Computer Vision Winter Workshop 2007, Michael Grabner, Helmut Grabner (eds.)
St. Lambrecht, Austria, February 6–8
Graz Technical University

Unified Framework for Semiring-based Arc Consistency and Relaxation Labeling

Tomáš Werner and Alexander Shekhovtsov

Center for Machine Perception, Dept. of Cybernetics, Czech Technical University, Prague

Abstract *Constraint Satisfaction Problem (CSP), including its soft modifications, is ubiquitous in artificial intelligence and related fields. In computer vision and pattern recognition, the crisp CSP is more known as the consistent labeling problem and certain soft CSPs as certain inference problems in Markov Random Fields. Many soft CSPs can be seen as special cases of the semiring-based CSP (SCSP), using two abstract operations that form a semiring.*

A fundamental concept to tackle the CSP, as well as the SCSPs with idempotent semiring multiplication, are arc consistency algorithms, also known as relaxation labeling. Attempts have been made to generalize arc consistency for soft CSPs with non-idempotent semiring multiplication. We achieve such generalization by generalizing max-sum diffusion of Kovalevsky and Koval, used to decrease Schlesinger's upper bound on the max-sum CSP. We formulate the proposed generalized arc consistency in the semiring framework. Newly, we introduce sum-product arc consistency and give its relation to max-sum arc consistency and optimal max-sum arc consistency.

1 Introduction

The[1] constraint satisfaction problem (CSP) [39, 26] seeks to find states of discrete variables that satisfy a given set of constraints (relations). This formulation is too rough for many applications and the attention is turning to soft CSPs, where one seeks to optimize soft constraints rather than satisfy crisp ones. A number of formulations of soft CSPs have been proposed, such as the max-CSP, the fuzzy, partial, weighted, and probabilistic CSP. References and terminology can be found e.g. in [5, 2, 27].

Many properties of soft CSPs can be studied in a unified algebraic framework. This is done by introducing the semiring-based CSP (SCSP), which uses two abstract operations that form a commutative semiring [35, 2, 5, 33].

The constraints research is multidisciplinary. This brings the problem that closely related or identical things may have several names. Thus, in computer vision the CSP is more known as the consistent labeling problem by Waltz [39] and certain soft CSPs as certain inference tasks in Markov random fields (MRFs) or undirected graphical models [24, 38, 36]. Unlike in AI, in computer vision much larger but typically binary and sparse instances occur.

Inference tasks in MRFs have recently attracted a lot of attention. In particular, maximum *a posteriori* (MAP) inference in MRFs (which we will call the **max-sum CSP** in this paper) finds applications in low level computer vision tasks, such as image segmentation, matching and restoration, and 3D reconstruction. It is sometimes called energy minimization [34, 20]. Its special cases are dynamic programming and max-flow/min-cut algorithms [21].

Ubiquitous in CSP is the concept of arc consistency, first proposed probably by Waltz [39]. It is the simplest one of local consistencies, surveyed e.g. in [12]. The algorithm achieving arc consistency is also known as (discrete) relaxation labeling by Rosenfeld *et al*. [29]. Considerable effort has been devoted to generalizing arc consistency to soft constraints [3, 10, 9, 4]. The earliest such work seems to be done already by Rosenfeld [29] for the fuzzy CSP. For the SCSPs with idempotent semiring multiplication (typically, when the semiring is a distributive lattice), local consistency algorithms are known to converge in a finite number of local operations and the result does not depend on their order [5].

We contribute to an old and not widely known approach by Schlesinger *et al*. [31] to the max-sum CSP, based on linear programming relaxation. We surveyed this approach in [41, 40]; we recommend reading [41, 40] prior to our paper. We will be interested especially in the max-sum diffusion algorithm by Kovalevsky and Koval [23] (independently by Flach [13]), which decreases Schlesinger's upper bound on the max-sum CSP. Max-sum diffusion resembles belief propagation [28] but it is in fact fundamentally different. It computes the same approximation as the sequential tree-reweighted message passing (TRW-S) by Wainwright, Kolmogorov *et al*. [19, 36], which is the most promising algorithm for the approximative max-sum CSP [34].

We show that arc consistency algorithms for SCSPs on a distributive lattice can be seen as the same thing as max-sum diffusion. We conjecture that in the same manner, arc consistency can be generalized even for a wider class of SCSPs.

To extend the list, we newly introduce sum-product arc consistency. We show how the max-sum arc consistency can be obtained as a limit ('tropicalization', 'zero temperature limit') of the sum-product arc consistency, which corresponds to replacing the log-sum-exp function with ordinary maximum. This sheds light on the question of non-optimal max-sum arc consistent states, which is of great practical interest because it aims at obtaining better upper bounds on the max-sum CSP. Importantly, we give a network algorithm to test for optimality of max-sum arc consistency.

We will denote a set by $\{\cdots\}$, an ordered tuple by $\langle\cdots\rangle$, non-negative reals by $\mathbb{R}_{0+}$, and positive reals by $\mathbb{R}_{++}$.

[1] We thank grants FP6-IST-004176 (COSPAL) and INTAS 04-77-7347 (PRINCESS) for kind support and Mirko Navara for valuable comments.

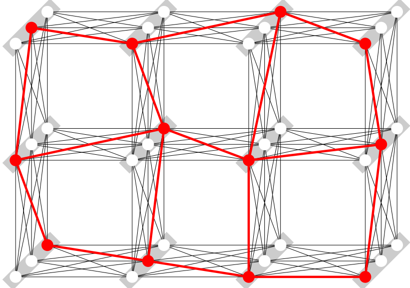

Figure 1: An example of a binary CSP. Graph $\langle T, E\rangle$ is the 3×4 grid graph, thus it has $|T| = 12$ nodes. There are $|X| = 3$ labels. An example of labeling $\mathbf{x}$ is emphasized.

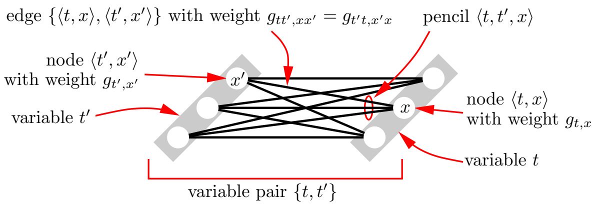

Figure 2: Illustration of our notation and terminology.

2 Semiring-based Constraint Satisfaction

For the sake of clarity, we will consider only binary CSPs and use notation more usual in computer vision than in the CSP literature. However, the results can be straightforwardly extended for problems with any arity.

Let T be a finite set of **variables** and $E \subseteq \binom{T}{2}$ a set of **variable pairs**, thus $\langle T, E\rangle$ is an undirected graph. By $N_t = \{\, t' \mid \{t,t'\} \in E \,\}$ we will denote the neighbors of variable $t \in T$. Each variable $t \in T$ is assigned a single **label** $x_t \in X$, where the label domain X is a finite set. **Labeling** $\mathbf{x} = \langle\, x_t \mid t \in T \,\rangle \in X^T$ assigns label x_t to each variable t. This is illustrated in Figure 1.

