# What Is Decreased by the Max-sum Arc Consistency Algorithm?

Tomáš Werner          WERNER@CMP.FELK.CVUT.CZ

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

## Abstract

Inference tasks in Markov random fields (MRFs) are closely related to the constraint satisfaction problem (CSP) and its soft generalizations. In particular, MAP inference in MRF is equivalent to the weighted (max-sum) CSP. A well-known tool to tackle CSPs are arc consistency algorithms, a.k.a. relaxation labeling. A promising approach to MAP inference in MRFs is linear programming relaxation solved by sequential tree-reweighted message passing (TRW-S). There is a not widely known algorithm equivalent to TRW-S, max-sum diffusion, which is slower but very simple. We give two theoretical results. First, we show that arc consistency algorithms and max-sum diffusion become the same thing if formulated in an abstract-algebraic way. Thus, we argue that max-sum arc consistency algorithm or max-sum relaxation labeling is a more suitable name for max-sum diffusion. Second, we give a criterion that strictly decreases during these algorithms. It turns out that every class of equivalent problems contains a unique problem that is minimal w.r.t. this criterion.

## 1. Introduction

A promising approach to (approximate) computing the maximum of a discrete Gibbs probability distribution or, equivalently, MAP inference in a discrete Markov random field (MRF, a.k.a. undirected graphical model) is the linear programming relaxation proposed independently by (Schlesinger, 1976b; Koster et al., 1998; Chekuri et al., 2001; Wainwright et al., 2005). It is known that rather than to solve this linear programming relaxation directly, it is more efficient first to solve its linear programming dual. This

dual corresponds to minimizing a convex upper bound by reparameterizations. The resulting linear program cannot be solved by general-purpose (such as simplex or interior point) algorithms. Instead, efficient *network* algorithms, utilizing the specific structure of the task, are needed. The most well-known repres+ant of such algorithms is the sequential tree-reweighted message passing (TRW-S) by (Kolmogorov, 2006).

A less known such algorithm is *max-sum diffusion* (Kovalevsky & Koval, approx 1975). It is an algorithm to decrease Schlesinger's upper bound (Schlesinger, 1976b; Schlesinger & Giginjak, 2007), which is equivalent to the bound decreased by TRW-S. These works are published only in Russian or unpublished but they have been recently surveyed[1] in (Werner, 2005; Werner, 2007). Max-sum diffusion is fascinatingly simple (although slower than TRW-S on monotonic chains), hence very easy to implement and suitable for mathematical analysis. In fact, it can be seen as a special case of TRW-S where the trees are the individual nodes and edges of the problem graph. Max-sum diffusion resembles belief propagation but there is an essential difference: it has been always observed to converge and it seeks to minimize a convex function. It is also different from the edge-based reparameterizations in (Wainwright et al., 2005). It shares with TRW-S the drawback that its fixed points not necessarily correspond to the minima of the upper bound.

From the optimization point of view, inference tasks in MRF are closely related to *constraint satisfaction problems* (CSPs). The ordinary (crisp) CSP (Waltz, 1972; Mackworth, 1991), which is also known as the consistent labeling problem (Haralick & Shapiro, 1979), seeks to assign states to discrete variables that satisfy a given set of crisp constraints (i.e., relations). As this formulation is too rough for many applications, a number of soft modifications of CSPs have been proposed, where one seeks to optimize soft constraints

---

[1] My knowledge of max-sum diffusion and my works (Werner, 2005; Werner, 2007) are based on unpublished lectures of M. I. Schlesinger and my personal communication with him.

rather than satisfy crisp ones. Examples are the max-CSP, the fuzzy, partial, weighted, probabilistic, and possibilistic CSP; references and terminology can be found e.g. in (Bistarelli et al., 1999b; Bistarelli, 2004; Meseguer et al., 2006). MAP inference in MRF coincides with what we further call the *max-sum CSP*.

Many properties of the CSP and soft CSPs can be studied in a unified algebraic framework. Particular CSPs are obtained by choosing the appropriate algebraic structures. Two approaches exist: the *valued CSP* uses a totally ordered set with a single operation (Schiex et al., 1995; Bistarelli et al., 1999b), the *semiring-based CSP* uses two abstract operations that form a commutative semiring (Bistarelli et al., 1999b; Bistarelli, 2004; Schlesinger & Flach, 2000). This algebraic formulation is similar to semiring unification of dynamic programming and message passing by (Verdú & Poor, 1987; Aji & McEliece, 2000).

As the CSP is NP-complete, there is no hope for a polynomial algorithm to prove or disprove its satisfiability, unless P=NP. However, there is a number of polynomial conditions necessary (but not sufficient) for satisfiability. They are typically based on local consistencies, surveyed e.g. in (Bessiere, 2006). Of them, the simplest and most important is *arc consistency* (Hentenryck et al., 1992). The analogical concept is frequent also in MRFs and graphical models, though usually under different names.

Arc consistency was originally formulated for the crisp CSP (Waltz, 1972). The algorithm achieving arc consistency is also known as *discrete relaxation labeling* by (Rosenfeld et al., 1976). Considerable effort has been devoted to generalizing arc consistency to soft constraints. The earliest such work was done probably by (Rosenfeld et al., 1976) for the fuzzy CSP, and later e.g. by (Bistarelli et al., 1999b; Bistarelli et al., 1999a; Bistarelli & Gadducci, 2006; Cooper & Schiex, 2004; Cooper et al., 2007).

We present two theoretical results.

