Marginal Consistency: Upper-Bounding Partition Functions over Commutative Semirings

Tomáš Werner

Abstract—Many inference tasks in pattern recognition and artificial intelligence lead to partition functions in which addition and multiplication are abstract binary operations forming a commutative semiring. By generalizing max-sum diffusion (one of convergent message passing algorithms for approximate MAP inference in graphical models), we propose an iterative algorithm to upper bound such partition functions over commutative semirings. The iteration of the algorithm is remarkably simple: change any two factors of the partition function such that their product remains the same and their overlapping marginals become equal. In many commutative semirings, repeating this iteration for different pairs of factors converges to a fixed point when the overlapping marginals of every pair of factors coincide. We call this state marginal consistency. During that, an upper bound on the partition function monotonically decreases. This abstract algorithm unifies several existing algorithms, including max-sum diffusion and basic costraint propagation (or local consistency) algorithms in constraint programming. We further construct a hierarchy of marginal consistencies of increasingly higher levels and show than any such level can be enforced by adding identity factors of higher arity (order). Finally, we discuss instances of the framework for several semirings, including the distributive lattice and the max-sum and sum-product semirings.

Index Terms—Partition function, commutative semiring, graphical model, Markov random field, linear programming relaxation, message passing, max-sum diffusion, soft constraint satisfaction, local consistency, constraint propagation

1 Introduction

A *partially separable function* is the product $\prod_{A\in E} f_A(x_A)$ where $E \subseteq 2^V$ is a hypergraph and each *factor* f_A is a function of variables $x_A = (x_i)_{i\in A}$. The sum of the values of this function over all the variables $x_V = (x_i)_{i\in V}$ is the *partition function*

$$\sum_{x_V} \prod_{A\in E} f_A(x_A). \tag{1}$$

E.g., for $V = \{1,2,3,4\}$ and $E = \{\{1,3,4\},\{1,2\},\{2,3\}\}$ the partition function is the number

$$\sum_{x_1,x_2,x_3,x_4} f_{134}(x_1,x_3,x_4) \times f_{12}(x_1,x_2) \times f_{23}(x_2,x_3),$$

where we abbreviated $f_{\{1,3,4\}}$ by f_{134}, etc.

It is known [1], [6], [7], [35], [52], [59] that many inference tasks in pattern recognition and artificial intelligence lead to expressions of the form (1) where $+$ and $\times$ are not the ordinary arithmetic operations but abstract binary operations on some set S such that both operations are associative and commutative and $\times$ distributes over $+$. Such a structure $(S,+,\times)$ is known as the *commutative semiring* [24], [26].

The simplest instance is obtained for the *or-and semiring* $(\{0,1\},\max,\min)$. Here the semiring operations have the meaning of logical disjunction and conjunction and (1) is the decision problem asking whether there is a configuration satisfying all the predicates f_A. This problem is known in computer vision and pattern recognition as the *consistent labeling problem* [27], [28] and in artificial intelligence and constraint programming [55] as the *constraint satisfaction problem* (CSP) [20], [45]. The latter name is more widely used today. Here, the factors f_A are usually called *constraints* and the collection f_A, $A \in E$, a *network of constraints*.

The ordinary 'crisp' CSP has been generalized to handle 'soft' constraints, which can be partially satisfied rather than completely satisfied or completely violated [48]. This leads to optimization problems. One important such formulation is obtained for the *max-min semiring* $([0,1],\max,\min)$. This problem was first proposed in [54] but it is more widely known as the *fuzzy CSP* [17]. Another important formulation is obtained for the *max-sum semirings* $(\mathbb{R},\max,+)$ and $(\mathbb{R}\cup\{-\infty\},\max,+)$ where $+$ is the ordinary addition. In computer vision and pattern recognition, this problem has been called a two-dimensional grammar [58], [61], discrete energy minimization [8], [33], [66], MAP inference in graphical models [69], or the max-sum labeling problem [72]. In constraint programming, it has been called the partial [40], weighted [48] or valued [68] CSP. Several other soft CSP formulations exist [5], [48].

Two abstract frameworks have been proposed in constraint programming to unify various formulations of soft CSPs, [56] and [6], [7]. The latter is strictly more general than the former [5] and closely related to our formulation. The main difference to our formulation is that [6], [7], [56] assume that the semiring addition is idempotent ($a + a = a$ for all $a \in S$) because otherwise (1) might no longer be an optimization problem (such as if it is the ordinary partition function).

For the max-sum semiring, a class of approaches to problem (1) is based on linear programming relaxation [10], [37], [40], [61], [67], [69], [72] or, equivalently, dual

• *The author is with the Department of Cybernetics, Czech Technical University, Karlovo náměstí 13, 12135, Praha, Czech Republic. E-mail: werner@cmp.felk.cvut.cz.*

Manuscript received 20 Sept. 2013; revised 10 Sept. 2014; accepted 13 Oct. 2014. Date of publication 16 Oct. 2014; date of current version 5 June 2015. Recommended for acceptance by A.J. Storkey.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2014.2363833

0162-8828 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

decomposition [32], [39], [64]. To solve this LP relaxation, convergent message-passing algorithms have been proposed [23], [32], [37], [41], [47], [64] that monotonically decrease a convex upper bound on (1) by minimizations over blocks of variables. These algorithms are closely related to one another: they converge in infinite time to a fixed point, which is a local (with respect to block-coordinate moves) optimum of the relaxation. The simplest and oldest of them is *max-sum diffusion*, proposed in 1970's for purely unary constraints but never published [41], and recently revisited in [72], [74].

Convergent message passing algorithms repeat a simple local operation which propagates information through the network and monotonically decreases some quantity. In this respect, they resemble *local consistency* (or *constraint propagation*) algorithms [4], [15], [50], used in constraint programming to prune the search space of the CSP. The most widely known local consistency is *arc consistency* [46], [4, Section 4]. In computer vision, an algorithm equivalent to enforcing arc consistency was proposed by Waltz [70] and revisited by Rosenfeld et al. [54], who called it *discrete relaxation labeling*. Unlike message passing algorithms, local consistency algorithms in CSP converge in polynomial time, so they can be easily maintained during search.

In constraint programming, the question appeared whether local consistency algorithms can be extended from the ordinary CSP to soft CSPs. This turns out to be straightforward for soft CSPs with idempotent semiring multiplication ($a \times a = a$ for all $a \in S$), such as in the max-min semiring. The resulting algorithms are polynomial and their fixed point does not depend on the order of updates [5], [6], [7], [54], [56]. However, for non-idempotent semiring multiplication (such as in the max-sum semiring), no finite network algorithm has been found that would naturally generalize classical local consistency algorithms [5], [6], [7], [12], [56]. Motivated primarily by efficiency in branch-and-bound search, several finite algorithms have been proposed [13] but they provide weaker bounds than convergent message-passing algorithms.

1.1 Contribution

In this paper, we show that max-sum diffusion can be naturally generalized to the abstract commutative semiring. The update of the resulting algorithm is remarkably simple:

> *Change two factors f_A and f_B such that the function $f_A \times f_B$ is preserved and the overlapping marginals of f_A and f_B become equal,*

i.e., change f_A and f_B such that $f_A(x_A) \times f_B(x_B)$ is preserved for all $x_{A \cup B}$ and $\sum_{x_{A \setminus B}} f_A(x_A) = \sum_{x_{B \setminus A}} f_B(x_B)$ for all $x_{A \cap B}$. In many semirings, repeating this operation for different pairs of factors converges to a fixed point. This results in the observation that has never before been clearly formulated:

> *In many commutative semirings, every partially separable function can be reparameterized by local operations to a state when the overlapping marginals of each pair of factors coincide.*

We call this state *marginal consistency* and the algorithm *enforcing marginal consistency*. This terminology agrees with that in constraint programming [4], which distinguishes, e.g., *arc consistency* (a property of a network) and *enforcing arc consistency* (an algorithm to achieve this property). Marginal consistency can be enforced in a number of commutative semirings, including the or-and, max-min, max-sum, and sum-product semiring. As special cases, we obtain basic local consistency algorithms in CSP, including arc consistency.

We further show that $\prod_{A \in E} \sum_{x_A} f_A(x_A)$ is an upper bound on the semiring partition function, with respect to the *canonical order* on the semiring [26]. If the semiring satisfies the Cauchy-Schwarz inequality, the upper bound monotonically decreases during the algorithm.

For the max-sum semiring, it was observed that the basic LP relaxation of (1) [10], [40], [61], [67], [72] can be made tighter at the expense of more computational effort [32], [38], [40], [42], [62], [63], [65], [69], [74]. Some researchers proposed whole hierarchies of increasingly tighter LP relaxations [69, Section 8.5], [63], [74]. This is similar to using increasingly larger subproblems in dual decomposition [39]. In [19], [74], we constructed such a hierarchy by adding 'dummy' zero constraints of higher arities. Zero constraints can be added incrementally during max-sum diffusion in a dual cutting-plane fashion. We generalize this technique to other semirings, obtaining a hierarchy of marginal consistencies of increasingly higher levels. For the or-and semiring, this hierarchy contains (strong) k-consistency in CSP [4], [21].

Even for idempotent semiring multiplication our algorithm is simpler than local consistency algorithms proposed for soft CSPs [5], [6], [7], [56].

Our framework does not cover *belief propagation* (or the *sum-product algorithm*) [51], [69], which computes, in polynomial time for acyclic networks, the exact partition function and marginals. This algorithm (and its junction-tree version) can be generalized to any commutative semiring [1]. This is straightforward because its update rule uses only the operations of the sum-product semiring. In contrast, the max-sum diffusion update [72, Section VI.A] uses not only the operations 'max' and 'sum' but also 'minus' and 'divide by 2', which have no counterparts in some semirings. Another difference to [1] is that our algorithm does not compute marginals even on trees, e.g., no simple way is known to extract max-marginals from a max-sum diffusion fixed point.

2 Preliminaries

In the sequel, sets are denoted by $\{\cdots\}$ and ordered tuples by $(\cdots)$. Real closed, open and semiopen intervals are $[a, b]$, (a, b) and $[a, b)$, respectively. Non-negative and positive reals are $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_{++} = (0, \infty)$, respectively. The set of all subsets of a set A is denoted by 2^A and the set of all its k-element subsets by $\binom{A}{k}$. Newly defined concepts are typed in **boldface**.

2.1 Commutative Semigroups and Semirings

Definition 1. *A **commutative semigroup** is a set S endowed with a binary operation $+$ that is associative and commutative. We denote a commutative semigroup by $(S, +)$.*

Definition 2. *A **commutative semiring** is a set S endowed with binary operations $+$ and $\times$ such that $+$ is associative and commutative, $\times$ is associative and commutative, and $\times$ distributes over $+$. We denote it by $(S, +, \times)$.*

A commutative semiring can be seen as two commutative semigroups, $(S, +)$ and $(S, \times)$, coupled by distributivity. A commutative semiring may have an **identity element** 1, satisfying $a \times 1 = a$ for all $a \in S$. If an identity element exists, it is unique. A commutative semiring may have a **zero element** 0, satisfying $a + 0 = a$ and $a \times 0 = 0$ for all $a \in S$. If it exists, it is unique.

