  $\mathbf{x}$ , each having a non-zero coefficient. By Corollary 2, the value of  $\mathbf{x}$ ,  $\mathcal{AC}(\mathbf{x}) = f(\mathbf{x})$ , is the sum of these coefficients. An MPE algorithm that does not perform this sum cannot be correct.<sup>7</sup>

Before we further discuss the impact of relaxing determinism, we point to a new class of arithmetic circuits, the Probabilistic Sentential Decision Diagram (PSDD) (Kisa et al., 2014), which imposes *stronger* versions of decomposability and determinism. This enables the multiplication of two ACs in polytime, which is otherwise hard under the standard versions of these properties (Shen et al., 2016).

## 5. The Impact of Relaxing Determinism

We now consider two new implications of relaxing determinism, one positive and one negative. We also address an apparent paradox: How could a representation (decomposable and smooth ACs) allow the computation of marginals easily (a PP-complete problem), yet not allow the computation of MPE easily (an NP-complete problem)? Recall that the complexity class NP is included in the class PP.

The positive implication is that relaxing determinism can lead to exponentially smaller arithmetic circuits.

**Theorem 4 (Separation)** *There is a family of factors  $f_n(\mathbf{X}_n)$  where (1) there exists a decomposable and smooth arithmetic circuit  $\mathcal{AC}_n(\mathbf{X}_n)$  that computes the marginals of  $f_n$ , with a size polynomial in  $n$ ; (2) every deterministic, decomposable and smooth circuit that computes the marginals of factor  $f_n$  must have a size exponential in  $n$ .*

**Proof** (Bova et al., 2016) identifies a family of Boolean functions (the Sauerhoff functions)  $S_n$  that have decomposable NNFs (DNNFs) with sizes polynomial in  $n$ , but where their deterministic DNNFs (d-DNNFs) must have sizes exponential in  $n$ . Previously known separations were conditional on the polynomial hierarchy not collapsing (Darwiche & Marquis, 2002), but (Bova et al., 2016) does not make such an assumption (and neither do we).

<sup>6</sup>Smoothness is not strictly needed for this algorithm, except that it ensures that a full variable instantiation is returned.

<sup>7</sup>This MPE algorithm was used on selective SPNs (equivalent to deterministic and decomposable ACs) in (Peharz et al., 2016). It was also adapted to algebraic model counting (AMC) in (Kimmig et al., 2016) and to Sum-Product Functions (SPFs) in (Friesen & Domingos, 2016). Determinism was not required in (Kimmig et al., 2016). This is sound since AMC problems correspond to Boolean circuits where the weight of an instantiation is a product of literal weights, and is independent of how many times the instantiation appears as a subcircuit.

Let  $g_n$  denote a polysize DNNF for function  $S_n$  and let  $\mathcal{AC}_n$  denote the polysize decomposable and smooth arithmetic circuit obtained by: replacing the inputs of  $g_n$  with the corresponding indicator variables, replacing conjunctions and disjunctions by products and sums, respectively, then smoothing if necessary. The resulting arithmetic circuit  $\mathcal{AC}_n$  has a positive value (possibly  $> 1$ ) on input  $\mathbf{x}$  iff the original function  $S_n$  evaluates to true. We now show that if  $f_n$  is the factor computed by arithmetic circuit  $\mathcal{AC}_n$ , then any deterministic, decomposable and smooth AC that computes  $f_n$  must have an exponential size.

Let  $\mathcal{AC}'_n$  be such a circuit. Consider the d-DNNF  $g'_n$  obtained by: replacing the indicator variables with the corresponding literals of variables  $\mathbf{X}$ , replacing products and sums with conjunctions and disjunctions, respectively, and by replacing positive parameters with true and zero parameters with false. Note that  $g'_n(\mathbf{x})$  is true iff  $\mathcal{AC}'_n(\mathbf{x}) > 0$ , i.e., a complete subcircuit for  $g'_n$  evaluates to true iff the corresponding subcircuit for  $\mathcal{AC}'_n$  has a positive coefficient. Hence, if  $\mathcal{AC}'_n$  had a sub-exponential size, then function  $S_n$  would have a sub-exponentially sized d-DNNF, which we know does not exist (Bova et al., 2016).  $\square$

We now get to a newly identified, negative implication of relaxing determinism. It pertains to compiling ACs from probabilistic models and requires the following notion.