Let $\langle T \times X, E_X\rangle$ be an undirected graph with edges $E_X = \{\, \{\langle t,x\rangle, \langle t',x'\rangle\} \mid \{t,t'\} \in E,\ x,x' \in X \,\}$. By **nodes** and **edges** we will refer to this graph, while the nodes and edges of $\langle T, E\rangle$ will be called variables and variable pairs. The set of all nodes and edges is $I = (T \times X) \cup E_X$. All edges leading from a node $\langle t, x\rangle$ to all nodes of a neighboring object $t' \in N_t$ is a **pencil** $\langle t, t', x\rangle$. The set of all pencils is $P = \{\, \langle t,t',x\rangle \mid \{t,t'\} \in E,\ x \in X \,\}$. These concepts are illustrated in Figure 2.

Let A be a set of **weights**. A weight $g_{t,x} \in A$ is assigned to each node $\langle t, x\rangle$ and a weight $g_{tt',xx'} \in A$ to each edge $\{\langle t,x\rangle, \langle t',x'\rangle\}$, where we adopt $g_{tt',xx'} = g_{t't,x'x}$. By $\mathbf{g} \in A^I$ we will denote the vector with components $g_{t,x}$, $g_{tt',xx'}$.

Let A be closed under an associative and commutative operation $\times$, i.e., $\langle A, \times\rangle$ be a commutative semigroup. Let

$$F(\mathbf{x}\,|\,\mathbf{g}) = \Big(\prod_{t\in T} g_{t,x_t}\Big) \times \Big(\prod_{\{t,t'\}\in E} g_{tt',x_t x_{t'}}\Big) \tag{1}$$

Let $+$ be another associative and commutative operation, forming a commutative semigroup $\langle A, +\rangle$. Let

$$F(\mathbf{g}) = \sum_{\mathbf{x}\in X^T} F(\mathbf{x}\,|\,\mathbf{g}) \tag{2}$$

Here, symbols $\prod$ and $\sum$ refer to operations $\times$ and $+$. We will often denote $a \times b = ab$, $a + a = 2a$, $aa = a^2$.

Associativity and commutativity of operations $\times$ and $+$ is necessary for expressions (1) and (2) to be well-defined. Later we will also need that $\times$ distributes over $+$. An algebraic structure $\langle A, +, \times\rangle$ where operations $+$ and $\times$ are associative, commutative and distributive is known as a **commutative semiring**. Calculating expression (2) is called the (binary) **semiring-based CSP (SCSP)**. Its instance is given by $\langle T, E, X, A, +, \times, \mathbf{g}\rangle$ but when the rest is known from context we will refer to an instance only as 'problem $\mathbf{g}$'.

Some authors require that a semiring in addition has the zero element 0 and the unit element 1, satisfying $a + 0 = a$, $a \times 1 = a$, $a \times 0 = 0$ [16]. Our definition is also used [18] and suits us better since 0 and 1 will not always be needed.

3 Equivalence of SCSP Instances

The concept of equivalent problems is useful because it may allow to transform a given CSP to an equivalent one with a simpler solution (e.g. [32]). Usually two problems are understood equivalent iff they have equal solution sets. Following [31], we consider a stronger form of equivalence.

Definition 1. *Problems* $\mathbf{g}, \mathbf{g}' \in A^I$ *are* **equivalent** *(denoted by* $\mathbf{g} \sim \mathbf{g}'$*) if the functions* $F(\,\cdot\,|\,\mathbf{g})$ *and* $F(\,\cdot\,|\,\mathbf{g}')$ *are equal, i.e., if* $F(\mathbf{x}\,|\,\mathbf{g}) = F(\mathbf{x}\,|\,\mathbf{g}')$ *for all* $\mathbf{x} \in X^T$.

Note that equivalence is defined only with respect to the semigroup $\langle A, \times\rangle$ because the operation $+$ is absent in (1).

Two tasks related to equivalence naturally arise: (i) *test* whether two given problems are equivalent; (ii) *enumerate* elements of an equivalence class. These tasks may be easy or hard depending on $\langle A, \times\rangle$.

Example 1. Let $\langle A, \times\rangle = \langle\{0,1\}, \min\rangle$. Testing whether a vector $\mathbf{g}$ is equivalent to the all-zero vector $\mathbf{0}$ means testing the crisp CSP for satisfiability, hence it is NP-hard.

To reduce this complexity, we define a weaker concept.

Definition 2. *A* **local equivalent transformation** *of problem* $\mathbf{g}$ *on pencil* $\langle t, t', x\rangle$ *is a replacement of weights* $g_{t,x}$ *and* $\{\, g_{tt',xx'} \mid x' \in X \,\}$ *with some other weights* $g'_{t,x}$ *and* $\{\, g'_{tt',xx'} \mid x' \in X \,\}$ *such that*

$$g'_{t,x}\, g'_{tt',xx'} = g_{t,x}\, g_{tt',xx'} \quad \forall x' \in X. \tag{3}$$

By (1), a local equivalent transformation preserves the function $F(\,\cdot\,|\,\mathbf{g})$. Given equivalent problems $\mathbf{g}$ and $\mathbf{g}'$, three cases can arise: (i) $\mathbf{g}$ can be changed to $\mathbf{g}'$ by a finite sequence of local equivalent transformations; (ii) there is an infinite sequence of locally equivalent transformations of $\mathbf{g}$ that converges to $\mathbf{g}'$ (see Example 3 later on); (iii) $\mathbf{g}$ cannot be changed to $\mathbf{g}'$ by local equivalent transformations.

An important special case is when $\langle A, \times\rangle$ happens to be a group, i.e., we have division, a/b, and the unit element, 1. Then any local equivalent transformation is given by a weight $\varphi_{tt',x} \in A$, assigned to pencil $\langle t, t', x\rangle$, as

$$g'_{t,x} = g_{t,x}/\varphi_{tt',x}\,, \qquad g'_{tt',xx'} = g_{tt',xx'}\,\varphi_{tt',x} \tag{4}$$

Composing (4) for all pencils $\langle t, t', x\rangle \in P$ yields

$$g'_{t,x} = g_{t,x}\Big/\prod_{t'\in N_t} \varphi_{tt',x} \tag{5a}$$

$$g'_{tt',xx'} = g_{tt',xx'}\,\varphi_{tt',x}\,\varphi_{t't,x'} \tag{5b}$$

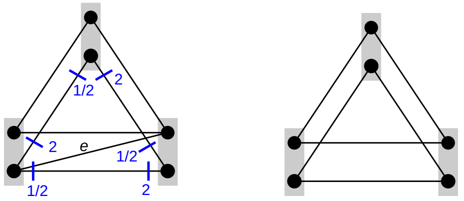

Figure 3: Cyclic repetition of the depicted local equivalent transformations decreases edge e to an arbitrarily small value.

Following [38, 19], we call (5) a **reparameterization**. Any equivalence class is fully covered by reparameterizations, as given by the following theorem [32].

Theorem 1. *Let $\langle A, \times\rangle$ be a group and let the graph $\langle T, E\rangle$ be connected. Then $\mathbf{g} \sim \mathbf{g}'$ iff there are $\varphi_{tt',x} \in A$ such that (5) holds.*

Proof. Can be found in [40, 41]. ∎

Another important case is the semigroup $\langle A' \cup \{0\}, \times\rangle$, where $\langle A', \times\rangle$ is a group and 0 is the identity element, satisfying $a \times 0 = 0$ for $a \in A'$. This is very close to $\langle A, \times\rangle$ being a group, yet Theorem 1 does not hold. Below are two examples of failure of Theorem 1.