We will usually abbreviate $a \times b$ by ab. We define $a^n = a \times \cdots \times a$ (n-times) and $na = a + \cdots + a$ (n-times).

2.2 Functions of Blocks of Variables

Let V be a finite set of **variables**. Each variable $i \in V$ attains **states** $x_i \in X_i$, where X_i is a finite **domain** of the variable. A **joint state** (configuration) of variables $A \subseteq V$ is an element x_A of the Cartesian product $X_A = \prod_{i \in A} X_i$. The order of factors in this Cartesian product is given by some fixed total order on V (e.g., for $V = \{1, \ldots, n\}$ we can take the natural arithmetic order).

In the sequel, by the symbol x_A we will always denote a joint state, i.e., the ordered tuple $(x_i)_{i \in A} \in X_A$. Moreover, we adopt the following 'implicit restriction' convention: for $B \subset A$, whenever symbols x_A and x_B occur in the same expression then x_B denotes the restriction of x_A to variables B. This convention is often tacitly used and in fact self-evident: if, e.g., $A = \{1, 2, 3\}$ and $B = \{1, 2\}$, then $x_B = (x_1, x_2)$ is *indeed* the restriction of $x_A = (x_1, x_2, x_3)$ to variables $\{1, 2\}$.

For $A \subseteq V$, consider an S-valued function of variables A, i.e., a function $X_A \to S$. We call A the **scope** of the function and $|A|$ its **arity** (often called *order*). We define the following two operations on such functions:

1) The **combination** of functions $\phi\colon X_A \to S$ and $\psi\colon X_B \to S$ is the function

$$\phi \times \psi\colon X_{A \cup B} \to S, \quad (\phi \times \psi)(x_{A \cup B}) = \phi(x_A) \times \psi(x_B).$$

2) The **marginalization** (also known as *projection*) of a function $\phi\colon X_A \to S$ onto variables $B \subseteq A$ (or over variables $A \setminus B$) is the function

$$\phi|_B\colon X_B \to S, \quad \phi|_B(x_B) = \sum_{x_{A \setminus B}} \phi(x_A).$$

Example 1. Let $A = \{1, 2, 3\}$, $B = \{3, 4\}$, $\phi\colon X_A \to S$, $\psi\colon X_B \to S$. The combination of functions ϕ and ψ is the function $(\phi \times \psi)(x_1, x_2, x_3, x_4) = \phi(x_1, x_2, x_3) \times \psi(x_3, x_4)$. The marginalization of function ϕ onto variables $C = \{2, 3\}$ is the function $\phi|_C(x_2, x_3) = \sum_{x_1} \phi(x_1, x_2, x_3)$.

For two functions $\phi, \psi\colon X_A \to S$, we will write $\phi = \psi$ to denote that $\phi(x_A) = \psi(x_A)$ for all $x_A \in X_A$.

The operators of combination and marginalization are often explicitly used in constraint programming [6], [7], [48]. The set of functions $X_A \to S$ for all $A \subseteq V$ endowed with combination and marginalization is an example of the *valuation algebra* [35], [36], [52], [60]. We state here three of the axioms of the valuation algebra:

1) Combination is associative and commutative.
2) For $\phi\colon X_A \to S$ and $C \subseteq B \subseteq A$,

$$(\phi|_B)|_C = \phi|_C. \tag{2}$$

3) For $\phi\colon X_A \to S$, $\psi\colon X_B \to S$, and $A \cap B \subseteq C \subseteq A \cup B$,

$$(\phi \times \psi)|_C = \phi|_{A \cap C} \times \psi|_{B \cap C}. \tag{3}$$

2.3 Semiring Partition Function

Let $E \subseteq 2^V$ be a hypergraph over V. Let each hyperedge $A \in E$ be assigned a function $f_A\colon X_A \to S$. The partially separable function $\prod_{A \in E} f_A\colon X_V \to S$ can be seen as the combination of the functions f_A and the partition function (1) is the marginal of this function over all the variables, thus it can be written also as $(\prod_{A \in E} f_A)|_\emptyset$.

We refer to each function f_A as a **factor** and to the collection f_A, $A \in E$, as a **network of functions** or simply a **network**. A network can be seen as a map

$$f\colon X_E \to S,$$
$$(A, x_A) \mapsto f_A(x_A),$$

where

$$X_E = \{\, (A, x_A) \mid A \in E,\ x_A \in X_A \,\}$$

is the set of **tuples**. Note the abuse of notation: X_i for $i \in V$, X_A for $A \subseteq V$, and X_E for $E \subseteq 2^V$ denote three different things.

3 Enforcing Marginal Consistency

Here we generalize max-sum diffusion and related concepts to the abstract commutative semiring.

3.1 Equivalent Networks and Reparameterizations

Definition 3. *Let $E, E' \subseteq 2^V$. Networks $f\colon X_E \to S$ and $f'\colon X_{E'} \to S$ are **equivalent** if $\prod_{A \in E} f_A = \prod_{A \in E'} f'_A$.*

Note that the operation $+$ does not appear in the definition, thus network equivalence is defined only with respect to the semigroup $(S, \times)$. Equivalent networks have the same set of variables V and domains X_i, $i \in V$, but they can have different hypergraphs and factors. When $E = E'$, the networks differ only in the values of the factors. In this case, we say that f' is a **reparameterization** of f. Deciding whether two given networks are reparameterizations of each other can be easy or hard, depending on the semigroup $(S, \times)$.

Some reparameterizations are local, in the sense that they are restricted only to a part of the network. The simplest such reparameterization is restricted to a subnetwork containing only two factors, f_A and f_B.

Definition 4. *A **reparameterization of a pair** $\{f_A, f_B\}$ is a change of f_A and f_B that preserves the function $f_A \times f_B$. A reparameterization of any single pair of factors of a network is a **local reparameterization** of the network.*

Example 2. Let $A = \{1, 2\}$ and $B = \{2, 3\}$. A reparameterization of the pair $\{f_A, f_B\}$ is any change of f_{12} and f_{23} that preserves the value $f_{12}(x_1, x_2) \times f_{23}(x_2, x_3)$ for all $x_1 \in X_1$, $x_2 \in X_2$, $x_3 \in X_3$.

Local reparameterizations allow us to traverse through a class of equivalent networks $X_E \to S$. However, some reparameterizations may not be compositions of local reparameterizations. This depends on the semigroup $(S, \times)$. In Section 5.1 we shall discuss properties of reparameterizations for several concrete semigroups $(S, \times)$.

3.2 Enforcing Marginal Consistency of a Pair

Definition 5. *A pair $\{f_A, f_B\}$ is **marginal consistent** if $f_A|_{A\cap B} = f_B|_{A\cap B}$.*

Example 3. For $A = \{1, 2\}$ and $B = \{2, 3\}$, $f_A|_{A\cap B} = f_B|_{A\cap B}$ reads $\sum_{x_1} f_{12}(x_1, x_2) = \sum_{x_3} f_{23}(x_2, x_3)$ for all $x_2 \in X_2$.

Note that marginal consistency is defined only with respect to the semigroup $(S, +)$, the operation $\times$ does not appear in Definition 5.

Definition 6. ***Enforcing marginal consistency** of a pair $\{f_A, f_B\}$ is a reparameterization of this pair that makes it marginal consistent.*

Enforcing marginal consistency of a pair $\{f_A, f_B\}$ means replacing this pair with a solution $\{f'_A, f'_B\}$ to the equation system

$$f'_A \times f'_B = f_A \times f_B, \tag{4a}$$

$$f'_A|_{A\cap B} = f'_B|_{A\cap B}. \tag{4b}$$

In expanded form, this reads

$$\begin{aligned} f'_A(x_A) \times f'_B(x_B) &= f_A(x_A) \times f_B(x_B) && \forall x_{A\cup B} \in X_{A\cup B}, \\ \sum_{x_{A\setminus B}} f'_A(x_A) &= \sum_{x_{B\setminus A}} f'_B(x_B) && \forall x_{A\cap B} \in X_{A\cap B}. \end{aligned}$$

Note that the system in fact breaks into several smaller independent systems, one for each $x_{A\cap B}$.

As we are in the abstract commutative semiring, it is not clear how many (if any) solutions system (4) has and how to find them. It would be desirable to characterize semirings in which the system is solvable and to give an algorithm to find all its solutions in any such semiring. We have not been able to do this.

It is easy to obtain a partial solution to (4). Using (3), marginalizing (4a) onto variables $A \cap B$ yields

$$f'_A|_{A\cap B} \times f'_B|_{A\cap B} = f_A|_{A\cap B} \times f_B|_{A\cap B}.$$

Substituting (4b) into this yields

$$\left(f'_A|_{A\cap B}\right)^2 = \left(f'_B|_{A\cap B}\right)^2 = f_A|_{A\cap B} \times f_B|_{A\cap B}, \tag{5}$$

where, for a function ϕ, we abbreviated $\phi^2 = \phi \times \phi$. Similarly, marginalizing (4a) onto variables A yields

$$f'_A \times f'_B|_{A\cap B} = f_A \times f_B|_{A\cap B}. \tag{6}$$

Equation (5) is solvable if the semiring has a square root. The square root may not be unique, thus (5) can have mutliple solutions. Unfortunately, having $f'_B|_{A\cap B}$ we may not be able to solve (6) for f'_A because the semiring may not have division. In fact, it can happen that (5) is solvable but (4) is not (see Example 16).

We shall see in Section 5 that in many semirings, system (4) has a solution and this solution is often unique.

3.3 Marginal Consistency Algorithm

We now formulate a simple algorithm (Algorithm 1) that iteratively enforces marginal consistency of different pairs of factors in a network $f\colon X_E \to S$. Let these pairs be given by a set $J \subseteq \{\{A, B\} \mid A, B \in E\} = \binom{E}{2}$, which can be seen as an undirected graph over E. The order of updates is given by an infinite sequence $(\{A_k, B_k\})_{k=1}^{\infty}$ of hyperedge pairs, such that each pair $\{A, B\} \in J$ occurs in the sequence an infinite number of times. We call this sequence the **update schedule**.

Algorithm 1. (Marginal Consistency Algorithm.)

for $k = 1, \ldots, \infty$ **do**
  Enforce marginal consistency of pair $\{f_{A_k}, f_{B_k}\}$.
end for

It turns out that in many commutative semirings, the algorithm converges to a fixed point when all pairs $\{f_A, f_B\}$ for $\{A, B\} \in J$ are marginal consistent. It would be desirable to characterize commutative semirings in which this fact holds and provide a rigorous proof. This is difficult in full generality and we have not done it.[1] We discuss convergence of the algorithm for a number of concrete semirings in Section 5.

3.4 Higher Levels of Marginal Consistencies

We say that a network has marginal consistency level $J \subseteq \binom{E}{2}$ if $f_A|_{A\cap B} = f_B|_{A\cap B}$ for all $\{A, B\} \in J$. We now extend this definition to levels higher than $\binom{E}{2}$.