First, we generalize equivalence of instances and the upper bound given in (Schlesinger, 1976b) and max-sum diffusion to the valued CSP. In this algebraic formulation, max-sum diffusion becomes the same thing as arc consistency algorithms (relaxation labeling). This relates max-sum diffusion (and message passing in general) to the large literature on arc consistency. In a way, this can be seen as a similar unification for arc consistency as (Aji & McEliece, 2000) was for message passing. It can be also seen as a continuation of the work by (Rosenfeld et al., 1976) on relaxation labeling. Based on this observation, we argue that max-sum diffusion could be called *max-sum arc consistency*
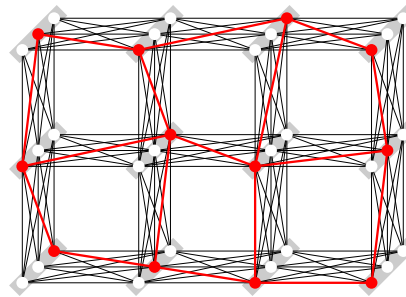


*Figure 1.* An example of a binary CSP. Graph $\langle T, E \rangle$ is the $3 \times 4$ grid graph, thus it has $|T| = 12$ nodes. There are $|X| = 3$ labels. An example of labeling $\mathbf{x}$ is emphasized.
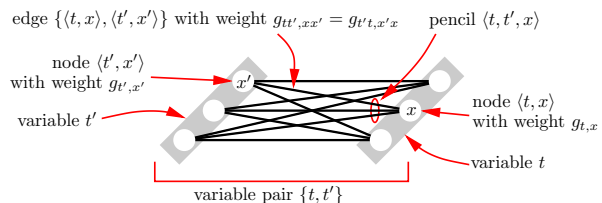


*Figure 2.* Illustration of the used terminology.

algorithm or *max-sum relaxation labeling*.

Second, we give a criterion that decreases after every iteration of the arc consistency algorithm. This contrasts the fact that the known convex upper bounds on the objective function, sought to be minimized by the max-sum diffusion and TRW-S, either decrease or remain unchanged after every iteration. Every set of equivalent instances contains at most one instance that is minimal w.r.t. the new criterion. We will show that the arc consistency algorithm can be interpreted as a block-coordinate descent w.r.t. to the new criterion.

## 2. Constraint Satisfaction Problems

In the sequel, we will denote a set by $\{\cdots\}$ and an ordered tuple by $\langle \cdots \rangle$. Ordered tuples will be typed in bold-face and their components in italics, as $\mathbf{g} = \langle g_i \rangle$. For clarity, we will consider only CSPs with unary and binary constraints. However, the results can be straightforwardly extended for problems of 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. We will denote by $N_t = \{ t' \in T \mid \{t, t'\} \in E \}$ the neighbors of $t$. Each variable $t \in T$ is assigned a single **label** $x_t \in X$, where the label domain $X$ is a finite set. An assignment of a single label $x_t$ to each variable $t$ is a **labeling** $\mathbf{x} = \langle x_t \mid t \in T \rangle \in X^T$. See 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 graph $\langle T \times X, E_X \rangle$, while the nodes and edges of graph $\langle T, E \rangle$ will be called

**variables** and **variable pairs**. The set of all nodes and edges is $I = (T \times X) \cup E_X$. The $|X|$ edges leading from a node $\langle t, x \rangle$ to all nodes of a neighboring variable $t' \in N_t$ is a **pencil** $\langle t, t', x \rangle$. The set of all pencils is $P = \{\, \langle t, t', x \rangle \mid \{t, t'\} \in E, \; x \in X \,\}$.

Let $A$ be a set of **values**. Each node $\langle t, x \rangle$ is assigned a value $g_{t,x} \in A$ and each edge $\{\langle t, x \rangle, \langle t', x' \rangle\}$ a value $g_{tt',xx'} \in A$, where we adopt $g_{tt',xx'} = g_{t't,x'x}$. By $\mathbf{g} \in A^I$ we denote the vector with components $g_{t,x}$, $g_{tt',xx'}$. These concepts are illustrated in Figure 2.

**Definition 1.** *A* **valuation structure** *(Schiex et al., 1995; Bistarelli et al., 1999b) is a triplet $\langle A, \le, \otimes \rangle$ where $\le$ is a total order on $A$, and $\otimes$ is a binary associative and commutative operation that is closed in $A$ and monotonic w.r.t. $\le$, i.e.*

$$(a \le b) \implies [(a \otimes c) \le (b \otimes c)] \qquad (1)$$

Note that $\langle A, \max \rangle$ and $\langle A, \otimes \rangle$ are commutative semigroups and $\langle A, \max, \otimes \rangle$ is a commutative semiring[2].

Let max denote the maximum with respect to $\le$. Let

$$F(\mathbf{x} \mid \mathbf{g}) = \left( \bigotimes_{t \in T} g_{t,x_t} \right) \otimes \left( \bigotimes_{\{t,t'\} \in E} g_{tt',x_t x_{t'}} \right) \quad (2a)$$

$$F(\mathbf{g}) = \max_{\mathbf{x} \in X^T} F(\mathbf{x} \mid \mathbf{g}) \qquad (2b)$$

Note, (2a) can be seen as a function $F(\cdot \mid \mathbf{g})\colon X^T \to A$ that is pairwise-decomposable using operation $\otimes$. Expression (2b) is known as the **valued CSP (VCSP)**.