Example 2. Let $A = \mathbb{R}_{0+}$ and $\times$ be the ordinary multiplication. This semigroup has subsemigroup $\langle\{0, 1\}, \times\rangle$, which is the same as $\langle\{0, 1\}, \min\rangle$ in Example 1. Hence, testing whether two problems are equivalent is NP-hard and Theorem 1 does not hold.

Example 3. Let $\langle A, \times\rangle$ be as in Example 2. Consider two problems in Figure 3, where the nodes have weights 1, the shown edges have weights 1, and the edges not shown have weights 0. Clearly, these problems are equivalent but are not related by (5) for any $\varphi_{tt',x} \in \mathbb{R}_{0+}$.

However, there is a sequence of locally equivalent transformations of the left problem that converges to the right problem. This sequence is the infinite cyclic repetition of transformations (4) with weights $\varphi_{tt',x}$ written next to pencils $\langle t, t', x\rangle$. Alternatively, such a sequence is obtained by running the arc consistency algorithm on the left problem, on semiring $\langle\mathbb{R}_{0+}, +, \times\rangle$ (given later on).

4 Arc Consistency

Arc consistency and relaxation labeling were originally formulated for the crisp CSP [39, 29, 17] and later underwent many generalization to soft CSPs [29, 5, 3, 4, 10, 9]. We observed that for a wide class of CSPs, arc consistency and the fixed point condition of max-sum diffusion can be seen as a single property. Based on this, we define arc consistency and an algorithm to achieve it in abstract semiring-based framework. We conjecture that the arc consistency algorithm makes sense and works for a certain class of semirings, which we call AC-semiring.

We will abbreviate **marginalization** (also known as **projection** in the CSP literature) over variable t, pencil $\langle t, t', x\rangle$, and variable pair $\{t, t'\}$ respectively by

$$g_{t,+} = \sum_{x \in X} g_{t,x}, \quad g_{tt',x+} = \sum_{x' \in X} g_{tt',xx'}, \quad g_{tt',++} = \sum_{x,x' \in X} g_{tt',xx'}$$

Definition 3. *A pencil $\langle t, t', x\rangle$ is* **arc consistent** *if*

$$g_{t,x} = g_{tt',x+} \tag{6}$$

A problem $\mathbf{g}$ is arc consistent if all its pencils are arc consistent.

Definition 4. **Local arc consistency transformation** *on pencil $\langle t, t', x\rangle$ is a local equivalent transformation that makes the pencil arc consistent.*

To do the local arc consistency transformation on pencil $\langle t, t', x\rangle$ given weights $g_{t,x}$ and $\{\, g_{tt',xx'} \mid x' \in X \,\}$, we need to find weights $g'_{t,x}$ and $\{\, g'_{tt',xx'} \mid x' \in X \,\}$ such that

$$g'_{t,x}\, g'_{tt',xx'} = g_{t,x}\, g_{tt',xx'} \qquad \forall x' \in X \tag{7a}$$

$$g'_{t,x} = g'_{tt',x+} \tag{7b}$$

Definition 5 introduces the class of semirings for which there exists a unique local arc consistency transformation, i.e., solution to (7). We omit the constant indices t, t', x in the definition, denoting $a = g_{t,x}$, $i = x'$, $b_i = g_{tt',xx'}$.

Definition 5. *Semiring $\langle A, +, \times\rangle$ is an* **AC-semiring** *if for any $a, b_i \in A$ there exist unique $a', b'_i \in A$ satisfying equations $a'b'_i = ab_i$ (for $i = 1, \ldots, n$) and $a' = \sum_{i=1}^n b'_i$.*

We do not know yet how to characterize AC-semirings by more elementary semiring properties. Nevertheless, one property of AC-semirings is apparent easily. Summing the first equation over i and substituting to the second one gives $(a')^2 = a\sum_{i=1}^n b_i$. As a and b_i are arbitrary, for any $c \in A$ the equation $x^2 = c$ must have a unique solution, $x = c^{1/2}$. Thus, AC-semirings must allow for square roots.

Definition 6. *For an SCSP on an AC-semiring, the* **arc consistency algorithm** *is as follows:*

1. *Choose an arc inconsistent pencil. This choice can be arbitrary, provided that every pencil has a non-zero probability to be chosen. If all pencils are arc consistent, stop.*
2. *Do local arc consistency transformation on the pencil. Go to 1.*

Conjecture 1. *For an SCSP on an AC-semiring, the arc consistency algorithm converges to an arc consistent problem.*

4.1 Upper Bound

Let us generalize Schlesinger's upper bound on the max-sum CSP [31, 40] to SCSPs,

$$\bar F(\mathbf{g}) = \Big(\prod_{t \in T} g_{t,+}\Big) \times \Big(\prod_{\{t,t'\} \in E} g_{tt',++}\Big) \tag{8}$$

This quantity is useful for two reasons, given by Theorems 2 and 3.

Definition 7. *We define the relation $\preceq$ on $\langle A, +\rangle$ as follows: $a \preceq b$ iff either $a = b$ or there is $c \in A$ such that $a + c = b$.*

The relation $\preceq$ is reflexive and transitive, i.e., a preorder [18]. It is often called the **natural preorder** on $\langle A, +\rangle$. For many semirings it becomes a partial or even a total order.

Theorem 2. *For any semiring $\langle A, +, \times\rangle$ and any $\mathbf{g} \in A^I$, we have $F(\mathbf{g}) \preceq \bar{F}(\mathbf{g})$.*

Proof. Multiplying out the factors in (8) shows that $\bar{F}(\mathbf{g})$ contains the terms present in (2) plus some additional terms. By Definition 7, we have $F(\mathbf{g}) \preceq \bar{F}(\mathbf{g})$. ∎

In other words, $\bar{F}(\mathbf{g})$ is an upper bound on $F(\mathbf{g})$. This has great practical importance because it gives an approximation (often even the exact solution) of the SCSP to solve.

We observed that all AC-semirings we discovered satisfied

$$2ab \preceq a^2 + b^2 \tag{9}$$

for any $a, b \in A$. It can be easily verified that (9) implies that any vectors $\mathbf{a}, \mathbf{b} \in A^n$ satisfy

$$\Big(\sum_i a_i\, b_i\Big)^2 \preceq \Big(\sum_i a_i^2\Big)\Big(\sum_i b_i^2\Big) \tag{10}$$

This relation is a semiring generalization of well-known Jensen's inequality, characterizing convexity. For the special choice of the semiring in §6.1, (10) becomes ordinary Jensen's inequality [6] applied to the log-sum-exp function. Thus, (10) characterizes 'semiring convexity' of semiring addition.

Theorem 3. *Let $\langle A, +, \times\rangle$ be an AC-semiring and let $\{t, t'\} \in E$. Doing $|X|$ local arc consistency transformations successively on pencils $\{\, \langle t, t', x\rangle \mid x \in X \,\}$ does not increase $\bar{F}(\mathbf{g})$.*

Proof. Before the transformations, the upper bound is

$$\bar{F}(\mathbf{g}) = c\, g_{t,+}\, g_{tt',++} \tag{11}$$

where c are the factors that do not depend on the weights $\{\, g_{t,x} \mid x \in X \,\}$ and $\{\, g_{tt',xx'} \mid x, x' \in X \,\}$.