Consider a collection of functions f_A, $A \subseteq V$, i.e., a network over the complete hypergraph 2^V. We say that this network is globally marginal consistent if $f_A = f_V|_A$ for all $A \subseteq V$. But then $f_A|_{A\cap B} = f_B|_{A\cap B}$ for all A, $B \subseteq V$. This immediately follows from (2) because $f_A|_{A\cap B} = (f_V|_A)|_{A\cap B} = f_V|_{A\cap B} = (f_V|_B)|_{A\cap B} = f_B|_{A\cap B}$. By imposing the constraints $f_A|_{A\cap B} = f_B|_{A\cap B}$ for only a subset $J \subseteq \{\{A, B\} | A, B \subseteq V\} = \binom{2^V}{2}$ of all possible pairs $\{A, B\}$, we obtain various levels of marginal consistency, which are necessary (but not sufficient) for global marginal consistency.

When we have a network over a hypergraph $E \subset 2^V$ rather than $E = 2^V$, there is a problem that for some pairs $\{A, B\} \in J$, the function f_A or f_B may not be in the network. In that case, we require that these missing functions exist outside of the network. This leads to the following definitions.

Definition 7. *A network $f\colon X_E \to S$ is **globally marginal consistent** if there exists a function $f_V\colon X_V \to S$ such that $f_A = f_V|_A$ for every $A \in E$. Here the function f_V can either be in the network ($V \in E$) or not ($V \notin E$).*

1. In fact, we cannot speak about convergence yet because we have not defined a metric or topology on the abstract commutative semiring. Endowing a semiring with a topology has been considered in mathematics [24], [26] but this is out of scope of our paper.

Example 4. Let $V = \{1,2,3,4\}$ and $E = \{\{1,3,4\}, \{1,2\}, \{2,3\}\}$. A network $f\colon X_E \to S$ is globally marginal consistent if there exists a function f_{1234} such that $f_{134} = f_{1234}|_{134}$, $f_{12} = f_{1234}|_{12}$, $f_{23} = f_{1234}|_{23}$.

Definition 8. *A network $f\colon X_E \to S$ has **marginal consistency level** $J \subseteq \binom{2^V}{2}$ if there exist functions $f_A\colon X_A \to S$, $A \subseteq V$, $A \notin E$, such that $f_A|_{A\cap B} = f_B|_{A\cap B}$ for all $\{A,B\} \in J$.*

Example 5. Let us have a network with $V = \{1,2,3,4\}$ and $E = \{\{1,2\}, \{1,3\}, \{1,4\}, \{3.4\}, \{2,3\}\}$. Let $J = \{\{\{1,2\}, \{1,3\}\}, \{\{1,2\},\{2,3\}\}, \{\{1,3\},\{2,3\}\}, \{\{1,3,4\},\{1,3\}\}, \{\{1,3,4\},\{1,4\}\}\}$. The network has marginal consistency of level J if $f_{12}|_1 = f_{13}|_1$, $f_{12}|_2 = f_{23}|_2$, $f_{13}|_3 = f_{23}|_3$ and there exists a function f_{134} such that $f_{134}|_{13} = f_{13}$, $f_{134}|_{14} = f_{14}$.

Some levels of marginal consistency are implied by lower levels. E.g., every level is implied by some level $J \subseteq \{\{A,B\} \mid B \subset A \subseteq V\}$ because $f_B|_{A\cap B} = f_B|_{A\cap B}$ is implied by $f_A|_{A\cap B} = f_B|_{A\cap B}$ and $f_B|_{A\cap B} = f_{A\cap B}$.

All possible subsets $J \subseteq \binom{2^V}{2}$ form a partially ordered hierarchy of marginal consistencies. The least element of the hierarchy has level $\emptyset$, the top element has level $\binom{2^V}{2}$. Global marginal consistency has level $\{\{V,A\} \mid A \in E\}$ but, by (2), this already implies the top level $\binom{2^V}{2}$. There are two natural intermediate levels:

1) *hyperedge-to-variable coupling*

$$J = \{\{A,\{i\}\} \mid i \in A \in E\}, \tag{7}$$

2) *hyperedge-to-hyperegde coupling* $J = \binom{E}{2}$.

Algorithm 1 can in general enforce marginal consistency levels not greater than $\binom{E}{2}$. We now describe a simple technique (proposed for max-sum diffusion in [19], [74]) how to enforce higher levels. Suppose our semiring has the identity element 1. We call f_A an **identity factor** if $f_A(x_A) = 1$ for every $x_A \in X_A$ (in short, $f_A = 1$). Suppose we extend E by some $A \notin E$ and set $f_A = 1$. Since $1 \times a = a$ for all $a \in S$, this yields an equivalent network. We call this operation **adding an identity factor** to the network. By Definition 8, adding one or more identity factors (of possibly higher arities) and running Algorithm 1 allows us to enforce an arbitrary level of marginal consistency, at the expense of enlarging the network. We shall see in Section 5.2.1 that in some semirings this is possible even without enlarging the network.

Remark 1. It might seem that adding an identity factor f_A requires to store $|X_A|$ numbers in memory, which may be prohibitive. But this can be alleviated by performing reparameterizations by 'messages' during Algorithm 1, rather than modifying factors 'in place'. This is common in the max-sum semiring [19], [37], [72], [74] but it is possible also in other semirings [73].

Remark 2. Recall that marginal consistency is defined on the semigroup $(S,+)$, so it can be studied independently on the operation $\times$. In the semigroup $(\mathbb{R}_+,+)$ where $+$ is the ordinary addition, the set of globally marginal consistent networks is (up to normalization conditions $f_A|_\emptyset = 1$, $A \in E$) known as the *marginal polytope* and the set of networks with marginal consistency level (7) as the *local marginal polytope* [69]. If E is acyclic, these polytopes are equal [69, Proposition 4.1]. This suggests a question: does this fact extend to other semigroups? Precisely, is it true that for acyclic networks, marginal consistency level (7) implies global marginal consistency? Though for some semigroups the answer is known, in general the question is open.

4 Upper Bound on Partition Function

Max-sum diffusion monotonically decreases an upper bound on the true max-sum partition function. Unlike the partition function, this bound is tractable to compute. At a fixed point of max-sum diffusion, it often happens that the bound is tight (i.e., equal to (1)). In this section, we generalize these concepts to other semirings.

4.1 Canonical Order on a Commutative Semiring

To formulate the upper bound, we first need to define a suitable partial order on the commutative semiring. The standard way of doing this is as follows [26].

Definition 9. *The **canonical preorder** on a commutative semigroup $(S,+)$ is the relation $\le$ on S defined by*

$$a \le b \iff (a = b) \text{ or } (\exists c \in S)(a + c = b). \tag{8}$$

Note that the condition $a = b$ is redundant if the semigroup $(S,+)$ has a neutral element 0. The relation $\le$ is reflexive and transitive, hence a preorder. It naturally extends to the semiring $(S,+,\times)$ as follows.

Theorem 1. *The semiring operations are monotone with respect to $\le$, i.e., for all $a,b,c \in S$ we have*

$$a \le b \implies a + c \le b + c, \quad ac \le bc. \tag{9}$$

Proof. Suppose $a \le b = a + d$. Then $b + c = a + d + c \ge a + c$ and $bc = (a+d)c = ac + dc \ge ac$. □

In general, the relation $\le$ is not antisymmetric, therefore it may not be a partial order. Theorem 2 gives some simple conditions sufficient for $\le$ to be an order or not.

A binary operation $+$ is **idempotent** if $a + a = a$ for all $a \in S$. It is **selective** [26] (also known as *conservative* [11]) if $a + b \in \{a,b\}$ for all $a,b \in S$. Clearly, any selective operation is idempotent. A commutative semigroup $(S,+)$ is **cancellative** if $a + c = b + c$ implies $a = b$ for all $a,b,c \in S$. Cancellation and idempotency exclude each other (by cancellation, $a + a = a$ implies $a = 0$).

Theorem 2. *Let $\le$ be the canonical preorder on $(S,+)$.*

1) If $(S,+)$ is a group, then $\le$ is an equivalence, therefore it is not a partial order.

2) If $+$ is idempotent (i.e., $(S,+)$ is a semilattice), then $\le$ is a partial order and we have

$$a \le b \iff a + b = b. \tag{10}$$

Moreover, $+$ is the least upper bound with respect to $\le$.

3) If $+$ is selective, then $\le$ is a total order. Moreover, $+$ is the maximum with respect to $\le$.

Proof.

1) Suppose $a \leq b = a + c$. Since $(S, +)$ is a group, c has an inverse, therefore $b + (-c) = a \geq b$. This shows that $\leq$ is symmetric.

2) Suppose $a \leq b = a + c$. Then $a + b = a + a + c = a + c = b$, which proves (10). Antisymmetry holds by (10). Proving that $a + b$ is the least upper bound of a, b means proving that $a \leq c$ and $b \leq c$ implies $a + b \leq c$. By (10), this means that $a + c = c$ and $b + c = c$ implies $a + b + c = c$. This is true because $a + b + c = (a + c) + (b + c) = c + c = c$.

3) For any a, b, we have either $a + b = a$ or $a + b = b$. By (10), this means either $b \leq a$ or $a \leq b$. □

If the canonical preorder is antisymmetric, we call it the **canonical order**. We shall see in Section 5 that this is so in many concrete instances.

4.2 The Bound

Now we can introduce a tractable upper bound on the semiring partition function (1).

Theorem 3. *We have*

$$\sum_{x_V} \prod_{A \in E} f_A(x_A) \leq \prod_{A \in E} \sum_{x_A} f_A(x_A) = \prod_{A \in E} f_A|_\emptyset. \tag{11}$$

Proof. Using distributivity, multiply the factors on the right-hand side. This yields all the terms on the left-hand side plus some additional terms. The inequality follows from (8), where c are the additional terms. □

4.3 The Effect of Enforcing Marginal Consistency

Suppose that enforcing marginal consistency of a pair $\{f_A, f_B\}$ is possible, i.e., there exist $\{f'_A, f'_B\}$ satisfying (4). In this section, we show that, under a certain assumption on the semiring, enforcing marginal consistency of the pair never increases the upper bound (11).

Since enforcing marginal consistency of $\{f_A, f_B\}$ affects only the two factors in the bound corresponding to A and B, we want to show that

$$f'_A|_\emptyset \times f'_B|_\emptyset \leq f_A|_\emptyset \times f_B|_\emptyset. \tag{12}$$

From (4b) and using (2) we have

$$f'_A|_\emptyset = \left(f'_A|_{A \cap B}\right)|_\emptyset = \left(f'_B|_{A \cap B}\right)|_\emptyset = f'_B|_\emptyset. \tag{13}$$

Recall that if system (4) has a solution, the functions $f'_A|_{A \cap B} = f'_B|_{A \cap B}$ satisfy (5). Suppose the semiring has a square root. It need not be unique, we only require that *some* unary operation $\sqrt{\cdot}$ exists on S satisfying $(\sqrt{a})^2 = a$ (but not necessarily $\sqrt{a^2} = a$) for all $a \in S$. Then (5) has a solution

$$f'_A|_{A \cap B} = f'_B|_{A \cap B} = \sqrt{f_A|_{A \cap B} \times f_B|_{A \cap B}}, \tag{14}$$

where $\sqrt{\phi}$ denotes component-wise application of $\sqrt{\cdot}$ to a function ϕ. Using (13) and (14), inequality (12) reads

$$\left(\sqrt{f_A|_{A \cap B} \times f_B|_{A \cap B}}|_\emptyset\right)^2 \leq f_A|_\emptyset \times f_B|_\emptyset.$$

Denoting $x_{A \cap B} = i$, $|X_{A \cap B}| = n$, $f_A|_{A \cap B}(x_{A \cap B}) = a_i$, $f_B|_{A \cap B}(x_{A \cap B}) = b_i$, this can be written as

$$\left(\sum_{i=1}^n \sqrt{a_i b_i}\right)^2 \leq \left(\sum_{i=1}^n a_i\right)\left(\sum_{i=1}^n b_i\right). \tag{15}$$

To summarize, we have the following result.