## 3. Equivalent Instances

It is an important observation that function (2a) can be identical for different value vectors $\mathbf{g}$. This leads to the concept of equivalent value vectors and closely related local equivalent transformations. Equivalence is important because it allows to transform value vector to a form most convenient for (approximate) calculation of expression (2b).

The following definition generalizes equivalence of max-sum CSP instances given by (Schlesinger, 1976b).

**Definition 2.** *Value vectors $\mathbf{g}, \mathbf{g}' \in A^I$ are* **equivalent** *(denoted by $\mathbf{g} \sim \mathbf{g}'$) if $F(\mathbf{x} \mid \mathbf{g}) = F(\mathbf{x} \mid \mathbf{g}')$ for all $\mathbf{x} \in X^T$, i.e., if functions $F(\cdot \mid \mathbf{g})$ and $F(\cdot \mid \mathbf{g}')$ equal.*

In the definition, set $I$ and functions $F(\mathbf{x} \mid \mathbf{g})$ and $F(\mathbf{x} \mid \mathbf{g}')$ are considered w.r.t. the same $T$, $E$, $X$, $A$, $\otimes$. Note, equivalence does not depend on the order $\le$.

The equivalence relation $\sim$ partitions the set $A$ into the set $A/\!\!\sim$ of equivalence classes.

---

[2]Valuation structures and commutative semiring are often defined also with the zero and unit elements. But we will not need the zero and unit.
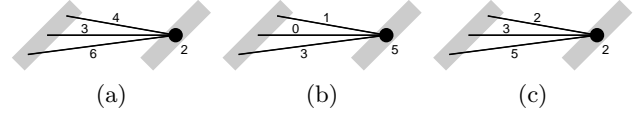


*Figure 3.* For $\langle A, \otimes \rangle = \langle \mathbb{R}, + \rangle$, pencils (a) and (b) are related by a local equivalent transformation because $4 + 2 = 1 + 5$, $3 + 2 = 0 + 5$, and $6 + 2 = 3 + 5$. For $\langle A, \otimes \rangle = \langle \mathbb{R}, \min \rangle$, pencils (a) and (c) are related by a local equivalent transformation because $\min\{4, 2\} = \min\{2, 2\}$, $\min\{3, 2\} = \min\{3, 2\}$, and $\min\{6, 2\} = \min\{5, 2\}$.

Two tasks related to equivalence naturally arise: *test* whether two given instances are equivalent and *enumerate* the elements of an equivalence class. These tasks may be easy or hard depending on $\langle A, \otimes \rangle$.

**Example 1.** Let $\langle A, \otimes \rangle = \langle \{0, 1\}, \times \rangle$. 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-complete. ∎

**Definition 3.** *A* **local equivalent transformation** *of problem $\mathbf{g}$ on pencil $\langle t, t', x \rangle$ is a replacement of values $g_{t,x}$ and $\{\, g_{tt',xx'} \mid x' \in X \,\}$ with values $g'_{t,x}$ and $\{\, g'_{tt',xx'} \mid x' \in X \,\}$ such that the latter satisfy*

$$g'_{t,x} \otimes g'_{tt',xx'} = g_{t,x} \otimes g_{tt',xx'} \quad \forall x' \in X \qquad (3)$$

Note, a local equivalent transformation affects only the $|X|$ edges in a *single* pencil $\langle t, t', x \rangle$ and the node $\langle t, x \rangle$, incident with the pencil. Figure 3 shows examples.

By (2a) and (3), local equivalent transformations preserve the function $F(\cdot \mid \mathbf{g})$. Thus, they allow to traverse through the equivalence class by *local* operations. Sometimes all elements of the equivalence class can be enumerated like this, sometimes not. In detail, three cases can occur for equivalent value vectors $\mathbf{g}$ and $\mathbf{g}'$:

1. $\mathbf{g}$ can be transformed to $\mathbf{g}'$ by a finite sequence of local equivalent transformations;

2. there is an infinite sequence of local equivalent transformations of $\mathbf{g}$ that converges to $\mathbf{g}'$ (as shown by Example 3 below);

3. $\mathbf{g}$ cannot be transformed to $\mathbf{g}'$ by local equivalent transformations (as shown by Example 1).

An important special case is when $\langle A, \otimes \rangle$ is a group (e.g., $\langle \mathbb{R}, + \rangle$ or $\langle \mathbb{R}_{++}, \times \rangle$, see §6.3). Thus, we have division (denoted by $\oslash$), the unit element, and inverse. In this case, all possible local equivalent transformations on pencil $\langle t, t', x \rangle$ can be parameterized by a single value $\varphi_{tt',x} \in A$ (assigned to pencil $\langle t, t', x \rangle$) as

$$g'_{t,x} \quad = g_{t,x} \quad \oslash \varphi_{tt',x} \qquad (4a)$$

$$g'_{tt',xx'} = g_{tt',xx'} \otimes \varphi_{tt',x} \qquad (4b)$$

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

$$g'_{t,x} \quad = g_{t,x} \quad \oslash \left( \bigotimes_{t' \in N_t} \varphi_{tt',x} \right) \quad (5a)$$

$$g'_{tt',xx'} = g_{tt',xx'} \otimes \varphi_{tt',x} \otimes \varphi_{t't,x'} \quad (5b)$$

Theorem 1 shows[3] that of the three cases above, Case 1 occurs for groups.

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

*Proof.* See (Werner, 2005; Werner, 2007). ∎

Values $\varphi_{tt',x}$ are often called *messages* (Pearl, 1988) and transformations (4) and (5) *reparameterizations* (Wainwright & Jordan, 2003; Kolmogorov, 2006).