Marginalizing (7a) over x' and substituting from (7b) yields $g'_{t,x} = g'_{tt',x+} = (g_{t,x}\, g_{tt',x+})^{1/2}$. Hence, after the transformations the upper bound is

$$\bar{F}(\mathbf{g}) = c\, g'_{t,+}\, g'_{tt',++} = c\,\Big[\sum_{x\in X} (g_{t,x}\, g_{tt',x+})^{1/2}\Big]^2 \tag{12}$$

Setting $i = x$, $a_i^2 = g_{t,x}$, and $b_i^2 = g_{tt',x+}$ in (10), we obtain that (12) is not greater than (11). ∎

Theorem 3 shows that the arc consistency algorithm can be partially understood as decreasing $\bar{F}(\mathbf{g})$ by local equivalent transformations. As we minimize over a subset of variables in each iteration, it can be interpreted as a co-ordinate descent [6]. However, this interpretation of the arc consistency algorithm has difficulties. First, not every transformation need to decrease $\bar{F}(\mathbf{g})$. Second, the algorithm need not find the global minimum of $\bar{F}(\mathbf{g})$. Thus Theorem 3 is only an evidence but not a proof that Conjecture 1 is true.

Important remark. In the previous sections, symbols $+$, $\times$, $\sum$, $\prod$, 0, 1 denoted abstract semiring operations. In the sequel, they will have their ordinary meaning as addition, multiplication, *etc*. Furthermore, we will denote $F(\mathbf{x} \,|\, \mathbf{g})$, $F(\mathbf{g})$, $\bar{F}(\mathbf{g})$ defined for particular concrete operations $+$ and $\times$ by $F_\times(\mathbf{x} \,|\, \mathbf{g})$, $F_{+,\times}(\mathbf{g})$, $\bar{F}_{+,\times}(\mathbf{g})$, respectively.

5 Examples

We will give examples of AC-semirings, covering most of the existing SCSPs and also some new ones. For all of them, we verified Conjecture 1 and Theorem 3 experimentally.

5.1 Distributive Lattice

Let $\preceq$ be a partial order on A. Let $\langle A, \vee, \wedge\rangle$ be a distributive lattice defined by $\preceq$, where $\vee$ is the supremum (least upper bound), and $\wedge$ the infimum (greatest lower bound). By its definition, a distributive lattice is a commutative semiring. It is even an AC-semiring since it is easy to verify that

$$g'_{tt',xx'} = g_{t,x} \wedge g_{tt',xx'}\,, \quad g'_{t,x} = g_{t,x} \wedge g_{tt',x\vee} \tag{13}$$

is the unique solution to (7). Expression (9), which reads $a \wedge b \preceq a \vee b$, clearly also holds true.

Recall, a binary operation $\circ$ is **idempotent** if $a \circ a = a$ for $a \in A$. Here, both operations $\vee$ and $\wedge$ are idempotent.

Distributive lattices cover several important CSPs.

Setting $A = \{0, 1\}$ yields the crisp CSP [26]. The algorithm from Definition 6 becomes the well-known CSP arc consistency algorithm (discrete relaxation labeling) [29].

Setting $A = [0, 1]$ yields the fuzzy CSP. Setting $A = \mathbb{R} \cup \{-\infty, +\infty\}$ yields the bottleneck (minimax) algebra [11] and the corresponding CSP.

In these examples, $\preceq$ was a total order. An important case when $\preceq$ is only a partial order is when A is a set of subsets of some set, $\vee$ is the set union and $\wedge$ is the set intersection.

It is known [5, 4] that if the semiring multiplication is idempotent, local consistency algorithms converge in finite time. However, exact algorithms differ from paper to paper [5, 4, 10] and we have yet to unify these results with ours.

5.2 Max-sum Semiring

Semiring $\langle \mathbb{R}, \max, +\rangle$ is an AC-semiring. Often one sets $A = \mathbb{R} \cup \{-\infty\}$, which again yields an AC-semiring, as one can verify. This semiring gives rise to the max-plus algebra and tropical mathematics, surveyed e.g. in [14, 25]. The corresponding CSP is the max-sum CSP, also known as the weighted CSP, MAP inference in MRFs [28, 24, 38, 36], or finding the mode of the Boltzmann/Gibbs distribution [15].

Choosing A the rational rather than real numbers again yields an AC-semiring. Choosing A the integers does not yield an AC-semiring since we lose the square root.

For semiring $\langle \mathbb{R} \cup \{-\infty\}, \max, +\rangle$, the arc consistency algorithm was proposed to decrease Schlesinger's upper bound $\bar{F}_{\max,+}$ by Koval and Kovalevsky [23]. Strictly speaking, the algorithm belonged to the 'folklore' knowledge in Kiev pattern recognition group in 1970's as 'max-sum diffusion' and the exact authorship was partially forgotten [32]. We will return to it in detail in §7.

In practice, minimizing $\bar{F}_{\max,+}$ yields a very useful upper bound on $F_{\max,+}$ (quite often, this bound is tight). Minimizing $\bar{F}_{\max,+}$ is dual to a linear programming relaxation of the max-sum CSP, considered by many others [22, 8, 36].

Lexicographic max-sum semiring. Another AC-semiring is $\langle \mathbb{R}^n, \max_{\text{lex}}, +\rangle$, where $\max_{\text{lex}}$ is maximum with respect to lexicographic ordering on $\mathbb{R}^n$ induced by the total order $\preceq$ on $\mathbb{R}$, and $+$ is the component-wise addition on $\mathbb{R}^n$.

5.3 Sum-product Semiring

The CSP on semiring $\langle \mathbb{R}_{++}, +, \times \rangle$ is equivalent to computing the partition function of MRFs [28, 24, 38, 7] or Boltzmann/Gibbs distribution [15]. Often, setting $A = \mathbb{R}_{0+}$ instead of $A = \mathbb{R}_{++}$ is useful and yields an AC-semiring. However, $A = \mathbb{R}$ does not yield an AC-semiring.

In §6, we prove Conjecture 1 for semiring $\langle \mathbb{R}_{++}, +, \times \rangle$. Arc consistency for this semiring has not been described. The closest work is the [38, 37], giving upper bounds on the partition function based on convex combination of trees.

The upper bound $\bar{F}_{+,\times}$ is not a very good approximation of $F_{+,\times}$, better upper bounds can be found in [38, 37]. Its main significance is in its relation to arc consistency.

6 Sum-product Arc Consistency

In this section, we will introduce sum-product arc consistency. The main result is the following theorem, which we will prove in the rest of the section.

Theorem 4. *On the semiring $\langle \mathbb{R}_{++}, +, \times \rangle$, there is a unique arc consistent problem in every equivalence class. This problem can be found by the arc consistency algorithm.*

6.1 Log-sum-exp Function

Rather than with $\langle \mathbb{R}_{++}, +, \times \rangle$, in §6 we will work with semiring $\langle \mathbb{R}, \oplus, + \rangle$ where

$$a \oplus b = \log(\mathrm{e}^a + \mathrm{e}^b) \tag{14}$$

denotes the log-sum-exp function [6]. The two semirings are isomorphic via the mapping $a \mapsto \log a$.