**Definition 10** *A set of parameters  $\Theta$  is complete for factor  $f(\mathbf{X})$  iff for every instantiation  $\mathbf{x}$ , the parameter  $f(\mathbf{x})$  can be expressed as a product of parameters in  $\Theta$ .*

The parameters of a Bayesian network are complete for its distribution; those of a Markov network are complete for its partition function; and the parameters  $\Theta = \{0, 1\}$  are complete for Boolean factors:  $f(\mathbf{X})$  with  $f(\mathbf{x}) \in \{0, 1\}$ .

We will write  $\mathcal{AC}(\mathbf{X}, \Theta)$  to denote an arithmetic circuit whose parameters are in  $\Theta$ . The following theorem states a key property which is lost when relaxing determinism.

**Theorem 5 (Completeness)** *Consider factor  $f(\mathbf{X})$  and complete parameters  $\Theta$ . There must exist an arithmetic circuit  $\mathcal{AC}(\mathbf{X}, \Theta)$  that computes the factor marginals and is deterministic, decomposable and smooth.*

**Proof** Consider the factor polynomial  $\sum_{\mathbf{x}} f(\mathbf{x}) \prod_{x \sim \mathbf{x}} \lambda_x$  in Definition 4 and replace each  $f(\mathbf{x})$  by a product of parameters from  $\Theta$ . The result can be represented by an AC that is deterministic, decomposable and smooth.  $\square$

The standard methods for compiling Bayesian networks, and graphical models more generally, into arithmetic circuits do indeed limit the circuit parameters to those appearing in the model factors. Hence, the compilation process amounts only to finding a (small) circuit structure since

the circuit parameters are already predetermined. As mentioned earlier, these methods can yield relatively small circuits for some graphical models with very high treewidth (Chavira & Darwiche, 2008; Darwiche et al., 2008).

The above property is lost if one insists on constructing arithmetic circuits that are decomposable and smooth, but not deterministic. This is shown in the following theorem, which refers to *dead* circuit nodes: ones that appear only in complete subcircuits that have zero coefficients.<sup>8</sup>

**Theorem 6 (Parametric Incompleteness)** *Let  $f(\mathbf{X})$  be a Boolean factor and  $\Theta = \{0, 1\}$  ( $\Theta$  is complete for  $f$ ). A circuit  $\mathcal{AC}(\mathbf{X}, \Theta)$  cannot compute  $f(\mathbf{X})$  if it is decomposable, smooth and free of dead nodes, but not deterministic.*

**Proof** If the AC has no  $+$ -node, then it is vacuously deterministic. Otherwise, it has a  $+$ -node. Since the circuit is not deterministic, there is a  $+$ -node  $n$  that violates determinism. This node is included in some complete subcircuit with a non-zero coefficient (otherwise, the node  $n$  is dead). Since node  $n$  violates determinism, we can find two distinct  $\mathbf{x}$ -subcircuits, with non-zero coefficients, that differ by the branch selected at node  $n$ . Since the circuit computes factor  $f(\mathbf{X})$ , Lemma 3 implies that the coefficients of these  $\mathbf{x}$ -subcircuits must add up to  $f(\mathbf{x}) = 1$ . There must then exist an  $\mathbf{x}$ -subcircuit whose coefficient is in  $(0, 1)$ , exclusive, i.e., the circuit has a parameter not in  $\{0, 1\}$ .  $\square$

In other words, if a decomposable and smooth circuit  $\mathcal{AC}(\mathbf{X}, \{0, 1\})$  computes the marginals of a Boolean factor, it must also be non-trivially deterministic. This result has a major implication on compiling probabilistic graphical models into ACs that are not deterministic. That is, one cannot generally restrict the circuit parameters to those appearing in the model, otherwise a circuit may not exist.