Another important special case is the semigroup $\langle A, \otimes \rangle = \langle A' \cup \{0_\otimes\}, \otimes \rangle$, where $\langle A', \otimes \rangle$ is a group and $0_\otimes$ satisfies $a \otimes 0_\otimes = 0_\otimes$ for all $a \in A'$. Here, reparameterizations (5) cover only the part of the equivalence class obtained for $\boldsymbol{\varphi} \in (A')^P$. Letting $\boldsymbol{\varphi} \in A^P$ would not help because expressions $a \oslash 0_\otimes$ and $0_\otimes \oslash 0_\otimes$ are undefined and cannot be defined consistently. Theorem 1 does not hold and Cases 2 or 3 can occur, as given by the two examples below.

**Example 2.** Let $\langle A, \otimes \rangle = \langle \mathbb{R} \cup \{-\infty\}, + \rangle$ (see §6.3). This semigroup has the subsemigroup $\langle \{-\infty, 0\}, + \rangle$, which is isomorphic to $\langle \{0, 1\}, \times \rangle$. By Example 1, testing whether two instances are equivalent is NP-hard and Theorem 1 does not hold. ∎

**Example 3.** Let $\langle A, \otimes \rangle = \langle \mathbb{R} \cup \{-\infty\}, + \rangle$. Consider two instances in Figure 4. These instances are equivalent but not related by (5) for any $\varphi_{tt',x} \in \mathbb{R}$. However, there is a sequence of local equivalent transformations of problem (a) that converges to problem (b). This sequence is the infinite repetition of reparameterizations (5) with values $\varphi_{tt',x}$ shown in the figure. ∎

## 4. Criteria to Minimize

Approximate computation of $F(\mathbf{g})$ is approached via upper bounds. Theorem 2 gives two such upper bounds on $F(\mathbf{g})$, obtained by generalizing upper bound by (Schlesinger, 1976b) (see §6.3) to VCSPs.

**Theorem 2.** *Denote*

$$\bar{F}(\mathbf{g}) = \left( \bigotimes_{t \in T} \max_{x \in X} g_{t,x} \right) \otimes \left( \bigotimes_{\{t,t'\} \in E} \max_{x,x' \in X} g_{tt',xx'} \right) \quad (6)$$

*For any $\mathbf{g} \in A^I$, we have*

$$F(\mathbf{g}) \leq \bar{F}(\mathbf{g}) \leq \left( \max_{i \in I} g_i \right)^{|T|+|E|} \quad (7)$$

---
[3]Theorem 1 and its proof was given in unpublished lectures by M. I. Schlesinger, attended by the author.
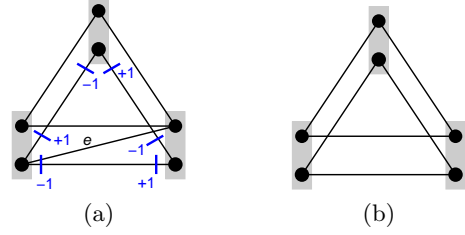


(a)          (b)

*Figure 4.* In these instances with $\langle A, \otimes \rangle = \langle \mathbb{R} \cup \{-\infty\}, + \rangle$, the nodes and the shown edges have values 0, and the edges not shown have values $-\infty$. Cyclic repetition of the depicted local equivalent transformations decreases the value of edge $e$ to an arbitrarily small value, keeping all other values unchanged.

*where the power in (7) is taken using operation $\otimes$.*

*Proof.* Using that $\otimes$ distributes over max, open the parentheses in (6) and write it as a single maximum of elements. It turns out that these elements will be all the elements present in (2b) plus some additional ones, i.e., $\bar{F}(\mathbf{g}) = \max\{F(\mathbf{g}), \cdot\}$. Hence $F(\mathbf{g}) \leq \bar{F}(\mathbf{g})$. It is easy to verify that any $a, b \in A$ satisfy

$$a \otimes b \leq (\max\{a,b\})^2 \quad (8)$$

To see this, suppose $a \leq b$ (if not, swap $a$ and $b$). Since $\max\{a,b\} = b$, (8) reads $a \otimes b \leq b^2$. But this is true, by setting $c = b$ in the axiom (1).

By induction, (8) can be generalized to $\bigotimes_{i=1}^n a_i \leq (\max_{i=1}^n a_i)^n$. Hence $\bar{F}(\mathbf{g}) \leq (\max_i g_i)^{|T|+|E|}$. ∎

A good approximation of $F(\mathbf{g})$ can be obtained by minimizing (or at least decreasing) the upper bounds over the equivalence class of $\mathbf{g}$ (or its subset). This can be done using a sequence of suitable local equivalent transformations, in other words, by a block-coordinate descent. Recall, *(block-)coordinate descent* is a method that in every iteration minimizes a given criterion over a single variable (or a set of variables), keeping the other variables fixed. In our case, the free variables are the values $g_{t,x}$ and $\{g_{tt',xx'} \mid x' \in X\}$ on a single pencil $\langle t, t', x \rangle$, while all other values $g_i$ are fixed. That is, every iteration is a local equivalent transformation.

As will be shown later, arc consistency algorithms and max-sum diffusion monotonically decrease the quantity $\max_i g_i$, hence the second upper bound in (7). It is tempting to interpret these algorithms as a block-coordinate descent to minimize $\max_i g_i$. This interpretation is flawed because the bound does not decrease strictly, i.e., in every iteration the bound either decreases or remains unchanged. We will show that what these algorithms do is a block-coordinate descent on a *different* criterion.