We denote $s(\mathbf{a}) = \bigoplus_{i=1}^n a_i$, where a_i are the components of vector $\mathbf{a} \in \mathbb{R}^n$. It is well-known [6] that s is convex, i.e, it satisfies Jensen's inequality (10). It is strictly convex in every direction except direction $\langle 1, \dots, 1 \rangle$, because $s(\mathbf{a} + b) = s(\mathbf{a}) + b$ for all $b \in \mathbb{R}$.

The derivative of the log-sum-exp function is

$$\frac{\partial(a \oplus b)}{\partial a} = \exp[a - (a \oplus b)] = \frac{\mathrm{e}^a}{\mathrm{e}^a + \mathrm{e}^b} \tag{15a}$$

$$\frac{\mathrm{d}s(\mathbf{a})}{\mathrm{d}\mathbf{a}} = \exp[\mathbf{a} - s(\mathbf{a})] = \frac{\exp \mathbf{a}}{\sum_{i=1}^n \exp a_i} \tag{15b}$$

Since vector (15b) is non-negative and sums up to one it has the properties of a probability distribution. It is sometimes called soft maximum because it says "to what extent each element of $\mathbf{a}$ is maximal".

6.2 Minimizing the Upper Bound

On $\langle \mathbb{R}, \oplus \rangle$, the pre-order $\preceq$ is the ordinary total order on $\mathbb{R}$.

Since $\langle \mathbb{R}, + \rangle$ is a group, by Theorem 1 every equivalence class can be covered using (5). Reparameterization (5) reads

$$g'_{t,x} = g_{t,x} - \sum_{t' \in N_t} \varphi_{tt',x}, \tag{16a}$$

$$g'_{tt',xx'} = g_{tt',xx'} + \varphi_{tt',x} + \varphi_{t't,x'}. \tag{16b}$$

We will rewrite (16) into a matrix form as $\mathbf{g}' = \mathbf{g} + \boldsymbol{\varphi}\mathbf{A}$, where $\boldsymbol{\varphi} \in \mathbb{R}^P$ is the vector with the components $\varphi_{tt',x} \in \mathbb{R}$ and $\mathbf{A}$ is the appropriate matrix with entries in $\{-1, 0, +1\}$. Vectors $\mathbf{g}$ and $\boldsymbol{\varphi}$ are *row* vectors.

Minimizing the sum-product upper bound reads

$$\min\{\, \bar{F}_{\oplus,+}(\mathbf{g} + \boldsymbol{\varphi}\mathbf{A}) \mid \boldsymbol{\varphi} \in \mathbb{R}^P \,\} \tag{17}$$

Since the log-sum-exp function is convex, $\bar{F}_{\oplus,+}$ is convex, too. Thus, (17) is a convex minimization task; in fact, it is a special case of geometric programming [6]. Since $\bar{F}_{\oplus,+}$ is smooth, the condition necessary and sufficient for global minimum can be found by calculus as follows.

Theorem 5. *In a class of equivalent problems $\mathbf{g}$, the upper bound $\bar{F}_{\oplus,+}(\mathbf{g})$ is minimal iff $\mathbf{g}$ satisfies*

$$g_{t,x} - g_{t,\oplus} = g_{tt',x\oplus} - g_{tt',\oplus\oplus} \tag{18}$$

for all pencils $\langle t, t', x \rangle \in P$.

Proof. By (15), the derivative $\mathrm{d}\bar{F}_{\oplus,+}(\mathbf{g})/\mathrm{d}\mathbf{g}$ reads

$$\frac{\partial \bar{F}_{\oplus,+}(\mathbf{g})}{\partial g_{t,x}} = \exp(g_{t,x} - g_{t,\oplus}) = \frac{\exp g_{t,x}}{\sum_x \exp g_{t,x}} \tag{19a}$$

$$\frac{\partial \bar{F}_{\oplus,+}(\mathbf{g})}{\partial g_{tt',xx'}} = \exp(g_{tt',xx'} - g_{tt',\oplus\oplus}) = \frac{\exp g_{tt',xx'}}{\sum_{x,x'} \exp g_{tt',xx'}} \tag{19b}$$

Denoting $\mathbf{g}' = \mathbf{g} + \boldsymbol{\varphi}\mathbf{A}$, one verifies by the chain rule that

$$\frac{\partial \bar{F}_{\oplus,+}(\mathbf{g}')}{\partial \varphi_{tt',x}} = \frac{\partial \bar{F}_{\oplus,+}(\mathbf{g}')}{\partial g'_{t,x}} - \sum_{x' \in X} \frac{\partial \bar{F}_{\oplus,+}(\mathbf{g}')}{\partial g'_{tt',xx'}} \tag{20}$$

The minimum of (17) is attained at $\boldsymbol{\varphi}$ for which (20) equals zero. This reveals the remarkable fact that in the optimum, the derivatives of the upper bound are arc consistent. Substituting for $\mathbf{g}'$ in (20) from (19) yields (18). ∎

The condition (18) is invariant to transformations

$$g'_{t,x} = g_{t,x} + \psi_t, \quad g'_{tt',xx'} = g_{tt',xx'} + \psi_{tt'} \tag{21}$$

for any $\psi_t, \psi_{tt'} \in \mathbb{R}$, i.e., to adding constants to variables and variable pairs. If $\sum_t \psi_t + \sum_{tt'} \psi_{tt'} = 0$, the transformation (21) preserves also $\bar{F}_{\oplus,+}(\mathbf{g})$. Ambiguity in choosing ψ_t and $\psi_{tt'}$ can be fixed by imposing the constraint $g_{t,\oplus} = g_{tt',\oplus\oplus}$ on all $\{t, t'\} \in E$. It follows that (18) can be written simply as the consistency condition (6).

It remains to prove that any class of equivalent problems contains a unique arc consistent problem. To do it, we will show that adding constants to variables and variable pairs is the *only* equivalent transformation that preserves (18).

6.3 Lagrangian Dual of Upper Bound Minimization

Let there be numbers $\mu_{t,x}, \mu_{tt',xx'} \in \mathbb{R}_{0+}$, assigned to nodes $\langle t, x \rangle$ and edges $\{\langle t, x \rangle, \langle t', x' \rangle\}$. By $\boldsymbol{\mu} \in \mathbb{R}^I_{0+}$ we denote the *column* vector whose components are $\mu_{t,x}, \mu_{tt',xx'}$.

Theorem 6. *Let $\boldsymbol{\mu} \in \mathbb{R}^I_{0+}$ with $\mu_{t,+} = 1$, $\mu_{tt',++} = 1$. Let $\mathbf{g} \in \mathbb{R}^I$. Then*

$$\bar{F}_{\oplus,+}(\mathbf{g}) + (\log \boldsymbol{\mu}^\top)\boldsymbol{\mu} - \mathbf{g}\boldsymbol{\mu} \geq 0 \tag{22}$$

(Here, $\boldsymbol{\mu}^\top$ is the matrix transpose and the logarithm is meant component-wise.) Equality in (22) occurs iff

$$\boldsymbol{\mu} = \frac{\mathrm{d}\bar{F}_{\oplus,+}(\mathbf{g})}{\mathrm{d}\mathbf{g}} \tag{23}$$

which in turn occurs iff

$$\mu_{t,x} = \mu_{tt',x+} \tag{24}$$

Proof. One easily verifies that (22) is invariant to equivalent transformations (16). Thus, without loss of generality we set $g_{t,\oplus} = 0$ and $g_{tt',\oplus\oplus} = 0$. Hence $\bar{F}_{\oplus,+}(\mathbf{g}) = 0$. With $\bar{F}_{\oplus,+}(\mathbf{g}) = 0$, (22) is the Kullback-Leibler divergence between $\boldsymbol{\mu}$ and $\exp \mathbf{g}$, which is non-negative.