Theorem 4. *Let $\sqrt{\cdot}$ be a unary operation on S that satisfies $(\sqrt{a})^2 = a$ for all $a \in S$ and (15) for all $a_i, b_i \in S, i = 1, \ldots, n$. Then enforcing marginal consistency of any pair of factors (if it is possible) does not increase the upper bound (11).*

Inequality (15) is a form of the *Cauchy-Schwarz inequality* on the semiring. When the square root is unique, we have $(\sqrt{a})^2 = a = \sqrt{a^2}$ for all $a \in S$ and therefore (15) can be written in the more familiar form

$$\langle a, b\rangle^2 \leq \langle a, a\rangle\langle b, b\rangle \tag{16}$$

for all $a, b \in S^n$, where $\langle a, b\rangle = a_1 b_1 + \cdots + a_n b_n$ is the 'inner product' on the semiring. Moreover, it can be verified that (16) is implied by the inequality

$$2ab \leq a^2 + b^2 \tag{17}$$

for all $a, b \in S$ (however, (16) does not imply (17) in some semirings). Let us emphasize that when $\sqrt{a^2} = a$ for all $a \in S$ does not hold, (17) may not imply (15).

Theorem 4 says that, under a reasonable assumption on the semiring, every iteration of Algorithm 1 either decreases the upper bound or keeps it unchanged. Given this result, one might think that the algorithm is nothing more than a (block-)coordinate descent to minimize the upper bound by local reparameterizations. However, this does not fully explain Algorithm 1 because a coordinate descent is expected to *strictly* decrease its objective in every iteration, whereas an iteration of Algorithm 1 can keep the bound unchanged. Yet we cannot omit such iterations because they may modify the network in such a way that some later iterations decrease the bound strictly. This is very obvious in the or-and semiring but it is true also in other semirings.

Of course, monotonic decrease of the bound during Algorithm 1 is neither sufficient nor necessary for its convergence to a fixed point. Although in many semirings these two properties occur together, there can be exceptions (see Section 5.5).

4.4 The Effect of Adding Identity Factors

In Section 3.4 we showed how higher levels of marginal consistency can be achieved by adding identity factors. What effect does this have on the upper bound?

When the operation $+$ is idempotent, adding an identity factor $f_A = 1$ to a network preserves the upper bound because $f_A|_\emptyset = \sum_{x_A} 1 = 1$. Suppose Algorithm 1 is at its fixed point. If we now add one or more identity factors to the network, extend the set J, and run the algorithm again, the upper bound may further decrease. Indeed, this is because the added factors extended the space of reparameterizations that Algorithm 1 can reach by local reparameterizations. Identity factors can be added incrementally during Algorithm 1 in a cutting-plane fashion, similarly as in [19], [74]. This incremental scheme ensures monotonic improvement of the bound.

When $+$ is not idempotent, adding an identity factor may increase the upper bound because $\sum_{x_A} 1 \geq 1$. Therefore, adding identity factors has no obvious advantage in, e.g., the sum-product semiring.

Remark 3. There is another, very obvious technique how to tighten the upper bound arbitrarily at the expense of more computational effort: by merging several factors into one. In the max-sum semiring, this corresponds to using larger subproblems in dual decomposition [32], [39]. E.g., if $\{1,2\},\{2,3\},\{1,3\} \in E$, we can merge binary factors f_{12}, f_{23}, f_{13} into the ternary factor $f_{12} \times f_{23} \times f_{13}$. This decreases $|E|$ by two, keeps the network equivalent, and may decrease the upper bound. Future enforcing of marginal consistency may improve the bound even further. This technique is not limited to semirings with idempotent addition. However, it is less compatible with the concept of local consistencies (e.g., in the or-and semiring it does not lead to k-consistencies, Section 5.2.2).

4.5 When Is the Bound Tight?

In this section, we discuss two natural conditions on a network under which inequality (11) is tight (i.e., holds with equality). One condition will be given by Definition 10 and the other condition is global marginal consistency (Definition 7).

In the max-sum semiring, inequality (11) is tight if all constraints in the network agree on some common global configuration x_V. In [69, Section 8.4], this condition has been called *strong tree agreement*. We say a tuple $(A, x_A) \in X_E$ is **active** if $f_A(x_A) = f_A|_\emptyset$. It is known [12], [61], [72], [74] that deciding the condition leads to the CSP formed by the active tuples. The condition can be formulated for any commutative semiring as follows.

Definition 10. *A network $f\colon X_E \to S$ satisfies **active tuple agreement** if there exists a configuration $x_V \in X_V$ such that the tuple (A, x_A) is active for every $A \in E$.*

Note the implicit restriction (Section 2.2) in Definition 10: x_A is a restriction of x_V to variables A.

Example 6. Let $V = \{1,2,3,4\}$ and $E = \{\{1,3,4\},\ \{1,2\},\ \{2,3\}\}$. A network $f\colon X_E \to S$ satisfies active tuple agreement if there exist $x_1 \in X_1,\ x_2 \in X_2,\ x_3 \in X_3,\ x_4 \in X_4$ such that $f_{134}(x_1,x_3,x_4) = f_{134}|_\emptyset$, $f_{12}(x_1,x_2) = f_{12}|_\emptyset$, $f_{23}(x_2,x_3) = f_{23}|_\emptyset$. Here, e.g., $f_{12}|_\emptyset = \sum_{x_1,x_2} f_{12}(x_1,x_2)$.

Theorem 5. *Active tuple agreement is sufficient for inequality (11) to be tight.*

Proof. The claim follows from the chain

$$\prod_{A\in E} f_A(x_A) \overset{\text{(a)}}{\leq} \sum_{y_V} \prod_{A\in E} f_A(y_A) \overset{\text{(b)}}{\leq} \prod_{A\in E} f_A|_\emptyset \overset{\text{(c)}}{=} \prod_{A\in E} f_A(x_A),$$

where inequality (a) follows from (8), inequality (b) is (11), and equality (c) holds by the assumption. □

Theorem 6. *If the operation $+$ is selective and the semigroup $(S,\times)$ is cancellative, active tuple agreement is necessary for inequality (11) to be tight.*

Proof. First observe that by (9) and by cancellation, $a < b$ and $a' \leq b$ implies $aa' < bb'$ for every $a,b,a',b' \in S$. Suppose that for every $x_V \in X_V$ there exists some $A \in E$ such that $f_A(x_A) < f_A|_\emptyset$. Using (11), this implies $\prod_{A\in E} f_A(x_A) < \prod_{A\in E} f_A|_\emptyset$. Since $+$ is selective, this implies that inequality (11) is strict. □

Let us turn to the second condition, global marginal consistency (Definition 7).

Theorem 7. *If the operation $+$ is selective or the operation $\times$ is idempotent, global marginal consistency is sufficient for inequality (11) to be tight.*

Proof. Suppose a network f is globally marginal consistent, i.e., there is a function f_V such that $f_A = f_V|_A$ for every $A \in E$. Then inequality (11) reads

$$\sum_{x_V} \prod_{A\in E} f_V|_A(x_A) \leq \prod_{A\in E} \sum_{x_A} f_V|_A(x_A). \tag{18}$$

We have

$$\prod_{A\in E} \sum_{x_A} f_V|_A(x_A) = \prod_{A\in E} \sum_{x_V} f_V(x_V) = \Big[\sum_{x_V} f_V(x_V)\Big]^{|E|}.$$

By (8), we have $f_V(x_V) \leq f_V|_A(x_A)$ for every $x_V \in X_V$ and $A \subseteq V$. This simply says that a function cannot be greater than its marginal. Therefore,

$$\sum_{x_V} \prod_{A\in E} f_V|_A(x_A) \geq \sum_{x_V} \prod_{A\in E} f_V(x_V) = \sum_{x_V} \Big[f_V(x_V)^{|E|}\Big].$$

If the operation $\times$ is idempotent, then for any $n \in \mathbb{N}$ and $a \in S$ we have $a^n = a$. If the operation $+$ is selective, it is easy to show that for any $n \in \mathbb{N}$ and any $a_1,\ldots,a_n \in S$ we have $\left(\sum_i a_i\right)^n = \sum_i (a_i)^n$. In both cases, we have $\sum_{x_V} \left[f_V\ (x_V)^{|E|}\right] = \left[\sum_{x_V} f_V(x_V)\right]^{|E|}$. Combining this with (18) yields that inequality (18) is tight. □

We now compare the strength of active tuple agreement and global marginal consistency.

Theorem 8. *If the operation $+$ is selective, global marginal consistency implies active tuple agreement.*

Proof. By global marginal consistency, there is f_V such that $f_A = f_V|_A$ for all $A \in E$. Take any x_V such that $f_V(x_V) = f_V|_\emptyset$. Such x_V exists because $+$ is selective. We have $f_A(x_A) = f_V|_A(x_A) = f_V|_\emptyset$ (note the implicit restriction: x_A is the restriction of x_V). By (2), $f_A|_\emptyset = (f_V|_A)|_\emptyset = f_V|_\emptyset$. We conclude that $f_A(x_A) = f_A|_\emptyset$. □

When $+$ is selective, for every $a_1,\ldots,a_n \in S$ there is some j such that $a_j = \sum_{i=1}^n a_i$. However, such j may not exist when $+$ is not selective. In that case, active tuple agreement is not likely to hold because there may be some $A \in E$ such that no tuple (A, x_A) is active.

On the other hand, it can happen that active tuple agreement does not hold but inequality (11) is tight. For a simple example, take a network with a single unary factor, i.e., $|V| = 1$ and $E = \{V\}$. Trivially, any such network is globally marginal consistent. Let $+$ not be selective and $\times$ be idempotent (as in Example 11). Then constraint agreement may not hold but, by Theorem 7, inequality (11) is tight.

5 Instances of the Framework

Let us now discuss concrete instances of our framework. Since the properties of reparameterizations do not depend on the operation $+$, we find it useful to first discuss reparameterizations in concrete commutative semigroups $(S, \times)$. Then we turn to enforcing marginal consistency in concrete commutative semirings.

Further in Section 5, symbols $+, \times, 0, 1, \sqrt{\cdot}$ will have their ordinary (non-semiring) meaning. We will distinguish semigroups and semirings only up to isomorphism; e.g., the max-sum semiring $(\mathbb{R}, \max, +)$ and the max-product semiring $(\mathbb{R}_{++}, \max, \times)$ are isomorphic (via logarithm).