**Definition 4.** *For $\mathbf{g} \in A^I$, let sort $\mathbf{g}$ denote the ele-*

ment of $A^I$ obtained by sorting the components $g_i$ of $\mathbf{g}$ in the descending order w.r.t. $\leq$. Let $\leq_{\text{lex}}$ denote the lexicographic order on $A^I$ induced by $\leq$.

We define the relations $\sqsubseteq$ and $\sqsubset$ on $A^I$ by

$$(\mathbf{g}' \sqsubseteq \mathbf{g}) \quad \Longleftrightarrow \quad (\text{sort } \mathbf{g}' \leq_{\text{lex}} \text{sort } \mathbf{g})$$
$$(\mathbf{g}' \sqsubset \mathbf{g}) \quad \Longleftrightarrow \quad (\mathbf{g}' \sqsubseteq \mathbf{g}) \wedge (\text{sort } \mathbf{g}' \neq \text{sort } \mathbf{g})$$

It is easy to see that the relation $\sqsubseteq$ is a total order[4]. We say that $\mathbf{g} \in A^I$ is minimal w.r.t. $\sqsubseteq$ on a set $G \subseteq A^I$ iff $\mathbf{g} \sqsubseteq \mathbf{g}'$ for all $\mathbf{g}' \in G$. Theorem 3 is a trivial consequence of the fact that $\sqsubseteq$ is a total order.

**Theorem 3.** *Any set $G \subseteq A^I$ contains at most one element that is minimal w.r.t. $\sqsubseteq$.*

*Proof.* Set $G$ is totally ordered by $\sqsubseteq$. It is well-known that if a totally (unlike partially) ordered set contains a minimal element than this element is unique. ∎

By letting $G$ to be the set of value vectors reachable by all local equivalent transformations on a single pencil $\langle t, t', x \rangle$, we obtain the main result of our paper.

**Proposition 1.** *Arc consistency algorithms and max-sum diffusion can be interpreted as the block-coordinate descent where in every iteration, value vector $\mathbf{g}$ is minimized w.r.t. $\sqsubseteq$ by a local equivalent transformation.*

Let $\mathbf{g}$ and $\mathbf{g}'$ be the value vector before and after any non-vacuous iteration (i.e., such that $\mathbf{g}' \neq \mathbf{g}$). Then clearly $\mathbf{g}' \sqsubset \mathbf{g}$, meaning that $\mathbf{g}$ is strictly decreased w.r.t. $\sqsubseteq$. In contrast, we have the implication

$$(\mathbf{g}' \sqsubset \mathbf{g}) \implies \left( \max_{i \in I} g_i' \leq \max_{i \in I} g_i \right) \tag{9}$$

but $\mathbf{g}' \sqsubset \mathbf{g}$ does not imply $\max_i g_i' < \max_i g_i$. Hence, if $\mathbf{g}$ stricly decreases w.r.t. $\sqsubseteq$ then $\max_i g_i$ either decreases or remains unchanged.

Importantly, the order $\sqsubseteq$ allows to define optimal instances in a new way: rather than minimizing any of the bounds in (7) on the equivalence class, we minimize $\mathbf{g}$ w.r.t. $\sqsubseteq$. Doing this does not seem to have any immediate practical advantage but it is very natural matematically because (by Theorem 3) this new optimum is unique. This is in contrast to the fact that the value vector $\mathbf{g}$ minimizing any bound in (7) is typically not unique because many components of $\mathbf{g}$ can be freely changed (within the equivalence class). An interesting example of this follows.

**Example 4.** Let $G$ be the equivalence class of the instance in Figure 4a. This instance minimizes $\max_i g_i$

---

[4]As the logical expression $\mathbf{g}' \sqsubseteq \mathbf{g}$ is invariant to permuting the elements of $\mathbf{g}$ and $\mathbf{g}'$, it would be also possible to define the relation $\sqsubseteq$ on unordered tuples (multisets) rather than on ordered tuples of elements from $A$.
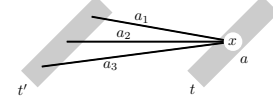


*Figure 5.* The values entering an arc consistency transformation on pencil $\langle t, t', x \rangle$, for $X = \{1, 2, 3\}$.

on $G$ but it is not minimal w.r.t. $\sqsubseteq$ because edge $e$ can be decreased by the depicted reparameterization. In contrast, the instance in Figure 4b is minimal w.r.t. $\sqsubseteq$ on $G$ (and, by (9), minimizes also $\min_i g_i$). ∎

## 5. Arc Consistency

This section defines arc consistency and an algorithm to achieve it in the algebraic framework developed in §2 and §3. It will turn out that this algorithm is the block-coordinate descent from Proposition 1 for a special class of valuation structures.

**Definition 5.** *Pencil $\langle t, t', x \rangle$ is **arc consistent** if*

$$g_{t,x} = \max_{x' \in X} g_{tt',xx'} \tag{10}$$

Note, arc consistency does not depend on $\otimes$.