The rest is verified by direct substitution and from the fact that (22) is convex in $\langle \mathbf{g}, \boldsymbol{\mu} \rangle$. ■

Theorem 7. *The concave maximization task*

$$\max\{\, (\mathbf{g} - \log \boldsymbol{\mu}^\top)\boldsymbol{\mu} \mid \boldsymbol{\mu} \geq \mathbf{0},\ \mu_{t,x} = \mu_{tt',x+},\ \mu_{t,+} = 1 \,\} \tag{25}$$

and the convex minimization task (17) are related by strong Lagrangian duality. Their optima equal iff (23) holds.

Proof. Follows from Theorem 6. Note, $\mu_{tt',++} = 1$ missing in (25) is implied by $\mu_{t,+} = 1$ and $\mu_{t,x} = \mu_{tt',x+}$. ■

Since the task (25) is strictly concave, it attains its optimum at a single point, $\boldsymbol{\mu}^*$. By (23), expression (19) must equal $\boldsymbol{\mu}^*$ in the optimum. Clearly, adding ψ_t to $g_{t,x}$ and adding $\psi_{tt'}$ to $g_{tt',xx'}$ is the only transformation preserving (19). This concludes the proof of Theorem 4.

Although Conjecture 1 has been proved for semirings $\langle \mathbb{R}_{++}, +, \times \rangle$ and $\langle \mathbb{R}, \oplus, + \rangle$, it holds also for $\langle \mathbb{R}_{0+}, +, \times \rangle$ and $\langle \mathbb{R} \cup \{-\infty\}, \oplus, + \rangle$. The difference is that the arc consistency algorithm may converge to 0 resp. $-\infty$, which happens when (17) is unbounded and (25) infeasible. In the product $\mathbf{g}\boldsymbol{\mu}$ in (25), we adopt convention $-\infty \times 0 = 0$.

7 Max-sum Arc Consistency

In this section, we will derive the CSP and arc consistency on $\langle \mathbb{R}, \max, + \rangle$ by a limit of those on $\langle \mathbb{R}, \oplus, + \rangle$.

7.1 From Soft to Hard Maximum

Ordinary (hard) maximum can be obtained as a limit of a sequence of smooth, non-idempotent functions. E.g., transition from semigroup $\langle \mathbb{R}_{0+}, + \rangle$ to $\langle \mathbb{R}_{0+}, \max \rangle$ corresponds to the limit of vector β-norms for $\beta \to \infty$. Transition from $\langle \mathbb{R} \cup \{-\infty\}, \oplus \rangle$ to $\langle \mathbb{R} \cup \{-\infty\}, \max \rangle$ is

$$\lim_{\beta \to \infty} \frac{(\beta a) \oplus (\beta b)}{\beta} = \max\{a, b\} \tag{26}$$

This is known as Maslov's dequantization or 'tropicalization', see e.g. [25, 14].

The transition for the 'soft maximum' (15) is

$$\boldsymbol{\mu} = \lim_{\beta \to \infty} \frac{\exp \beta \mathbf{a}}{\sum_i \exp \beta a_i} \tag{27}$$

Denoting $B = \{\, i \mid a_i = \max_i a_i \,\}$, one verifies that $\mu_i = 0$ if $i \notin B$ and $\mu_i = |B|^{-1}$ if $i \in B$. Note that even in the limit, (27) sums up to one. In statistical physics and often in machine learning [38], transition (27) is called the zero temperature limit (of a probability distribution).

7.2 From Sum-product to Max-sum CSP

Similarly as in (26), one can define the dequantization of a function $f(\mathbf{g})$ as

$$\lim_{\beta \to \infty} \frac{f(\beta \mathbf{g})}{\beta} \tag{28}$$

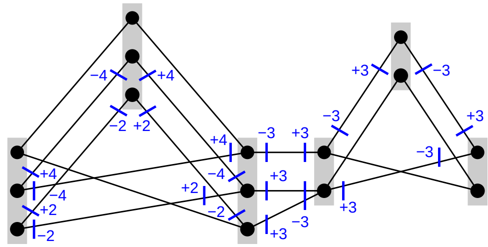

Figure 4: An arc consistent SCSP on semiring $\langle \mathbb{R}, \max, + \rangle$, which has non-minimal upper bound $\bar{F}_{\max,+}(\mathbf{g})$.

Dequantizing $F_\oplus(\mathbf{x} \,|\, \mathbf{g})$ yields $F_{\max}(\mathbf{x} \,|\, \mathbf{g})$ and dequantizing $F_{\oplus,+}(\mathbf{g})$ yields $F_{\max,+}(\mathbf{g})$.

Dequantizing $\bar{F}_{\oplus,+}(\mathbf{g})$ yields Schlesinger's upper bound $\bar{F}_{\max,+}(\mathbf{g})$ on the max-sum CSP. Its minimization,

$$\min\{\, \bar{F}_{\max,+}(\mathbf{g} + \boldsymbol{\varphi}\mathbf{A}) \mid \boldsymbol{\varphi} \in \mathbb{R}^P \,\} \tag{29}$$

is an unconstrained convex nonsmooth optimization task, which can be formulated as a linear program [31, 40, 41].

Dequantizing expression $(\mathbf{g} - \log \boldsymbol{\mu}^\top)\boldsymbol{\mu}$ yields $\mathbf{g}\boldsymbol{\mu}$ and hence dequantizing (25) yields the linear program

$$\max\{\, \mathbf{g}\boldsymbol{\mu} \mid \boldsymbol{\mu} \geq \mathbf{0},\ \mu_{t,x} = \mu_{tt',x+},\ \mu_{t,+} = 1 \,\} \tag{30}$$

This is the linear programming relaxation of the max-sum CSP by Schlesinger [31] and independently by others [22, 8, 36]. The programs (29) and (30) are mutually dual.

7.3 Arc Consistency Is Insufficient for Optimality

Dequantizing the stationary condition (18) yields

$$g_{t,x} - g_{t,\max} = g_{tt',x\max} - g_{tt',\max\max} \tag{31}$$

Imposing the constraint $g_{t,\max} = g_{tt',\max\max}$, analogical to $g_{t,\oplus} = g_{tt',\oplus\oplus}$, gives the arc consistency condition (6).

One could think that similarly as in the sum-product case, (31) is necessary and sufficient for minimality of $\bar{F}_{\max,+}(\mathbf{g})$. Surprisingly, this is false: (31) is neither sufficient nor necessary and (6) is not sufficient for this.

Example 4. (by Schlesinger [30]) In the max-sum CSP in Figure 4, the nodes have weights 0, the shown edges have weights 0, and the unshown edges have strictly negative weights. The problem is arc consistent but $\bar{F}_{\max,+}$ can be decreased by an equivalent transformation. The decreasing direction $\boldsymbol{\varphi}$ is depicted by numbers $\varphi_{tt',x}$ written near corresponding pencils $\langle t, t', x \rangle$.