5.1 Reparameterizations in Concrete Semigroups

Here we discuss reparameterizations in concrete commutative semigroups. We focus on two questions: (i) How hard is to decide whether two given networks are reparameterizations of each other? (ii) Which reparameterizations are compositions of local reparameterizations? We do not try to answer these questions for any commutative semigroup (which we believe would be difficult) but only for selected semigroups of our interest.

5.1.1 *Semilattice* $(S, \wedge)$

A commutative semigroup $(S, \wedge)$ in which the semigroup operation $\wedge$ is idempotent is a *semilattice* [14]. Equivalently, $\wedge$ is the greatest lower bound with respect to some partial order on S. Examples of semilattices are $(\{0,1\}, \min)$, $([0,1], \min)$, and $(2^U, \cap)$ where U is a set and $\cap$ is the set intersection.

Theorem 9. *In every non-trivial ($|S| > 1$) semilattice, deciding whether two networks are reparameterizations of each other is NP-hard.*

Proof. The claim holds for semilattice $(\{0,1\}, \min)$ because deciding whether a given network is a reparameterization of the zero network $f = 0$ (i.e., $f_A(x_A) = 0$ for all $(A, x_A) \in X_E$) is equivalent to the CSP, hence NP-complete. The general case holds because every non-trivial semilattice has a subsemilattice isomorphic to $(\{0,1\}, \min)$, namely $(\{a, a \wedge b\}, \wedge)$ for any $a, b \in S$. □

In a semilattice, not every reparameterization is as a composition of local reparameterizations. This is shown by the following example.

Example 7. Let $(S, \wedge) = (\{0,1\}, \min)$. Let $V = \{1,2,3\}$, $E = \{\{1,2\},\{2,3\},\{1,3\}\}$, $X_1 = X_2 = X_3 = \{1,2\}$, $f_{12}(1,1) = f_{12}(2,2) = f_{23}(1,1) = f_{23}(2,2) = f_{13}(1,2) = f_{13}(2,1) = 1$, $f_{12}(1,2) = f_{12}(2,1) = f_{23}(1,2) = f_{23}(2,1) = f_{13}(1,1) = f_{13}(2,2) = 0$. This network (the 'inconsistent cycle') forms an unsatisfiable CSP, i.e., it is a reparameterization of the zero network. But one easily checks that no local reparameterization is possible.

5.1.2 *Group*

Consider semigroup $(\mathbb{R}, +)$. It is a group, i.e., every element has an inverse. Here, every reparameterization $\{f'_A, f'_B\}$ of a pair $\{f_A, f_B\}$ can be written explicitly as

$$f'_A = f_A + \varphi_{AB} \tag{19a}$$

$$f'_B = f_B - \varphi_{AB} \tag{19b}$$

for some function $\varphi_{AB}\colon X_{A\cap B} \to \mathbb{R}$ (a 'message'). It is known [37, Lemma 6.3], [72, Theorem 3] that in $(\mathbb{R}, +)$, every reparameterization is a composition of a finite number of local reparameterizations and that these local reparameterizations (and whether they exist) can be found in polynomial time. Given (19), this shows that every reparameterization is an affine transformation.

Though this result has been proved for networks with unary and binary constraints f_A, it is natural to conjecture (cf. [74, Section 3.2]) that it extends to any network. Moreover, it is easy to verify that the proofs in [37], [72] apply not only to $(\mathbb{R}, +)$ but to any group.

5.1.3 *Semigroup* $(\mathbb{R} \cup \{-\infty\}, +)$

This semigroup has the sub-semigroup $(\{-\infty, 0\}, +)$ which is a semilattice, thus some reparameterizations are not compositions of local reparameterizations. It has also the sub-semigroup $(\mathbb{R}, +)$, thus some reparameterizations are compositions of a finite number of local reparameterizations, given by (19) for $\varphi_{AB}\colon X_{A\cap B} \to \mathbb{R}$. However, a new phenomenon appears [75], [76]: there exist networks f and f' that are reparameterizations of each other and there is an infinite (but no finite) sequence of local reparameterizations of f that converges to f'.

Example 8. Let $V = \{1,2,3\}$, $E = \{\{1,2\},\{2,3\},\{1,3\}\}$, $X_1 = X_2 = X_3 = \{1,2\}$, $f_{12}(1,1) = f_{12}(2,2) = f_{23}(1,1) = f_{23}(2,2) = f_{13}(1,1) = f_{13}(2,2) = f_{13}(1,2) = 0$, $f_{12}(1,2) = f_{12}(2,1) = f_{23}(1,2) = f_{23}(2,1) = f_{13}(2,1) = -\infty$. Consider the sequence of three local reparameterizations (19) where the functions $\varphi_{12,13}$, $\varphi_{23,12}$, $\varphi_{13,23}$ are given by $\varphi_{12,13}(1) = \varphi_{23,12}(1) = \varphi_{13,23}(1) = 1$, $\varphi_{12,13}(2) = \varphi_{23,12}(2) = \varphi_{13,23}(2) = 0$. This sequence decreases the value $f_{13}(1,2)$ by 1 and keeps all other values unchanged. Repeating the sequence therefore converges to $f_{13}(1,2) = -\infty$. However, no finite sequence of local reparameterizations can set $f_{13}(1,2)$ to $-\infty$.

5.2 Distributive Lattice $(S, \vee, \wedge)$

A commutative semiring $(S, \vee, \wedge)$ in which both operations are idempotent and satisfy the absorption law

$$a \vee (a \wedge b) = a = a \wedge (a \vee b) \tag{20}$$

is a *distributive lattice* [14]. Then $\vee$ is the least upper bound and $\wedge$ is the greatest lower bound with respect to the canonical order $\leq$. Equivalence (10) extends to

$$a \wedge b = a \iff a \leq b \iff a \vee b = b. \tag{21}$$

Example 9. The or-and semiring $(\{0,1\}, \max, \min)$ is a distributive lattice.

Example 10. The max-min semiring $([0,1], \max, \min)$ is a distributive lattice.

Example 11. In the or-and and max-min semirings, both operations are selective hence the canonical order is total. In some inference tasks, our preferences may be given by a partial order that is not total. An example is the distributive lattice $(2^U, \cup, \cap)$ for some set U (or a sublattice of this lattice), where $\leq$ is the inclusion relation $\subseteq$ on 2^U. In

this case, the value (1) is not exactly what we would like to obtain as the result of inference. We would rather like to find maximal elements of the partially ordered set $\{ \bigwedge_{A\in E} f_A(x_A) \mid x_V \in X_V \} \subseteq S$, while (1) is the least upper bound of this set. Discussion on how to find maximal elements of this set is out of scope of our paper. Nevertheless, enforcing marginal consistency may decrease the values of some tuples, i.e., simplify the problem.

As $(S, \wedge)$ is a semilattice, reparameterizations are described by Section 5.1.1. System (4) has the unique solution

$$f'_A = f_A \wedge f_B|_{A\cap B}, \tag{22a}$$

$$f'_B = f_B \wedge f_A|_{A\cap B}. \tag{22b}$$

Let us prove (22a). From (6) and (4b) we obtain $f'_A \wedge f'_A|_{A\cap B} = f_A \wedge f_B|_{A\cap B}$. But the absorption law (20) implies $f'_A \wedge f'_A|_{A\cap B} = f'_A$. By symmetry we get (22b).

The update (22) never increases the value of any tuple because, by (8) and (20), we have $a \wedge b \leq a$ for all $a, b \in S$. It follows that the upper bound (11) never increases. This agrees with the fact that the distributive lattice has a unique square root (the solution to $b \wedge b = a$ is $b = a$) which satisfies (17).

The behavior of Algorithm 1 is similar to local consistency algorithms for the crisp CSP and soft CSPs with idempotent semiring multiplication: it converges in finite time and its fixed point is unique. To formulate this statement more precisely, we extend the canonical order $\leq$ from tuples to networks: for $f, f'\colon X_E \to S$ we define $f \leq f'$ if $f_A(x_A) \leq f'_A(x_A)$ for all $(A, x_A) \in X_E$.

Theorem 10. *Algorithm 1 reaches in a finite number of iterations a fixed point. This fixed point is the greatest one among all fixed points that are not greater than the initial network, therefore it is independent of the update schedule.*

Our proof uses the technique proposed in [2]. It is similar to the proof of the well-known Knapster-Tarski fixed point theorem [14].

Proof. Enforcing marginal consistency of a single pair $\{f_A, f_B\}$ is a function that maps a network to a network. We denote this function by p_{AB}. It has the following properties:

$$\begin{aligned} &p_{AB}(p_{AB}(f)) = p_{AB}(f) && \text{(idempotency)}\\ &p_{AB}(f) \leq f && \text{(intensivity)}\\ &f \leq f' \implies p_{AB}(f) \leq p_{AB}(f') && \text{(monotonicity)} \end{aligned}$$

Note, these are the axioms of a *closure operator* [14].

Algorithm 1 produces a sequence of networks $(f^k)_{k=0}^{\infty}$ defined recurrently by $f^k = p_{A_kB_k}(f^{k-1})$, where f^0 is the initial network and $\{A_k, B_k\}$ is the kth element of the update schedule.

Any value that any tuple can ever attain during the algorithm belongs to the closure of the set of initial values $\{ f^0_A(x_A) \mid (A, x_A) \in X_E \} \subseteq S$ by the operations $\vee$ and $\wedge$. Due to the lattice structure, this closure has finite size. Therefore, by intensivity, the sequence f^k converges in a finite number of iterations.

Suppose a network f satisfies $f \leq f^0$ and $p_{AB}(f) = f$ for all $A, B \in E$. We will prove by induction that $f \leq f^k$ for any k. Suppose $f \leq f^{k-1}$. By monotonicity,

$$f = p_{A_kB_k}(f) \leq p_{A_kB_k}(f^{k-1}) = f^k.$$

We conclude that the convergence point of the sequence f^k is the greatest common fixed point of all the functions p_{AB}, $\{A, B\} \in J$, among all networks not greater than f^0. □

5.2.1 Adding Identity Factors

In Sections 3.4 and 4.4 we discussed how any level of marginal consistency can be achieved by adding identity factors to the network. Assume that our lattice has an identity element, 1. Distributive lattices have the following advantage, not shared by other semirings.

Theorem 11. *Let $E, F \subseteq 2^V$. Let $f\colon X_{E\cup F} \to S$ be a network such that $f_A = 1$ for every $A \in F$. Let $f'\colon X_{E\cup F} \to S$ be the fixed point of Algorithm 1 applied to f. Then*

$$\bigwedge_{A\in E\cup F} f'_A = \bigwedge_{A\in E} f'_A. \tag{23}$$

Proof. The claim is proved by the following chain:

$$\bigwedge_{A\in E} f'_A \overset{\text{(a)}}{\leq} \bigwedge_{A\in E} f_A \overset{\text{(b)}}{=} \bigwedge_{A\in E\cup F} f_A \overset{\text{(c)}}{=} \bigwedge_{A\in E\cup F} f'_A \overset{\text{(d)}}{\leq} \bigwedge_{A\in E} f'_A.$$

In (a) and (d), $\leq$ denotes the componentwise partial order. Inequality (a) holds because the update (22) cannot increase the value of any tuple, (b) holds because $f_A = 1$ for every $A \in F$, (c) holds because enforcing marginal consistency is a reparameterization, and (d) holds because $a \wedge b \leq a$ for every $a, b \in S$. □

The theorem says that if we add one or more identity factors to a network and run Algorithm 1, we can then *remove* the updated identity factors from the network because this yields a reparameterization of the initial network. In other words, identity factors can be added only temporarily and thus the level of marginal consistency can be increased without enlarging the network.