**Definition 6.** *An **arc consistency transformation** on pencil $\langle t, t', x \rangle$ is a local equivalent transformation that makes the pencil arc consistent.*

Doing an arc consistency transformation on pencil $\langle t, t', x \rangle$ means solving the system of two equations (3) and (10) for $g'_{t,x}$ and $\{ g'_{tt',xx'} \mid x' \in X \}$. Dropping the indices $t, t', x$ (which are constant because we consider only a single pencil $\langle t, t', x \rangle$) and denoting

$$a = g_{t,x}, \ \ a' = g'_{t,x}, \ \ a_{x'} = g_{tt',xx'}, \ \ a'_{x'} = g'_{tt',xx'} \tag{11}$$

(see Figure 5), this system reads

$$a' \otimes a'_{x'} = a \otimes a_{x'} \qquad \forall x' \in X \tag{12a}$$
$$a' = \max_{x' \in X} a'_{x'} \tag{12b}$$

Given $a$ and $a_{x'}$, (12) has to be solved for $a'$ and $a'_{x'}$.

The arc consistency transformation is well defined only on a subclass of valuation structures.

**Definition 7.** *Valuation structure $\langle A, \leq, \otimes \rangle$ is an **AC-valuation structure** if the system (12) has a single solution $a', a'_{x'}$ for any $a, a_{x'} \in A$.*

The solution of (12) cannot be given in closed form. Only $a'$ can be given in closed form. Maximizing (12a) over $x'$ and substituting (12b) in (12a) yields

$$a' = \max_{x' \in X} a'_{x'} = \left( a \otimes \max_{x' \in X} a_{x'} \right)^{1/2} \tag{13}$$

This shows that on an AC-valuation structure, semigroup $\langle A, \otimes \rangle$ must possess the unique square root, i.e.,

equation $x^2 = x \otimes x = a$ must have a unique solution $x = a^{1/2}$ for any $a \in A$.

**Algorithm 1.** (Arc consistency algorithm)

**loop**
- Choose an arc inconsistent pencil (if none exists, halt). This choice can be arbitrary such that every pencil has a non-zero probability to be chosen.
- Do arc consistency transformation on the pencil.

**end loop**

**Conjecture 1.** *Let $\langle A, \leq, \otimes \rangle$ be an AC-valuation structure. Then Algorithm 1 converges to a state where all pencils are arc-consistent.*

The proof of this conjecture is known if the operation $\otimes$ is idempotent[5] (more in §6.1) but it is unknown if $\langle A, \otimes \rangle = \langle \mathbb{R}, + \rangle$, i.e., for max-sum diffusion.

Theorem 4 below shows that the arc consistency transformation strictly decreases $\mathbf{g}$ w.r.t. $\sqsubseteq$. For its proof, we will need Lemma 1.

**Lemma 1.** *Let $J \subseteq I$. Let $\mathbf{g}, \mathbf{g}' \in A^I$ satisfy $g_i = g_i'$ for all $i \in I \setminus J$. Then the following implication holds:*

$$\left( \max_{i \in J} g_i' < \max_{i \in J} g_i \right) \implies (\mathbf{g}' \sqsubset \mathbf{g}) \qquad (14)$$

*Proof.* Obvious and hence omitted. ∎

**Theorem 4.** *Let $\mathbf{g}$ and $\mathbf{g}'$ be an instance before and after applying the arc consistency transformation on any arc inconsistent pencil, respectively. Then $\mathbf{g}' \sqsubset \mathbf{g}$.*

*Proof.* Let $\langle t, t', x \rangle$ be an arc inconsistent pencil and let $J = \{\langle t, x \rangle\} \cup \{ \{\langle t, x \rangle, \langle t', x' \rangle\} \mid x' \in X \}$.

Let $a$ and $a'$ be given as in (11) and $b = \max_{x'} g_{tt',xx'}$. By (13), we have $a' = (a \otimes b)^{1/2}$.

As $\langle A, \otimes \rangle$ has the unique square root (§5), (8) implies $(a \otimes b)^{1/2} \leq \max\{a, b\}$. As $\langle t, t', x \rangle$ is arc inconsistent, we have $a \neq b$. It can be shown that $a \neq b$ implies that the inequality in (8) is strict. Therefore we have $(a \otimes b)^{1/2} = a' < \max\{a, b\}$. But $a' < \max\{a, b\}$ is the same as $\max_{i \in J} g_i' < \max_{i \in J} g_i$. Apply Lemma 1. ∎

## 6. Particular Cases of VCSPs

This section will give concrete examples of the valuation structure $\langle A, \leq, \otimes \rangle$ to which the results obtained in the above developed algebraic framework apply.

### 6.1. Max-min CSP

Let $\langle A, \leq \rangle$ be an arbitrary totally ordered set and let $\otimes$ be the minimum induced by $\leq$. Then $\langle A, \leq, \otimes \rangle$ is an AC-valuation structure. We will call this VCSP the *max-min CSP*.

---

[5]An operation $\otimes$ is idempotent if $a \otimes a = a$.
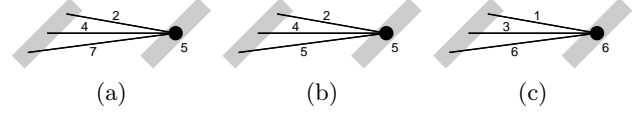


(a)    (b)    (c)

*Figure 6.* In the max-min CSP (i.e., $\langle A, \otimes \rangle = \langle \mathbb{R}, \min \rangle$), the arc consistency transformation of the pencil in (a) results in (b). In the max-sum CSP (i.e., $\langle A, \otimes \rangle = \langle \mathbb{R}, + \rangle$), the arc consistency transformation of (a) results in (c).