The phenomenon is caused by the fact that co-ordinate descent need not find the minimum of a nonstrictly convex function [1]. The function $\bar{F}_{\max,+}$ is piecewise linear, hence nonsmooth and nonstrictly convex. Figure 5 shows how the contours of $\bar{F}_{\max,+}(\mathbf{g} + \boldsymbol{\varphi}\mathbf{A})$ as a function of $\boldsymbol{\varphi}$ can look like, simplified to two dimensions.

Despite this, it has been justified by many experiments that Conjecture 1 holds for the max-sum CSP. Yet there may be many arc consistent problems in an equivalence class and the point of convergence of the arc consistency algorithm may depend on the order of local equivalent transformations.

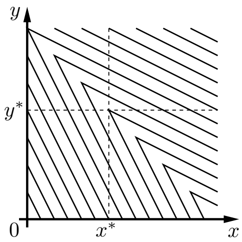

Figure 5: Co-ordinate descent did not find the minimum of a non-strictly convex function $f(x, y)$. Point $\langle x^*, y^* \rangle$ is not optimal.

7.4 Optimal Arc Consistency

It is of great practical interest to solve (29) and (30). General linear programming algorithms, such as simplex or interior point, are not usable for large instances (occurring e.g. in low-level computer vision) both for space and time inefficiency (e.g. [42]). Instead, one needs an efficient *network algorithm*. The max-sum arc consistency algorithm is a good candidate. However, we are facing the problem that max-sum arc consistency is insufficient for optimality. Given an arc consistent problem $\mathbf{g}$, we cannot even recognize whether it minimizes $\bar{F}_{\max,+}(\mathbf{g})$ on its equivalent class or not. We will show how to tackle these issues.

Let $\mathbf{g}' = \mathbf{C}_{\oplus,+}(\mathbf{g})$ denote the (unique) point of convergence of the arc consistency algorithm on $\langle \mathbb{R}, \oplus, + \rangle$. Let $\boldsymbol{\mu} = \mathbf{m}_{\oplus,+}(\mathbf{g})$ denote the (unique) solution of (25). Recall, they are related by (23). For increasing β, the sequences

$$\{\, \mathbf{C}_{\oplus,+}(\beta \mathbf{g})/\beta \,\}, \quad \{\, \mathbf{m}_{\oplus,+}(\beta \mathbf{g}) \,\} \tag{32}$$

converge to a solution of (29) and (30), respectively. The problem is that convergence of the arc consistency algorithm on $\langle \mathbb{R}, \oplus, + \rangle$ will be increasingly slower for larger β. This can be reduced by re-using solutions from previous iterations, as done by the following algorithm.

1: **loop**
2: $\quad \mathbf{g} := 2\,\mathbf{C}_{\oplus,+}(\mathbf{g});$
3: $\quad g_{t,x} := g_{t,x} - g_{t,\oplus};\;\; g_{tt',xx'} := g_{tt',xx'} - g_{tt',\oplus\oplus};$
4: **end loop**

The factor 2 on line 2 ensures increasing β to 2β, initially having $\beta = 1$. The algorithm converges to a weight vector $\mathbf{g}^*$ such that $\boldsymbol{\mu}^* = \exp \mathbf{g}^*$ is a solution to (30). It can be understood as a special interior point algorithm [6] to solve the linear program (30) because $\boldsymbol{\mu} = \exp \mathbf{g}$ is approaching the boundary of the feasible region of (30) with increasing β. The term $(\log \boldsymbol{\mu}^\top)\boldsymbol{\mu}$ is the barrier function. However, while a typical barrier function approaches ∞ near the boundary of the feasible region, $(\log \boldsymbol{\mu}^\top)\boldsymbol{\mu}$ is bounded for feasible $\boldsymbol{\mu}$ and only the magnitude of its derivative approaches ∞.

The algorithm is not fully practical due to its slow convergence and because it is not clear when to stop the arc consistency algorithm on line 2. Nevertheless, in our experiments it often achieved a better upper bound than the max-sum arc consistency in a reasonable time. To avoid using $\log$ and $\exp$ functions, the algorithm can be translated from $\langle \mathbb{R}, \oplus, + \rangle$ to $\langle \mathbb{R}_{++}, +, \times \rangle$.

7.5 Test for Optimality of Arc Consistency

Finally, we will show how to test whether a given $\mathbf{g}$ is optimal, i.e., whether $\bar{F}_{\max,+}(\mathbf{g})$ is minimal for all problems equivalent with $\mathbf{g}$. The side-result of this test will be:

- if $\mathbf{g}$ is optimal: a solution $\boldsymbol{\mu}^*$ to (30);
- if $\mathbf{g}$ is not optimal: a decreasing direction $\boldsymbol{\varphi}$.

First, set every node with $g_{t,x} < g_{t,\max}$ and every edge with $g_{tt',xx'} < g_{tt',\max\max}$ to $-\infty$. Set the other ones to 0. Clearly, this simplification does not change the situation.

Theorem 8. *Let* $\mathbf{g} \in \{-\infty, 0\}^I$.

- *If the problem (25) is feasible (i.e., (17) is bounded) then its solution* $\boldsymbol{\mu}^* = \mathbf{m}_{\oplus,+}(\mathbf{g})$ *is also a solution to (30).*
- *If there exists* $\boldsymbol{\varphi}$ *such that* $\bar{F}_{\oplus,+}(\mathbf{g} + \boldsymbol{\varphi}\mathbf{A}) < 0$ *then* $\bar{F}_{\max,+}(\mathbf{g} + \lambda\boldsymbol{\varphi}\mathbf{A}) < 0$ *for any* $0 < \lambda \le 1$.

Proof. Let (25) be feasible. Then its solution is $\boldsymbol{\mu}^*$. Since $\mathbf{g} \in \{-\infty, 0\}$ and since $(\log \boldsymbol{\mu}^\top)\boldsymbol{\mu} < 0$ for any feasible $\boldsymbol{\mu}$, we have $\mathbf{g}\boldsymbol{\mu}^* = 0$. Since $\mathbf{g} \in \{-\infty, 0\}$, the optimum of (30) can be either $-\infty$ or 0. Hence, $\boldsymbol{\mu}^*$ is a solution to (30).

Any $a, b \in \mathbb{R} \cup \{-\infty\}$ satisfy $\max\{a, b\} \le a \oplus b$. Hence, $\bar{F}_{\max,+}(\mathbf{g}) \le \bar{F}_{\oplus,+}(\mathbf{g})$ for any $\mathbf{g}$. It follows that $\bar{F}_{\max,+}(\mathbf{g} + \boldsymbol{\varphi}\mathbf{A}) \le \bar{F}_{\oplus,+}(\mathbf{g} + \boldsymbol{\varphi}\mathbf{A}) < 0$. Since $\bar{F}_{\max,+}$ is convex, we have $\bar{F}_{\max,+}(\mathbf{g} + \lambda\boldsymbol{\varphi}\mathbf{A}) < 0$ for $0 < \lambda \le 1$. ■

Thus, the test is as follows. Given $\mathbf{g} \in \{-\infty, 0\}^I$, we run the $\langle \oplus, + \rangle$ arc consistency algorithm. If $\bar{F}_{\oplus,+}$ becomes negative during the algorithm, we stop because it is already sure that $\mathbf{g}$ does not minimize $\bar{F}_{\max,+}$. In that case, we obtain a decreasing direction. If the algorithm converges, we obtain a solution to (30).

