A partial solution for D-unification based on a reduction to $AC1$-unification

Evelyne Contejean

Max-Planck-Institut für Informatik
Im Stadtwald
D-6600 Saarbrücken, Germany
contejea@mpi-sb.mpg.de

Abstract. We show that deciding unification modulo both-sided distributivity of a symbol $*$ over a symbol $+$ can be reduced to $AC1$-unification for all unification problems which do not involve the $+$ operator. Moreover, we can describe "almost all" solutions in a finite way, although there are in general infinitely many minimal solutions for such problems. As a consequence, $*$-problems appear as a good candidate for a notion of solved-form for D-unification.

1 Introduction

Equations are ubiquitous in mathematics as well as in computer science. Unification is solving equations in some particular domains, namely free term algebras or term algebras modulo an equational theory. Unification was first introduced by Herbrand [3], and rediscovered by Robinson as a basic mechanism for resolution in first order logic [10]. Unification is also the basic mechanism for computing critical pairs [7], and the main inference rule of the completion procedure.

On the other hand, the very old and famous 10^{th} Hilbert problem of solving Diophantine equations has been proved undecidable by Matijasevic [9]. The challenge now is to find the maximal subsets of Peano arithmetic for which unification is decidable. Such a candidate is actually provided by the two axioms of left and right distributivity of a symbol $*$ over a symbol $+$. Arnborg and Tidén have shown that one-sided (left *or* right) distributivity has a decidable unification problem, whereas adding a unit element for $*$ and associativity of $+$ makes unification undecidable [12]. Szabo has proved the same result when associativity of $+$ is added to both-sided distributivity [11].

In this paper, we give a partial solution for unification modulo both-sided distributivity. More precisely, we show that if a problem does not contain any $+$ function symbol, solvability modulo D and solvability in the free algebra are equivalent.

Describing all solutions of a problem P , however, is more difficult. In a first step, we show that all solutions (modulo D) are instances of the linearized form λ of the most general unifier of P in the free algebra. Not all of them, however, are solutions. The appropriate instances are topped by an $AC1$ solution of the problem $P\lambda$ in the theory where $*$ is $AC1$.

The paper is organized as follows: the second section contains the needed formal apparatus. Section 3 is devoted to $*$-problems. It introduces "indexed distributivity" as a main tool used throughout the paper. Section 4 studies the algebra of structures which are the terms used in order to represent the tops of solutions modulo distributivity. This algebra has a unique decomposition property similar to the unique decomposition property of natural numbers into prime numbers. Hence, the algebra of structures can be seen as a free commutative monoid. The description of almost all solutions of a $*$-problem follows as a corollary.

2 Definitions

We assume that the reader is familiar with the notions of *term algebra, position in a term, substitution, equational theory, equational proof, rewrite proof* surveyed in [2]. We recall briefly the most important ones. Given a *signature* $\mathcal{F}$, that is, a set of function symbols with their arity, and a set $\mathcal{X}$ of variables, the term algebra $\mathcal{T}(F, X)$ is the free $\mathcal{F}$-algebra over $\mathcal{X}$. An equational axiom is an unordered pair of terms $< l, r >$ of $\mathcal{T}(F, X)$, also denoted by $l = r$. An equational theory presented by a set of equational axioms E is the reflexive transitive closure of $\longleftrightarrow_E$, the smallest relation containing E which is compatible with the term structure of $\mathcal{T}(F, X)$. As in [2], we use postfixed notations for substitution applications.

In the following, we may not always distinguish a set of equational axioms with its associated equational theory.