As the minimum is idempotent, $a \otimes a = \min\{a, a\} = a$, we have $(\max \mathbf{g})^{|T|+|E|} = \max \mathbf{g}$ in (7).

The arc consistency transformation is given by

$$a' = \min\left\{ a, \max_{x' \in X} a_{x'} \right\} \qquad (15a)$$
$$a'_{x'} = \min\{ a, a_{x'} \} \qquad \forall x' \in X \qquad (15b)$$

An example is in Figure 6. It can be verified that (15) is the unique solution to (12).

It is known (Bistarelli et al., 1999b; Bistarelli & Gadducci, 2006; Cooper & Schiex, 2004) that for the max-min CSP, Algorithm 1 converges in a finite number of iterations and the result does not depend on their order. This proves Conjecture 1.

### 6.2. The Crisp CSP

The crisp CSP is the max-min CSP where $A$ has only two elements. The natural choice is $A = \{0, 1\}$ and $0 \leq 1$. The numbers $g_{t,x}$ and $g_{tt',xx'}$ encode unary and binary relations, and the task is to find a labeling $\mathbf{x} \in X^T$ that satisfies all the relations. The CSP instance is satisfiable if $F(\mathbf{g}) = 1$ and unsatisfiable if $F(\mathbf{g}) = 0$.

Theorem 2 has the following meaning. We have $\bar{F}(\mathbf{g}) = 0$ iff all nodes in any variable $t$ or all edges in any variable pair $\{t, t'\}$ are zero, i.e., $\max_x g_{t,x} = 0$ or $\max_{x,x'} g_{tt',xx'} = 0$. Clearly, in this case the instance is unsatisfiable, which is what inequality $F(\mathbf{g}) \leq \bar{F}(\mathbf{g})$ says. The second upper bound, $\max_i g_i$, is zero iff all nodes and edges in the whole instance are zero.

Algorithm 1 is the well-known discrete relaxation labeling (arc consistency algorithm) (Waltz, 1972; Rosenfeld et al., 1976; Hentenryck et al., 1992).

### 6.3. Max-sum CSP

Let $A = \mathbb{R} \cup \{-\infty\}$, let $\leq$ be the natural order, and let $\otimes$ be the ordinary addition. Then $\langle A, \leq, \otimes \rangle$ is an AC-valuation structure. We will call the resulting VCSP the *max-sum CSP*. Details on it can be found in survey (Werner, 2005; Werner, 2007).

The upper bound (6) becomes Schlesinger's max-sum upper bound (Schlesinger, 1976b; Werner, 2007),

$$\bar{F}(\mathbf{g}) = \sum_{t \in T} \max_{x \in X} g_{t,x} + \sum_{\{t,t'\} \in E} \max_{x,x' \in X} g_{tt',xx'} \qquad (16)$$

Minimizing (16) by reparameterizations (5) reads

$$\min \left\{ \bar{F}(\mathbf{g} + \mathbf{A}\boldsymbol{\varphi}) \mid \boldsymbol{\varphi} \in \mathbb{R}^P \right\} \qquad (17)$$

where $\mathbf{A}$ is the matrix with elements $\{-1, 0, +1\}$ such that the formula $\mathbf{g}' = \mathbf{g} + \mathbf{A}\boldsymbol{\varphi}$ is the formula (5) for $\otimes$ replaced with $+$ and $\oslash$ replaced with $-$. Problem (17) is convex and can be posed and a linear programming.

The linear programming dual of (17) is the well-known linear programming relaxation of the max-sum CSP (Schlesinger, 1976b; Koster et al., 1998; Chekuri et al., 2001; Wainwright et al., 2005)

$$\max \left\{ \mathbf{g}^\top \boldsymbol{\mu} \ \middle| \ \boldsymbol{\mu} \geq \mathbf{0}, \ \mu_{t,x} = \sum_{x' \in X} \mu_{tt',xx'}, \ \sum_{x \in X} \mu_{t,x} = 1 \right\}$$

where $\boldsymbol{\mu} \in \mathbb{R}^I$ is a vector with components $\mu_{t,x}$ and $\mu_{tt',xx'}$, often called pseudomarginals.

As shown in (Werner, 2007), the optimum of problem (17) is the same as the optimum of

$$\min \left\{ (|T| + |E|) \max(\mathbf{g} + \mathbf{A}\boldsymbol{\varphi}) \mid \boldsymbol{\varphi} \in \mathbb{R}^P \right\} \qquad (18)$$

where $\max \mathbf{g}'$, for $\mathbf{g}' = \mathbf{g} + \mathbf{A}\boldsymbol{\varphi}$, denotes $\max_i g'_i$. This is the second upper bound in (7).

The system (12) has the closed form solution

$$a' = \frac{1}{2}\left( a + \max_{x' \in X} a_{x'} \right) \qquad (19a)$$

$$a'_{x'} = a_{x'} + a - a' \qquad \forall x' \in X \qquad (19b)$$

An example is in Figure 6. The case when $a$ and/or $a_{x'}$ equals $-\infty$ must be handled separately but this is easy. Algorithm 1 becomes max-sum diffusion (Kovalevsky & Koval, approx 1975; Werner, 2005; Werner, 2007).

Algorithm 1 need not find the global minimum of problem (17), as observed by (Schlesinger, 1976a) and reviewed in (Werner, 2007). This is because (block-)coordinate descent is not guaranteed to find the global minimum of a *nonsmooth* convex problem (17) (Werner & Shekhovtsov, 2007).