Therefore, while relaxing determinism can lead to exponentially smaller circuits, finding these circuits is now more involved as it may require searching for parameters. This demands new techniques as all techniques we are aware of for compiling models into deterministic circuits assume that the circuit parameters come from model parameters.

Our last contribution relates to the following apparent paradox. Suppose we have a set of factors  $f_1(\mathbf{X}_1), \dots, f_n(\mathbf{X}_n)$ , representing a probabilistic graphical model that has a corresponding joint factor  $f = f_1 \cdots f_n$ . Consider now the following decision problems, over such probabilistic graphical models, which correspond to computing the MPE and marginals:

- **D-MPE:** Is there an instantiation  $\mathbf{x}$  where  $f(\mathbf{x}) > k$ ?
- **D-PR:** Is  $\sum_{\mathbf{x}} f(\mathbf{x}) > k$ ?

D-MPE is NP-complete, whereas D-PR is PP-complete. Moreover, the complexity class PP includes NP. Yet, decomposable and smooth ACs allow one to compute marginals in linear time, while computing MPE, which is no harder, is hard on these circuits!

To resolve this apparent paradox, one must observe the sometimes subtle distinction between a *representation* and the *computation* needed to produce that representation. The representation here is decomposable and smooth ACs, and the computation is the algorithm used to compile a graphical model into this representation. While the representation itself does not facilitate the computation of MPE, the compilation algorithm must be sufficient to compute the MPE query without additional complexity (beyond polynomial). To formalize this, we need the following lemma.

**Lemma 6** *D-MPE can be reduced to D-PR.*

We now have the following result, which implies that a polytime compilation algorithm for decomposable and smooth ACs can be used as a sub-routine for a polytime algorithm for computing MPEs (which we do not expect to exist, under typical complexity theoretic assumptions).

**Theorem 7** *Consider an algorithm  $\Xi$  that takes a set of factors  $f_1(\mathbf{X}_1), \dots, f_n(\mathbf{X}_n)$  as input and returns a decomposable and smooth arithmetic circuit that computes the marginals of factor  $f = f_1 \cdots f_n$ . Let  $s$  be the size of input factors and let  $O(t(s))$  be the time complexity of algorithm  $\Xi$ . One can compute the MPE of factor  $f$  in time  $O(t(\text{poly}(s)))$ .*

These findings highlight an interesting property of decomposable and smooth ACs. They “store” the results to an exponential number of marginal queries, where each result can be retrieved by a simple traversal of the circuit. Yet, they do not “store” the answers to MPE queries, even though these queries are easier. The implication of this can be seen from two angles, depending on whether these circuits are compiled from models or learned from data. In the former case, the compilation algorithm is readily available to answer MPE queries, but at the cost of invoking this algorithm for each query. In the latter case, however, answering MPE queries remains a challenge. Hence, learning circuits that are not deterministic needs to yield an additional benefit that compensates for this loss in tractability. This could be a simpler learning algorithm; a smaller learned circuit; or a learned circuit whose factor is superior from a statistical learning viewpoint.

<sup>8</sup>Dead nodes can be replaced by the constant zero without changing the factor computed by the circuit. One can relax determinism trivially by adding dead nodes, but that does not help as far as obtaining smaller circuits.



## Acknowledgements

This work has been partially supported by NSF grant #IIS-1514253, ONR grant #N00014-15-1-2339 and DARPA XAI grant #N66001-17-2-4032. We thank YooJung Choi, Umüt Oztok, Yujia Shen, and Guy Van den Broeck for comments and discussions on this paper.