Definition 2.1 *D is the equational theory presented by*

$$x * (y + z) = (x * y) + (x * z) \qquad (D_l)$$
$$(x + y) * z = (x * z) + (y * z) \qquad (D_r)$$

The axioms of D are usually oriented from left to right in this paper:

Definition 2.2 (Developed normal form) *Let t be a term in $\mathcal{T}(\{+, *\}, \mathcal{X})$. Its (possibly many) irreducible forms with respect to the following set of rules:*

$$x * (y + z) \to (x * y) + (x * z)$$
$$(x + y) * z \to (x * z) + (y * z)$$

are called developed normal forms*.*

Definition 2.3 *$AC1$ is the equational theory presented by*

$$x * (y * z) = (x * y) * z \qquad (A)$$
$$x * y = y * x \qquad (C)$$
$$x * 1 = x \qquad (1)$$

Definition 2.4 (E-Unification Problem) *Let $\mathcal{T}(F, X)$ be a term algebra, and E an equational theory over $\mathcal{T}(F, X)$. A E-*unification problem *is a first order formula, where the logical connectives are $\exists$, $\wedge$ and $\vee$, and the atoms are either T, or F, or some equation $s =^?_E t$ between two terms of $\mathcal{T}(F, X)$. The* *solutions σ of such a problem are the substitutions of $\mathcal{T}(F,X)$ that satisfy the formula when $=^?_E$ is interpreted as $=_E$.*

In this paper, we deal with unification problems modulo $AC1$, D, or in the free algebra.

Definition 2.5 ($*$-Problems) *A* $*$-equation *in $\mathcal{T}(\{+,*\},X)/D$ is an equation $s =^?_D t$, where s and t are terms of $\mathcal{T}(\{*\},X)$. A* $*$-problem *is a unification problem in the algebra $\mathcal{T}(\{+,*\},X)/D$ in which all equations are $*$-equations.*

Definition 2.6 (Variable Identification)
A variable identification *of $\mathcal{T}(F,X)$ is a substitution, the codomain of which is contained in the set of variables $\mathcal{X}$.*

Definition 2.7 (Well-balanced $*$-Problems) *A* well-balanced $*$-equation *in $\mathcal{T}(\{+,*\},X)/D$ is an equation $t\sigma =^?_D t\tau$, where t is a term of $\mathcal{T}(\{*\},X)$ and σ and τ are two variables identifications. A* well-balanced $*$-problem *is a unification problem in the algebra $\mathcal{T}(\{+,*\},X)/D$ in which all equations are well-balanced $*$-equations.*

One should notice that terms occurring in $*$-problems do not contain any $+$ function symbol. Moreover, in the well-balanced case, both terms of an equation are the same except for their leaves.

3 General $*$-Problems

This section is devoted to the proof of our first main result: if P is a unification $*$-problem then P is solvable modulo D if and only if it is solvable in the free algebra. Moreover all its solutions modulo D are instances of the linearized form of its most general unifier in the free algebra.

The proof of this theorem needs some technical tools, some of which are based on rewriting techniques. To record some information about the application of distributive steps, we will mark "distinct" $*$ function symbols by distinct indexes. Two occurrences of $*$ in a term will be considered identical if and only if they come from the duplication of a same $*$ symbol when using distributivity.

3.1 Indexation

Definition 3.1 (Marked term) *Let $\mathcal{X}$ be a set of variables, and $\mathcal{I}$ be a set whose members are called hereafter "indexes". A* marked term *with respect to $\mathcal{I}$ is either*
- a variable x in $\mathcal{X}$, or
- $s + t$ where s and t are marked terms with respect to $\mathcal{I}$, or
- $s *_i t$ where s and t are marked terms with respect to $\mathcal{I}$ and i is an index of $\mathcal{I}$.

Indexed distributivity, called DI in the following, is a version of distributivity which cops with marked terms. DI is presented by the indexed axioms:

$$x_I *_i (y_J + z_K) = (x_I *_i y_J) + (x_I *_i z_K) \qquad (DI_l)$$
$$(x_I + y_J) *_i z_K = (x_I *_i z_K) + (y_J *_i z_K) \qquad (DI_r)$$

where i is any index and the variables x_I, y_J and z_K can be instanciated by marked terms.