Application to the crisp CSP. The test can be used to disprove satisfiability of a crisp CSP. The crisp CSP is isomorphic to the SCPS on $\langle \{-\infty, 0\}, \max, + \rangle$. Therefore, if $\mathbf{g} \in \{-\infty, 0\}$ does not minimize $\bar{F}_{\max,+}$ on its equivalence class then the corresponding crisp CSP is unsatisfiable. To our knowledge, this test is new, and qualitatively different from usual tests based on local consistencies [12].

8 Conclusion

The contributions of the paper are threefold.

First, in §2–5 we have unified the concept of arc consistency, well-known for SCSP with idempotent semiring multiplication, with similar algorithms for SCSPs with non-idempotent semiring multiplications. In particular, this unifies Rosenfeld's discrete relaxation labeling with max-sum diffusion. Here, the main result is Conjecture 1. In the future, Conjecture 1 should either be proved or a counter-example found. Nevertheless, even if a counter-example is found, the result that arc consistency can be unified for many practically important CSPs would stay valid.

Second, in §6 we have newly introduced sum-product arc consistency. The main results are Theorems 4 and 5.

Third, in §7 we have derived the max-sum case as a limit of the sum-product case. We have shown where uniqueness and optimality of sum-product arc consistency are lost, and how the sum-product arc consistency can be used to compute and test for optimal max-sum arc consistency. The main result is Theorem 8 and the algorithm in §7.4. In the future, the algorithm should be made practical for large instances, occurring e.g. in low level vision.

References

[1] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1999.

[2] S. Bistarelli. *Semirings for Soft Constraint Solving and Programming*. Lecture Notes in Computer Science. Springer Verlag, 2004.

[3] S. Bistarelli, P. Codognet, Y. Georget, and F. Rossi. Labeling and partial arc consistency for soft constraints. In *Joint Conf. Declarative Programming (AGP), Italy*, pages 121–136, 1999.

[4] S. Bistarelli and F. Gadducci. Enhancing constraints manipulation in semiring-based formalisms. In *Eur. Conf. Artificial Intelligence (ECAI)*, pages 63–67. IOS Press, 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*, 4(3):199–240, 1999.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[7] A. Bulatov, P. Jeavons, and A. Krokhin. Classifying the complexity of constraints using finite algebras. *Computing*, 34(3):720–742, 2005.

[8] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *Symposium on Discrete Algorithms*, pages 109–118, 2001.

[9] M. Cooper, S. de Givry, and T. Schiex. Optimal soft arc consistency. In *Proc. IJCAI, Hyderabad, India*, 2007.

[10] M. Cooper and T. Schiex. Arc consistency for soft constraints. *Artificial Intelligence*, 154(1-2):199–227, 2004.

[11] R. A. Cuninghame-Green. *Minimax Algebra*. Springer-Verlag New York, 1979.

[12] R. Debruyne and C. Bessière. Domain filtering consistencies. *Journal of Artificial Intelligence Research*, (14):205–230, May 2001.

[13] B. Flach. A diffusion algorithm for decreasing energy of max-sum labeling problem. Fakultät Informatik, Technische Universität Dresden, Germany. Unpublished, 1998.

[14] S. Gaubert. Methods and applications of (max,+) linear algebra. Technical Report 3088, Institut national de recherche en informatique et en automatique (INRIA), 1997.

[15] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans. Pattern Analysis Machine Intelligence*, 6:721–741, 1984.

[16] J. S. Golan. *Semirings and their Applications*. Kluwer Academic, 1999.

[17] P. V. Hentenryck, Y. Deville, and C.-M. Teng. A generic arc-consistency algorithm and its specializations. *Artificial Intelligence*, 57(2–3):291–321, 1992.

[18] J. Kohlas and N. Wilson. Exact and approximate local computation in semiring induced valuation algebras. Technical Report 06-06, Department of Informatics, University of Fribourg, 2006.

[19] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

[20] V. Kolmogorov and C. Rother. Comparison of energy minimization algorithms for highly connected graphs. In *European Conf. Computer Vision (ECCV)*, pages II: 1–15, 2006.

[21] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? In *European Conf. Computer Vision (ECCV)*, pages 65–81. Springer-Verlag, 2002.

[22] A. Koster, C. P. M. van Hoesel, and A. W. J. Kolen. The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters*, 23(3–5):89–97, 1998.

[23] V. A. Kovalevsky and V. K. Koval. A diffusion algorithm for decreasing energy of max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR. Unpublished, approx. 1975.

[24] S. Lauritzen. *Graphical Models*. Oxford, 1996.

[25] G. L. Litvinov. The Maslov dequantization, idempotent and tropical mathematics: A brief introduction. *Mathematical Sciences*, 2006. Available from http://arXiv.org.

[26] A. Mackworth. Constraint satisfaction. In *Encyclopedia of Artificial Intelligence*, pages 285–292. Wiley, New York, 1991.

[27] P. Meseguer, F. Rossi, and T. Schiex. Soft constraints. In F. Rossi, P. van Beek, and T. Walsh, editors, *Handbook of Constraint Programming*, chapter 9. Elsevier, 2006.

[28] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, San Francisco, 1988.

[29] A. Rosenfeld, R. A. Hummel, and S. W. Zucker. Scene labeling by relaxation operations. *IEEE Trans. on Systems, Man, and Cybernetics*, 6(6):420–433, June 1976.

[30] M. I. Schlesinger. False minima of the algorithm for minimizing energy of max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR. Unpublished, 1976.

[31] M. I. Schlesinger. Sintaksicheskiy analiz dvumernykh zritelnikh signalov v usloviyakh pomekh (Syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika*, 4:113–130, 1976. In Russian.

[32] M. I. Schlesinger. Lectures on labeling problems attended by the authors, Kiev, Prague, Dresden, 1996-2006.

[33] M. I. Schlesinger and B. Flach. Some solvable subclasses of structural recognition problems. In *Czech Patt. Recog. Workshop*, 2000.

[34] 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. In *European Conf. Computer Vision (ECCV)*, pages II: 16–29, 2006.

[35] S. Verdú and H. V. Poor. Abstract dynamic programming models under commutativity conditions. *SIAM J. Control and Optimization*, 25(4):990–1006, July 1987.

[36] M. Wainwright, T. Jaakkola, and A. Willsky. MAP estimation via agreement on (hyper)trees: message passing and linear programming approaches. *IEEE Trans. Information Theory*, 51(11):3697–3717, 2005.

[37] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. A new class of upper bounds on the log partition function. *IEEE Trans. Information Theory*, 51(7):2313–2335, 2005.

[38] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical Report 649, UC Berkeley, Dept. of Statistics, 2003.

[39] D. L. Waltz. Generating semantic descriptions from drawings of scenes with shadows. Technical report, Massachusetts Institute of Technology, 1972.

[40] T. Werner. A linear programming approach to max-sum problem: A review. Technical Report CTU–CMP–2005–25, Center for Machine Perception, Czech Technical University, December 2005.

[41] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007. To appear.

[42] C. Yanover, T. Meltzer, and Y. Weiss. Linear programming relaxations and belief propagation: An empirical study. *Machine Learning Research*, 7:1887–1907, September 2006.