## A. Proofs

**Proof of Corollary 3** Given an instantiation  $\mathbf{y}$  and value  $k$ , we can test whether  $\sum_{\mathbf{z}} \mathcal{AC}(\mathbf{y}, \mathbf{z}) > k$ , by evaluating the circuit  $\mathcal{AC}$  in time linear in the size of the circuit. Hence, the problem is in NP. To show the problem is NP-hard, we reduce the (decision) problem of computing (partial) MAP in a naive Bayes network, as in the proof of Theorem 2.  $\square$

**Proof of Lemma 6.** We first reduce D-MPE to satisfiability on a CNF (this is essentially an instance of Cook’s theorem). Satisfying assignments of the CNF correspond to solutions of the D-MPE query, which we can count using a D-PR query.

First, we construct a Boolean circuit that takes inputs  $\mathbf{x}$ , and outputs true if  $f(\mathbf{x}) > k$  and false otherwise. We construct a (multiplexer) circuit for each factor  $f_i(\mathbf{X}_i)$ , which has inputs  $\mathbf{x}_i$  and outputs a bitstring  $\mathbf{y}_i$  representing a bit encoding of the value  $f_i(\mathbf{x}_i)$  (which we assume are rational values). We then construct a circuit that represents a multiplier, which takes as input the bitstrings  $\mathbf{y}_i$  and outputs another bitstring  $\mathbf{z}$  representing the product of  $f_i(\mathbf{x}_i)$ . Finally, we have another circuit that takes the bitstring  $\mathbf{z}$  as input, and outputs true if this bitstring represents a value that is greater than  $k$ , and false otherwise. Hence, the output of this circuit is true iff  $f(\mathbf{x}) > k$ . Each one of the constructed circuits has size polynomial in the size of the inputs, i.e., the aggregate size of the factors and the number of bits needed to represent their values.

We can reduce this circuit to a CNF by adding auxiliary variables  $\mathbf{Y}$ , using one new variable for the output of each gate, i.e., we reduce circuit satisfiability to 3-SAT; see, e.g., (Kleinberg & Tardos, 2006). This results in a set of Boolean factors  $g_1, \dots, g_m$ . If  $g = g_1 \cdots g_m$ , then  $g(\mathbf{x}) = \sum_{\mathbf{y}} g(\mathbf{x}, \mathbf{y}) > 0$  iff  $f(\mathbf{x}) > k$ , and  $\sum_{\mathbf{x}, \mathbf{y}} g(\mathbf{x}, \mathbf{y}) > 0$  iff there exists an input  $\mathbf{x}$  where  $f(\mathbf{x}) > k$ .  $\square$

**Proof of Theorem 7.** Given factors  $f_1, \dots, f_n$  of size  $s$ , we first construct a set of factors  $g_1, \dots, g_m$  of size  $\text{poly}(s)$ , as in Lemma 6. We invoke algorithm  $\Xi$  on factors  $g_1, \dots, g_m$ , obtaining a decomposable and smooth arithmetic circuit  $g$  representing  $g_1 \cdots g_m$  in time  $O(t(\text{poly}(s)))$ . The size of  $g$  is also  $O(t(\text{poly}(s)))$  (the size of the circuit cannot be larger than the time needed to construct it). The same amount of time is required to evaluate the marginal of  $g$ , hence the overall time to compute