This can be understood also as follows. Adding identity factors extends the space of reparameterizations reachable by local reparameterizations. In a distributive lattice, some reparameterizations cannot be composed of local reparameterizations. Adding identity factors, enforcing marginal consistency, and then removing the updated identity factors means performing a reparameterization of the initial network that may not be reachable by local reparameterizations.

In fact, if we *could* minimize the upper bound (11) over *all* reparameterizations, the bound would become tight. Indeed, we can add the identity factor $f_V = 1$ to the network and run Algorithm 1 with $J = \{ \{V, A\} \mid A \in E \}$. By Theorem 7, this makes inequality (11) tight. Now we remove the factor f_V.

5.2.2 Marginal Consistency in CSP

In the or-and semiring $(\{0, 1\}, \max, \min)$, inequality (11) evaluated at the fixed point of Algorithm 1 says the well-known fact that passing a local consistency test is necessary

for CSP satisfiability. Here, some levels of marginal consistency coincide with some basic local consistencies in CSP [4]. For a network with unary and binary constraints, local marginal consistency is *arc consistency* [46], [4, Section 4]. For any network, marginal consistency of level (7) is *generalized arc consistency* [4, Section 4]. For any network, local marginal consistency is *pairwise consistency* [31], [4, Section 5.4]. Adding appropriate identity constraints of arity less than or equal to k and enforcing pairwise consistency yields (strong) *k-consistency* [21], [4, Section 5.2].

5.3 Semirings of Max-Sum Type

5.3.1 Semiring $(\mathbb{R}, \max, +)$

In this semiring, reparameterizations are affine transformation of f (see Section 5.1.2) and the upper bound (11) is a piecewise-linear convex function of f. Therefore, minimizing the upper bound over all reparameterizations can be formulated as a linear program. This linear program is the natural LP relaxation of problem (1), considered (sometimes in dual form) by many researchers [10], [37], [40], [61], [67], [69].

System (4) has the unique solution

$$f'_A = f_A + \big(f_B|_{A\cap B} - f_A|_{A\cap B}\big)/2, \tag{24a}$$

$$f'_B = f_B + \big(f_A|_{A\cap B} - f_B|_{A\cap B}\big)/2, \tag{24b}$$

which immediately follows from (5) and (6). The semiring has a unique square root (the solution to $b + b = a$ is $b = a/2$), which satisfies conditions (17). Algorithm 1 is known as *max-sum diffusion* [41], [72], [74]. It is firmly believed that max-sum diffusion converges to a fixed point in an infinite number of iterations but this was never proved (a slightly weaker form of convergence has been proved in [57]).

For different update schedules, the algorithm can converge to different fixed points with different values of the bound. Therefore, in general it does not find the minimum upper bound over all reparameterization [37], [72]. Precisely, for some networks the bound cannot be decreased by any single local reparameterization but only by multiple local reparameterizations simultaneously. This is a manifestation of the fact that block-coordinate descent may not find the global minimum of a convex non-smooth function [3]. Note the difference to the distributive lattice, where some reparameterizations cannot be composed of local reparameterizations at all.

5.3.2 Semiring $(\mathbb{R} \cup \{-\infty\}, \max, +)$

This semiring, known as the *tropical semiring* [22], is obtained by adding the zero semiring element $-\infty$ to $(\mathbb{R}, \max, +)$. Minimizing the upper bound over local reparameterizations (19) again leads to a linear program. However, by Section 5.1.3, some reparameterizations are not compositions of local reparameterizations and so this does not yield the minimum upper bound over all reparameterizations. This is not surprising since the semiring has a subsemiring $(\{-\infty, 0\}, \max, +)$ isomorphic to $(\{0, 1\}, \max, \min)$, so this would solve the CSP.

The solution to (4) is unique, given by (24) where the operation '$-$' (minus) is extended from $\mathbb{R}$ to $\mathbb{R} \cup \{-\infty\}$ by defining $a - (-\infty) = -\infty$ for all $a \in \mathbb{R} \cup \{-\infty\}$. The semiring has a unique square root which satisfies (17).

Two stages can be discerned in Algorithm 1. After a finite number of iterations, the set of tuples with values $-\infty$ stops changing, which resolves the 'crisp' part of the problem. Then the algorithm changes only finite tuples, similarly as in the semiring $(\mathbb{R}, \max, +)$.

5.3.3 Max-Sum Semiring with Truncated Addition

This is the semiring $([-1, 0], \max, \otimes)$ where

$$a \otimes b = \max\{-1,\ a + b\}. \tag{25}$$

This semiring is isomorphic to semiring $([0, 1], \max, \otimes')$ where $a \otimes' b = \max\{a + b - 1, 0\}$ is the Łukasziewicz t-norm [34]. The resulting problem (1) is closely related to the *k-weighted CSP* [48, Section 9.2.2].

The semiring has a square root but it is not unique: $b \otimes b = a$ has always a solution but, e.g., for $a = -1$ the solutions are all $b \in [-1, -\frac{1}{2}]$. However, there exists a square root, $b = a/2$, satisfying (15). With this square root, system (4) has a solution[2]

$$f'_A = \max\{\,-1,\ f_A + (f_B|_{A\cap B} - f_A|_{A\cap B})/2\,\}, \tag{26a}$$

$$f'_B = \max\{\,-1,\ f_B + (f_A|_{A\cap B} - f_B|_{A\cap B})/2\,\}. \tag{26b}$$

In experiments on random networks, we observed that Algorithm 1 always converged to a fixed point.

5.3.4 Max-Sum Semiring with Lexicographic Maximum

This is the semiring $(\mathbb{R}^2, \oplus, \otimes)$ where

$$(a_1, a_2) \oplus (b_1, b_2) = \begin{cases} (b_1, b_2) & \text{if } a_1 < b_1, \\ (a_1, \max\{a_2, b_2\}) & \text{if } a_1 = b_1, \\ (a_1, a_2) & \text{if } a_1 > b_1, \end{cases}$$

$$(a_1, a_2) \otimes (b_1, b_2) = (a_1 + b_1, a_2 + b_2).$$

The operation $\oplus$ is the maximum with respect to the lexicographic order on $\mathbb{R}^2$, which is also the canonical order. The solution to (4) is unique, given by (24) where $(\max, +)$ is replaced by $(\oplus, \otimes)$.

The framework can be easily extended from $\mathbb{R}^2$ to $\mathbb{R}^n$.

5.3.5 Adding Identity Factors

As in Section 5.2.1, suppose we add identity factors to a network and then apply Algorithm 1 to the resulting network. Unfortunately, nothing like Theorem 11 holds in max-sum semirings, so we now cannot remove the updated identity factors because this might yield a network that is not equivalent to the initial network. Thus, in general, higher levels of marginal consistency can be achieved only at the expense of increasing the number of factors in the network.

This can be alternatively understood as follows. In semiring $(\mathbb{R}, \max, +)$, every reparameterization can be composed of local reparameterizations. Thus, the only way how to extend the space of reparameterizations reachable by local reparameterizations is to add new

2. Note that we cannot write $f'_A = f_A \otimes (f_B|_{A\cap B} - f_A|_{A\cap B})/2$ in (26a), because $(f_B|_{A\cap B} - f_A|_{A\cap B})/2$ may not be in $[-1, 0]$.

identity factors. This is in contrast with the distributive lattice (Section 5.2.1), where it suffices to add identity factors only temporarily.

5.4 Semirings of Sum-Product Type

5.4.1 Semiring $(\mathbb{R}_{++}, +, \times)$

In this semiring, system (4) has the unique solution

$$f'_A = f_A \times \sqrt{f_B|_{A\cap B} / f_A|_{A\cap B}}, \tag{27a}$$

$$f'_B = f_B \times \sqrt{f_A|_{A\cap B} / f_B|_{A\cap B}}. \tag{27b}$$

The semiring has a unique square root (the only solution to $b^2 = a$ is $b = \sqrt{a}$) which satisfies conditions (17).

The semiring is isomorphic (via logarithm) to semiring $(\mathbb{R}, \oplus, +)$ where

$$a \oplus b = \log(\mathrm{e}^a + \mathrm{e}^b) \tag{28}$$

is the *log-sum-exp operation*. In this semiring, reparameterizations are affine transformations of f and the upper bound (11) is a smooth convex function of f. Algorithm 1 is a block-coordinate descent method to minimize this function over reparameterizations and therefore it converges to its global minimum [3]. It can be shown [76] that the fixed point of the algorithm is unique.

This algorithm is not widely known, it was proposed in [76, Section 6] and also [47] noticed that max-sum diffusion can be formulated in the sum-product semiring. The minimum upper bound is usually very loose, therefore not useful to approximate (1). Even for acyclic E, the bound is not exact and no finite algorithm is known to compute the fixed point. The algorithm can be seen as a very simple representant of convergent message passing algorithms to minimize convex free energies [29], [30], [71], [69, Section 7], which can provide better bounds.

5.4.2 Semiring $(\mathbb{R}_+, +, \times)$

This semiring is obtained by adding the zero semiring element 0 to $(\mathbb{R}_{++}, +, \times)$. Since the semigroup $(\mathbb{R}_+, \times)$ is isomorphic to $(\mathbb{R} \cup \{-\infty\}, +)$, reparameterizations are described by Section 5.1.3. System (4) has a unique solution, given by (27) where we define $a/0 = 0$ for all $a \in \mathbb{R}_+$.

5.4.3 Relation to the Max-Sum Semiring

Define the operation $\oplus_t$ by

$$a \oplus_t b = \frac{(ta) \oplus (tb)}{t} = \frac{\log(\mathrm{e}^{ta} + \mathrm{e}^{tb})}{t}. \tag{29}$$

For every finite t, $(\mathbb{R}, \oplus_t, +)$ is a semiring isomorphic to $(\mathbb{R}, \oplus, +)$. In the limit $t \to \infty$, the operation $\oplus_t$ becomes $\max$. The semiring $(\mathbb{R}, \max, +)$ is no longer isomorphic to $(\mathbb{R}, \oplus, +)$. This process is known as *tropicalization* [22], *dequantization* [44] or the *zero temperature limit* [49].

We said in Section 5.3.1 that in semiring $(\mathbb{R}, \max, +)$, Algorithm 1 in general does not find the minimum upper bound over all reparameterizations. However, the sequence of fixed points of the algorithm in semirings $(\mathbb{R}, \oplus_t, +)$ for increasing t converges to the *optimal* upper bound [76]. This is the core of proximal projection methods with entropy distances for exactly solving the LP relaxation mentioned in Section 5.3.1 [53].