### 6.4. Modifications of Max-sum CSP

In the max-sum CSP, it is of interest to consider other value sets than $A = \mathbb{R} \cup \{-\infty\}$.

Setting $A = \mathbb{R}$ simplifies things since $\langle \mathbb{R}, + \rangle$ is a group. This case corresponds to MAP inference in MRF.

If $A$ are the rational numbers, $\langle A, \leq, \otimes \rangle$ remains to be an AC-valuation structure. This case is very similar to $A = \mathbb{R}$, since the optima of linear programs with rational coefficient are rational.

If $A$ are the integer numbers, $\langle A, \leq, \otimes \rangle$ is no longer an AC-valuation structure. Thus Algorithm 1 cannot be used. However, we can still do the block-coordinate

descent as given in Proposition 1.

## 7. Conclusion

We have given a unified algebraic framework including arc consistency (Hentenryck et al., 1992), discrete relaxation labeling (Rosenfeld et al., 1976), max-sum diffusion (Kovalevsky & Koval, approx 1975), and, more broadly, sequential message passing algorithms for MAP inference in MRF (Wainwright et al., 2005; Kolmogorov, 2006). This has also showed that max-sum diffusion can be naturally considered a soft generalization of crisp arc consistency algorithm, as discussed in (Cooper & Schiex, 2004; Cooper et al., 2007).

As part of this framework, we have given a criterion that strictly decreases during the arc consistency algorithm. Importantly, this allows to define optimal CSP instances in a new way, such that every equivalence/reparameterization class contains at most one optimal instance. It also shows that arc consistency algorithms, relaxation labeling, and max-sum diffusion can be seen as an algorithm to decrease this new criterion by (block)-coordinate descent.

## Acknowledgement

## References

Aji, S. M., & McEliece, R. J. (2000). The generalized distributive law. *IEEE Trans. on Information Theory*, *46*, 325–343.

Bessiere, C. (2006). Constraint propagation. In F. Rossi, P. van Beek and T. Walsh (Eds.), *Handbook of constraint programming*, chapter 3. Elsevier. Also technical report LIRMM 06020.

Bistarelli, S. (2004). *Semirings for soft constraint solving and programming*. Lecture Notes in Computer Science. Springer Verlag.

Bistarelli, S., Codognet, P., Georget, Y., & Rossi, F. (1999a). Labeling and partial arc consistency for soft constraints. *Joint Conf. Declarative Programming (AGP), Italy* (pp. 121–136).

Bistarelli, S., & Gadducci, F. (2006). Enhancing constraints manipulation in semiring-based formalisms. *Eur. Conf. Artificial Intelligence (ECAI)* (pp. 63–67). IOS Press.

Bistarelli, S., Montanari, U., Rossi, F., Schiex, T., Verfaillie, G., & Fargier, H. (1999b). Semiring-based CSPs and valued CSPs: Frameworks, properties, and comparison. *Constraints*, *4*, 199–240.

Chekuri, C., Khanna, S., Naor, J., & Zosin, L. (2001). Approximation algorithms for the metric labeling problem via a new linear programming formulation. *Symposium on Discrete Algorithms* (pp. 109–118).

Cooper, M., de Givry, S., & Schiex, T. (2007). Optimal soft arc consistency. *Proc. IJCAI, Hyderabad, India*.

Cooper, M., & Schiex, T. (2004). Arc consistency for soft constraints. *Artificial Intelligence*, *154*, 199–227.

Haralick, R. M., & Shapiro, L. G. (1979). The consistent labeling problem. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *1*, 173–184.

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

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

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

Kovalevsky, V. A., & Koval, V. K. (approx. 1975). A diffusion algorithm for decreasing energy of max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR. Unpublished.

Mackworth, A. (1991). Constraint satisfaction. In *Encyclopedia of artificial intelligence*, 285–292. New York: Wiley.

Meseguer, P., Rossi, F., & Schiex, T. (2006). Soft constraints. In F. Rossi, P. van Beek and T. Walsh (Eds.), *Handbook of constraint programming*, chapter 9. Elsevier.

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

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

Schiex, T., Fargier, H., & Verfaillie, G. (1995). Valued constraint satisfaction problems: Hard and easy problems. *Intl. Joint Conf. on Artificial Intelligence (IJCAI)*. Montreal.

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

Schlesinger, M. I. (1976b). Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika*, *4*, 113–130. In Russian.

Schlesinger, M. I., & Flach, B. (2000). Some solvable subclasses of structural recognition problems. *Czech Patt. Recog. Workshop*.

Schlesinger, M. I., & Giginjak, V. V. (2007). Solving (max,+) problems of structural pattern recognition using equivalent transformations. *Upravlyayushchie Sistemy i Mashiny (Control Systems and Machines), Kiev, Naukova Dumka*, *1 and 2*. In Russian.

Verdú, S., & Poor, H. V. (1987). Abstract dynamic programming models under commutativity conditions. *SIAM J. Control and Optimization*, *25*, 990–1006.

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

Wainwright, M. J., & Jordan, M. I. (2003). *Graphical models, exponential families, and variational inference* (Technical Report 649). UC Berkeley, Dept. of Statistics.

Waltz, D. L. (1972). *Generating semantic descriptions from drawings of scenes with shadows* (Technical Report). Massachusetts Institute of Technology.

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

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

Werner, T. s., & Shekhovtsov, A. (2007). Unified framework for semiring-based arc consistency and relaxation labeling. *Computer Vision Winter Workshop*. St. Lambrecht, Austria.