the MPE of factor  $f$  is  $O(t(\text{poly}(s)))$ .  $\square$

## References

- Bova, Simone, Capelli, Florent, Mengel, Stefan, and Slivovsky, Friedrich. Knowledge compilation meets communication complexity. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1008–1014, 2016.
- Chan, Hei and Darwiche, Adnan. On the robustness of most probable explanations. In *Proceedings of the 22nd Conference in Uncertainty in Artificial Intelligence (UAI)*, 2006.
- Chavira, Mark and Darwiche, Adnan. On probabilistic inference by weighted model counting. *Artificial Intelligence*, 172(6–7): 772–799, April 2008.
- Darwiche, Adnan. On the tractable counting of theory models and its application to truth maintenance and belief revision. *Journal of Applied Non-Classical Logics*, 11(1-2):11–34, 2001a.
- Darwiche, Adnan. Decomposable negation normal form. *Journal of the ACM*, 48(4):608–647, 2001b.
- Darwiche, Adnan. A logical approach to factoring belief networks. In *Proceedings of KR*, pp. 409–420, 2002.
- Darwiche, Adnan. A differential approach to inference in Bayesian networks. *J. ACM*, 50(3):280–305, 2003.
- Darwiche, Adnan. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.
- Darwiche, Adnan and Marquis, Pierre. A knowledge compilation map. *JAIR*, 17:229–264, 2002.
- Darwiche, Adnan, Dechter, Rina, Choi, Arthur, Gogate, Vibhav, and Otten, Lars. Results from the probabilistic inference evaluation of UAI-08. 2008.
- de Campos, Cassio Polpo. New complexity results for MAP in Bayesian networks. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2100–2106, 2011.
- Dennis, Aaron W. and Ventura, Dan. Greedy structure search for sum-product networks. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 932–938, 2015.
- Friesen, Abram L. and Domingos, Pedro M. The sum-product theorem: A foundation for learning tractable models. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 1909–1918, 2016.
- Kimmig, Angelika, Van den Broeck, Guy, and De Raedt, Luc. Algebraic model counting. *CoRR*, abs/1211.4475, 2012. URL <http://arxiv.org/abs/1211.4475>.
- Kimmig, Angelika, Van den Broeck, Guy, and De Raedt, Luc. Algebraic model counting. *International Journal of Applied Logic*, November 2016.
- Kisa, Doga, Van den Broeck, Guy, Choi, Arthur, and Darwiche, Adnan. Probabilistic sentential decision diagrams. In *KR*, 2014.

- Kleinberg, Jon M. and Tardos, Éva. *Algorithm design*. Addison-Wesley, 2006.
- Lowd, Daniel and Domingos, Pedro M. Learning arithmetic circuits. In *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence (UAI)*, pp. 383–392, 2008.
- Mauá, Denis Deratani and de Campos, Cassio Polpo. Approximation complexity of maximum a posteriori inference in sum-product networks. *CoRR*, abs/1703.06045, 2017.
- Park, James D. and Darwiche, Adnan. Complexity results and approximation strategies for MAP explanations. *J. Artif. Intell. Res. (JAIR)*, 21:101–133, 2004.
- Peharz, Robert, Gens, Robert, and Domingos, Pedro M. Learning selective sum-product networks. In *LTPM workshop*, 2014.
- Peharz, Robert, Gens, Robert, Pernkopf, Franz, and Domingos, Pedro. On the latent variable interpretation in sum-product networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2016.
- Poon, Hoifung and Domingos, Pedro M. Sum-product networks: A new deep architecture. In *UAI*, pp. 337–346, 2011.
- Rooshenas, Amirmohammad and Lowd, Daniel. Learning sum-product networks with direct and indirect variable interactions. In *ICML*, pp. 710–718, 2014.
- Roth, Dan. On the hardness of approximate reasoning. *Artif. Intell.*, 82(1-2):273–302, 1996.
- Shen, Yujia, Choi, Arthur, and Darwiche, Adnan. Tractable operations for arithmetic circuits of probabilistic models. In *Advances in Neural Information Processing Systems 29 (NIPS)*, 2016.
- Shimony, Solomon Eyal. Finding MAPs for belief networks is NP-hard. *Artif. Intell.*, 68(2):399–410, 1994.
- Zhao, Han, Poupart, Pascal, and Gordon, Geoffrey J. A unified approach for learning the parameters of sum-product networks. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pp. 433–441, 2016.