5.4.4 Application to CSP

Although in the sum-product semiring the minimum upper bound is usually very loose, in [75] we described an interesting situation when this bound *is* useful. We now revisit this result in the semiring context.

Let $f\colon X_E \to \{0,1\}$. In semiring $(\{0,1\}, \max, \min)$, expression (1) equals 1 if the CSP represented by f is satisfiable and 0 if not. In semiring $(\mathbb{R}_+, +, \times)$, expression (1) counts the number of solutions to the CSP represented by f. This problem is known as the *counting CSP* (#CSP) [9]. Note that $(\{0,1\}, +, \times)$ is not a semiring because the set $\{0,1\}$ is not closed under addition.

Let $U^{\mathrm{or,and}} \in \{0,1\}$ be the upper bound (11) at the fixed point of Algorithm 1 applied to the network f in semiring $(\{0,1\}, \max, \min)$. Let $U^{+,\times} \in \mathbb{R}_+$ be the upper bound at the fixed point of Algorithm 1 applied to f in semiring $(\mathbb{R}_+, +, \times)$. Clearly, $U^{\mathrm{or,and}} = 1$ is necessary for the CSP represented by f to be satisfiable (see Section 5.2.2). But $U^{+,\times} \geq 1$ is also necessary for this CSP to be satisfiable, requiring that the CSP has at least one solution.

The update rules in the semirings $(\{0,1\}, \max, \min)$ and $(\mathbb{R}_+, +, \times)$ treat zero tuples in the same way: if the former sets some tuple to zero, so does the latter.[3] It follows that $U^{\mathrm{or,and}} = 0$ implies $U^{+,\times} = 0$. Equivalently, $U^{+,\times} > 0$ implies $U^{\mathrm{or,and}} = 1$. However, the opposite implication does not hold: there are CSP instances for which $U^{\mathrm{or,and}} = 1$ and $U^{+,\times} = 0$ [75]. Therefore, Algorithm 1 in semiring $(\mathbb{R}_+, +, \times)$ yields a *strictly stronger* condition necessary for CSP satisfiability than in semiring $(\{0,1\}, \max, \min)$.

The algorithm has the drawback that when reparameterizations are represented by messages, some messages can grow unbounded [75]. This is a manifestation of the phenomenon described in Example 8.

5.5 Expectation Semiring

Expectation semirings, introduced in [18], [43], are dissimilar to any semiring we discussed above. An example is the commutative semiring $(\mathbb{R}_{++} \times \mathbb{R}, \oplus, \otimes)$ where

$$(a_1, a_2) \oplus (b_1, b_2) = (a_1 + b_1, a_2 + b_2),$$
$$(a_1, a_2) \otimes (b_1, b_2) = (a_1 b_1, a_1 b_2 + b_1 a_2).$$

As noted in [25, Example 7.3], this semiring can be seen as the semiring of matrices $\begin{bmatrix} a_1 & a_2 \\ 0 & a_1 \end{bmatrix}$ with the usual matrix addition and product. These matrices are positive definite, hence the semiring has multiplicative inverse and unique square root. Therefore, the solution to (4) can be written as (27) where $\times, /, \sqrt{\cdot}$ are matrix operations.

The canonical preorder (8) is not antisymmetric: e.g., we have both $(0,-1) \leq (0,1)$ and $(0,1) \leq (0,-1)$. Therefore the concepts of upper bound and its decrease are meaningless. Despite this, we observed in experiments on random networks that Algorithm 1 always converged to a fixed point.

3. A similar observation was made for loopy belief propagation [16], showing that it treats zero tuples the same way as the arc consistency algorithm.

5.6 Semirings That Do Not Admit Enforcing Marginal Consistency

Not every commutative semiring allows enforcing marginal consistency. For that, system (4) has to be solvable. Furthermore, it is reasonable to require that the canonical preorder $\leq$ is antisymmetric and the semiring satisfies the conditions of Theorem 4. Here we give examples of semirings that violate some of these requirements.

Example 12. In semiring $(\mathbb{R}, +, \times)$, the semigroup $(\mathbb{R}, +)$ is a group, therefore by Theorem 2 the relation $\leq$ is an equivalence rather than a partial order.

Example 13. In semiring $(\mathbb{N}, \max, +)$, system (4) is not always solvable. Indeed, this semiring does not have a square root because $a + a = b$ has no solution for odd b.

Example 14. Semiring $(2^U, \cup, \otimes)$ where 2^U is the set of all subsets of a vector space U, $\cup$ is the set union and $a \otimes b = \{\, x + y \mid x \in a,\ y \in b \,\}$ is the Minkowski set sum. This semiring does not have a square root: e.g., for $U = \mathbb{R}$, there is no $a \subseteq \mathbb{R}$ satisfying $a \otimes a = \{1, 2\}$.

Example 15. Semiring $(S, \oplus, \otimes)$ where S is the set of all convex subsets of a vector space U, $a \oplus b = \mathrm{conv}(a \cup b)$, and $a \otimes b = \{\, x + y \mid x \in a,\ y \in b \,\}$. This semiring has a unique square root: the solution to $b \otimes b = a$ is $b = \{\, x/2 \mid x \in a \,\}$. Thus equation (5) always has a unique solution. However, system (4) may not have a solution. This can happen already in the simple case $U = \mathbb{R}$, i.e., the elements of S are intervals. E.g., take $A = \{1, 2\}$, $B = \{1\}$, $X_1 = \{1\}$, $X_2 = \{1, 2\}$, $f_1(1) = \{0\}$, $f_{12}(1, 1) = \{-2\}$, $f_{12}(1, 2) = \{2\}$. The solution to (5) is $f'_1(1) = \mathrm{conv}\{-1, 1\}$. But (4a) requires that $f'_{12}(1, 1) \otimes f'_1(1) = f_{12}(1, 1) \otimes f_1(1) = \{-2\} \otimes \{0\} = \{-2\}$. Clearly, there is no such $f'_{12}(1, 1)$.

Example 16. [5, Section 2.4.5] Semiring $(\mathbb{R}^2, \oplus, \otimes)$ where

$$(a_1, a_2) \oplus (b_1, b_2) = \begin{cases} (b_1, b_2) & \text{if } a_1 < b_1, \\ (a_1, \max\{a_2, b_2\}) & \text{if } a_1 = b_1, \\ (a_1, a_2) & \text{if } a_1 > b_1, \end{cases}$$

$$(a_1, a_2) \otimes (b_1, b_2) = \begin{cases} (a_1, a_2) & \text{if } a_1 < b_1, \\ (a_1, a_2 + b_2) & \text{if } a_1 = b_1, \\ (b_1, b_2) & \text{if } a_1 > b_1. \end{cases}$$

The operation $\oplus$ is the same as in Section 5.3.4. The solution to the equation $(b_1, b_2) \otimes (b_1, b_2) = (a_1, a_2)$ is $(b_1, b_2) = (a_1, a_2/2)$, thus the semiring has a unique square root. Therefore equation (5) always has a unique solution. However, system (4) may not have a solution. This happens, e.g., for $A = \{1, 2\}$, $B = \{1\}$, $X_1 = \{1\}$, $X_2 = \{1, 2\}$, $f_1(1) = (0, 3)$, $f_{12}(1, 1) = (0, 2)$, $f_{12}(1, 2) = (2, 0)$.

For semirings that do not allow enforcing marginal consistency it is an interesting open question whether enforcing marginal consistency only approximately can yield useful upper bounds.

6 Summary

Our goal in this paper has been to theoretically investigate the simple algorithm defined in Section 3, first for the abstract commutative semiring and then for several concrete semirings. Let us review the algorithm once again. We are given a commutative semiring $(S, +, \times)$, a hypergraph $E \subseteq 2^V$, and a collection of functions $f_A \colon X_A \to S$, $A \in E$. The algorithm visits different pairs $\{f_A, f_B\}$ and changes every pair such that $f_A|_{A \cap B} = f_B|_{A \cap B}$ while preserving the function $f_A \times f_B$. In many semirings, repeating this operation converges to a fixed point when $f_A|_{A \cap B} = f_B|_{A \cap B}$ holds for *each* pair $\{f_A, f_B\}$. Every iteration either decreases or preserves the upper bound $\prod_{A \in E} f_A|_\emptyset$ on the semiring partition function $\left(\prod_{A \in E} f_A\right)|_\emptyset$.

We have extended this basic algorithm to achieve higher levels of consistency. This is done by adding identity factors $f_A = 1$ (typically of higher arities) to the network, which preserves the function $\prod_{A \in E} f_A$ but extends the set of reachable reparameterizations and thus may enable further improvement of the bound. This yields a hierarchy of consistencies of increasingly higher levels, necessary for global marginal consistency. For a wide class of semirings, global marginal consistency suffices for the upper bound to be tight.

We have discussed the properties of the algorithm in a number of concrete semirings. In a distributive lattice, the algorithm converges in finite time and its fixed point is unique. An example of a distributive lattice is the or-and semiring, for which various levels of marginal consistency correspond to several classical local consistencies in CSP. In semirings of max-sum type, the algorithm converges in an infinite time and its fixed point depends on the update schedule. It is known as max-sum diffusion. In semirings of sum-product type, the algorithm converges in infinite time and its fixed point is unique. In the log-domain, the algorithm minimizes a smooth convex function by block-coordinate descent. It is a simple example of message passing algorithms with convex free energies.

Finally, let us remark that our paper is relevant to two disciplines, pattern recognition and constraint programming, which use different terminology and communicate little with each other. We hope that our paper will narrow this undesirable interdisciplinary gap.

Acknowledgments

The author was supported by the Czech Science Foundation under project P202/12/2071 and by the European Commission under project FP7-ICT-270138.

References

[1] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 325–343, Mar. 2000.

[2] Krzysztof R. Apt, "The rough guide to constraint propagation," in *Proc. Conf. Principles Practice Constraint Program.*, 1999, pp. 1–23.

[3] Dimitri P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[4] C. Bessiere, "Constraint propagation," in *Handbook of Constraint Programming*, Rossi et al. Eds. Amsterdam, The Netherlands: Elsevier, 2006.

[5] S. Bistarelli, U. Montanari, F. Rossi, T. Schiex, G. Verfaillie, and H. Fargier, "Semiring-based CSPs and valued CSPs: Frameworks, properties, and comparison," *Constraints*, vol. 4, no. 3, pp. 199–240, 1999.

[6] S. Bistarelli, *Semirings for Soft Constraint Solving and Programming*. New York, NY, USA: Springer Verlag, 2004.
[7] S. Bistarelli, U. Montanari, and F. Rossi, "Semiring-based constraint satisfaction and optimization," *J. ACM*, vol. 44, no. 2, pp. 201–236, 1997.
[8] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[9] A. A. Bulatov and V. Dalmau, "Towards a dichotomy theorem for the counting constraint satisfaction problem," *J. Inf. Comput.*, vol. 205, no. 5, pp. 651–678, 2007.
[10] C. Chekuri, S. Khanna, J. Naor, and L. Zosin, "A linear programming formulation and approximation algorithms for the metric labeling problem," *SIAM J. Discrete Math.*, vol. 18, no. 3, pp. 608–625, 2005.
[11] D. Cohen and P. Jeavons, "The complexity of constraint languages," in *Handbook of Constraint Programming*. Amsterdam, The Netherlands. Elsevier, ch. 8, 2006.
[12] M. C. Cooper, S. de Givry, M. Sanchez, T. Schiex, M. Zytnicki, and T. Werner, "Soft arc consistency revisited," *Artif. Intell.*, vol. 174, nos. 7/8, pp. 449–478, 2010.
[13] M. Cooper and T. Schiex, "Arc consistency for soft constraints," *Artif. Intell.*, vol. 154, nos. 1/2, pp. 199–227, 2004.
[14] B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
[15] R. Debruyne and C. Bessière, "Domain filtering consistencies," *J. Artif. Intell. Res.*, vol. 14, pp. 205–230, 2001.
[16] R. Dechter and R. Mateescu, "A simple insight into iterative belief propagation's success," in *Proc. Conf. Uncertainty Artif. Intell.*, 2003, pp. 175–183.
[17] D. Dubois, H. Fargier, and H. Prade, "The calculus of fuzzy restrictions as a basis for flexible constraint satisfaction," in *Proc. IEEE 2nd Int. Conf. Fuzzy Syst.*, 1993, pp. 1131–1136.
[18] J. Eisner, "Parameter estimation for probabilistic finite-state transducers," in *Proc. Annu. Meeting Assoc. Comput. Linguistics*, 2002, pp. 1–8.
[19] V. Franc, S. Sonnenburg, and T. Werner, "Cutting plane methods in machine learning," in *Optimization for Machine Learning*. S. Sra, S. Nowozin, and S. J. Wright, Eds. Cambridge, MA, USA: MIT Press, 2012.
[20] E. Freuder and A. K. Mackworth, "Constraint satisfaction: An emerging paradigm," in *Handbook of Constraint Programming*, Rossi et al. Eds. Elsevier: Amsterdam, The Netherlands, 2006.
[21] E. C. Freuder, "Synthesizing constraint expressions," *Commun. ACM*, vol. 21, no. 11, pp. 958–966, 1978.
[22] S. Gaubert, "Methods and applications of (max,+) linear algebra," Institut National Recherche Informatique Automatique (INRIA), Tech. Rep. 3088, 1997.
[23] A. Globerson and T. Jaakkola, "Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations," in *Proc. Neural Inf. Process. Syst.*, 2008, pp. 553–560.
[24] J. S. Golan, *Semirings and Their Applications*. Norwell, MA, USA: Kluwer, 1999.
[25] J. S. Golan, *Semirings and Affine Equations over Them: Theory and Applications*. Norwell, MA, USA: Kluwer, 2003.
[26] M. Gondran and M. Minoux, *Graphs, Dioids and Semirings: New Models and Algorithms*. New York, NY, USA: Springer, 2008.
[27] R. M. Haralick and L. G. Shapiro, "The consistent labeling problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 173–184, Feb. 1979.
[28] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Reading, MA, USA: Addison-Wesley, 1992.
[29] T. Hazan and A. Shashua, "Convergent message-passing algorithms for inference over general graphs with convex free energies," in *Proc. Conf. Uncertainty Artif. Intell.*, 2008, pp. 264–273.
[30] T. Heskes, "Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies," *J. Artif. Intell. Res.*, vol. 26, pp. 153–190, 2006.
[31] P. Janssen, P. Jegou, B. Nouguier, and M. C. Vilarem, "A filtering process for general constraint satisfaction problems: Achieving pairwise consistency using an associated binary representation," in *Proc. IEEE Workshop Tools Artif. Intell.*, 1989, pp. 420–427.
[32] J. K. Johnson, D. M. Malioutov, and A. S. Willsky, "Lagrangian relaxation for MAP estimation in graphical models," in *Proc. Allerton Conf. Commun., Control Comput.*, 2007.
[33] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, and C. Rother, "A comparative study of modern inference techniques for discrete energy minimization problem," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1328–1335.
[34] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. New York, NY, USA: Springer, 2000.
[35] J. Kohlas, *Information Algebras: Generic Structures for Inference*. New York, NY, USA: Springer-Verlag, 2003.
[36] J. Kohlas and N. Wilson, "Semiring induced valuation algebras: Exact and approximate local computation algorithms," *Artif. Intell.*, vol. 172, no. 11, pp. 1360–1399, 2008.
[37] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.
[38] N. Komodakis and N. Paragios, "Beyond loose LP-relaxations: Optimizing MRFs by repairing cycles," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 806–820.
[39] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 531–552, Mar. 2011.
[40] A. Koster, C. P. M. van Hoesel, and A. W. J. Kolen, "The partial constraint satisfaction problem: Facets and lifting theorems," *Oper. Res. Lett.*, vol. 23, nos. 3–5, pp. 89–97, 1998.
[41] V. A. Kovalevsky and V. K. Koval, "A diffusion algorithm for decreasing the energy of the max-sum labeling problem," Glushkov Institute of Cybernetics, Kiev, USSR. Unpublished, approx. 1975.
[42] P. Kumar and P. H. S. Torr, "Efficiently solving convex relaxations for MAP estimation," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 680–687.
[43] Z. Li and J. Eisner, "First- and second-order expectation semirings with applications to minimum-risk training on translation forests," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2009, pp. 40–51.
[44] G. L. Litvinov, "Maslov dequantization, idempotent and tropical mathematics: A brief introduction," *J. Math. Sci.*, vol. 140, no. 3, pp. 426–444, 2007.
[45] A. Mackworth, "Constraint satisfaction," in *Encyclopaedia of Artificial Intelligence*. New York, NY, USA: Wiley, 1991, pp. 285–292.
[46] A. K. Mackworth, "Consistency in networks of relations," *Artif. Intell.*, vol. 8, no. 1, pp. 65–73, 1977.
[47] T. Meltzer, A. Globerson, and Y. Weiss, "Convergent message passing algorithms: A unifying view," in *Proc. Conf. Uncertainty Artif. Intell.*, 2009, pp. 393–401.
[48] P. Meseguer, F. Rossi, and T. Schiex, *Soft Constraints*, in *Handbook of Constraint Programming*, Rossi et al. Eds. Elsevier: Amsterdam, The Netherlands, 2006.
[49] M. Mézard and A. Montanari, *Information, Physics, and Computation*. New York, NY, USA: Oxford Univ. Press, 2009.
[50] U. Montanari, "Networks of constraints: Fundamental properties and application to picture processing," *Inform. Sci.*, vol. 7, pp. 95–132, 1974.
[51] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann, 1988.
[52] M. Pouly and J. Kohlas, *Generic Inference—A unifying Theory for Automated Reasoning*. New York, NY, USA: Wiley, 2011.
[53] P. Ravikumar, A. Agarwal, and M. J. Wainwright, "Message-passing for graph-structured linear programs: Proximal methods and rounding schemes," *J. Mach. Learn. Res.*, vol. 11, pp. 1043–1080, 2010.
[54] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Systems, Man, Cybern.*, vol. SMC-6, no. 6, pp. 420–433, Jun. 1976.
[55] F. Rossi, P. van Beek, and T. Walsh, Eds. *Handbook of Constraint Programming*. Elsevier. Amsterdam, The Netherlands, 2006.
[56] T. Schiex, H. Fargier, and G. Verfaillie, "Valued constraint satisfaction problems: Hard and easy problems," in *Proc. Int. Joint Conf. Artif. Intell.*, 1995, pp. 631–637.
[57] M. Schlesinger and K. Antoniuk, "Diffusion algorithms and structural recognition optimization problems," *Cybern. Syst. Anal.*, vol. 47, pp. 175–192, 2011.
[58] M. I. Schlesinger, *Matematicheskie sredstva obrabotki izobrazheniy (Mathematical Tools of Image Processing)*. Kiev, Ukraine: Naukova Dumka, 1989, in Russian.

[59] M. I. Schlesinger and B. Flach, "Some solvable subclasses of structural recognition problems," in *Proc. Czech Pattern Recognit. Workshop*, 2000, pp. 55–62.
[60] P. P. Shenoy and G. Shafer, "Axioms for probability and belief-function proagation," in *Proc. Conf. Uncertainty Artif. Intell.*, 1990, pp. 169–198.
[61] M. I. Shlezinger, "Syntactic analysis of two-dimensional visual signals in noisy conditions," *Cybernetics Syst. Anal.*, vol. 12, no. 4, pp. 612–628, 1976, Translation from Russian.
[62] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss, "Tightening LP relaxations for MAP using message passing," in *Proc. Conf. Uncertainty Artif. Intell.*, 2008, pp. 503–510.
[63] D. Sontag, "Approximate inference in graphical models using LP relaxations," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2010.
[64] D. Sontag, A. Globerson, and T. Jaakkola, "Introduction to dual decomposition for inference," in *Optimization for Machine Learning*, Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, Eds. Cambridge, MA, USA: MIT Press, 2012.
[65] D. Sontag and T. Jaakkola, "New outer bounds on the marginal polytope," in *Proc. Neural Inf. Process. Syst.*, 2007, pp. 1393–1400.
[66] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.
[67] J. Thapper and S. Živný, "The power of linear programming for valued CSPs," in *Proc. Symp. Found. Comput. Sci.*, 2012, pp. 669–678.
[68] S. Živný, *The Complexity of Valued Constraint Satisfaction Problems*. Cognitive Technologies. New York, NY, USA: Springer, 2012.
[69] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends Mach. Learn.*, vol. 1, nos. 1/2, pp. 1–305, 2008.
[70] D. L. Waltz, "Generating semantic descriptions from drawings of scenes with shadows," Massachusetts Institute of Technology, Cambridge, MA, USA, Tech. Rep. AI271, 1972.
[71] Y. Weiss, C. Yanover, and T. Meltzer, "MAP estimation, linear programming and belief propagation with convex free energies," in *Proc. Conf. Uncertainty Artif. Intell.*, 2007, pp. 416–425.
[72] T. Werner, "A linear programming approach to max-sum problem: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1165–1179, Jul. 2007.
[73] T. Werner, "Marginal consistency: Unifying constraint propagation on commutative semirings," in *Proc. Int. Workshop Preferences Soft Constraints*, 2008, pp. 43–57.
[74] T. Werner, "Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1474–1488, Aug. 2010.
[75] T. Werner, "How to compute primal solution from dual one in MAP inference in MRF?" *Control Syst. Comput.*, vol. 2, pp. 35–45, Mar./Apr., 2011.
[76] T. Werner and A. Shekhovtsov, "Unified framework for semiring-based arc consistency and relaxation labeling," in *Proc. 12th Comput. Vis. Winter Workshop*, St. Lambrecht, Austria, Feb. 2007, pp. 27–34.

Tomáš Werner received the PhD degree from the Czech Technical University in Prague, in 1999. Since then, he was a researcher at the Center for Machine Perception at the same university. In 2000-2001, he spent as a postdoc in the Visual Geometry Group at the Oxford University, United Kingdom. His interests include multiple view geometry, graphical models, optimization, and constraint satisfaction.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.