Indexed distributivity has been introduced in order to study distributivity. Hence we would like that D and DI have the "same" equivalence classes. More precisely, we mean that if t_I is a marked term, and t is the term obtained by erasing its indexes, then the class of t modulo D is exactly the class of t_I modulo DI after erasing all indexes. Unfortunately this is not always true. For example, the class modulo DI of $(x *_1 y) + (x *_2 z)$ is equal to $\{(x *_1 y) + (x *_2 z)\}$, whereas the class modulo D of $(x * y) + (x * z)$ is equal to $\{(x * y) + (x * z), x * (y + z)\}$.

Definition 3.2 (Compatible indexation) *Let t_I be a marked term and t be the unmarked term derived from t_I. t_I has a* compatible indexation *if any sequence of rewrite steps modulo D starting from t can be obtained by erasing all indexes in some sequence of rewrite steps modulo DI starting from t_I.*

Example 3.3 $(x *_1 y) + (x *_2 z)$ *has not a compatible indexation.*
$(x *_1 y) + (x *_1 z)$ *has a compatible indexation.*

Any term in $\mathcal{T}(\{+, *\}, \mathcal{X})$ can be marked in such a way that the obtained term has a compatible indexation. The obvious but useless solution is to mark all $*$ function symbols with the same index.

Definition 3.4 (Finer indexation) *Let t_I and t_J be two marked terms such that the terms obtained by erasing their indexes are equal to the same term t. t_I has a* finer indexation *than t_J if t_J is equal to t_I after applying an index identification.*

Lemma 3.5 *Let t be a term in $\mathcal{T}(\{+, *\}, \mathcal{X})$. t can be marked in such a way that the obtained term t_I has a unique minimal (finest) compatible indexation, up to renaming.*

Sketch of the proof : t has finitely many $*$ function symbols. Hence, up to renaming, t can be marked by finitely many distinct indexations. Among these indexations, there are some compatible ones, hence t has a finite non empty set of compatible indexations. This set owns at least a minimal indexation with respect to the above ordering. Two minimal indexations of t can be superimposed, yielding a finer compatible indexation. Hence the three indexations are equal up to an index renaming. □

Minimal compatible indexations are very useful in order to decide when a term can be factored out, that is, when an axiom of distributivity can be applied from right to left. Indeed when two $*$ function symbols are marked with the same index, such a factorization is possible, and we will show that there is no need to look at the subterms below them. This is important for unification since the terms that have to be unified are known, but the substitution that has to be applied in order to make them equal is *a priori* unknown.

Lemma 3.6 *Let t_I be a marked term with a compatible (respectively minimal compatible) indexation and s_J be a marked term equal modulo DI to the term t_I. Then s_J has a compatible (respectively minimal compatible) indexation.*

Lemma 3.7 *Let t_I be a marked term with a compatible indexation. Then all subterms of t_I have a compatible indexation.*

Our purpose, in this section, is to concentrate on $*$ function symbols. We define a new notion of position which takes only $*$ function symbols into account. Thanks to this notion, we can prove lemma 3.9 which is the key to proposition 3.12.

Definition 3.8 ($*$-position) *Let t_I be a marked term, and p a position of t_I. The* $*$-position *$\mathcal{D}_*(t_I, p)$ associated with p in t_I is equal to*
- Λ *if p is equal to Λ*
- $\mathcal{D}_*(t_I, q)$ *if p is equal to $q \cdot n$, where n is either 1 or 2, and $t_I(q) = +$,*
- $\mathcal{D}_*(t_I, q) \cdot (n, *_i)$ *if p is equal to $q \cdot n$, where n is either 1 or 2, and $t_I(q) = *_i$.*

Lemma 3.9 *Let t_I be a marked term with a minimal compatible indexation. For every index i occurring in t_I, the set $\mathcal{P}_*(t_I, i) = \{\mathcal{D}_*(t_I, p) \mid t_I(p) = *_i\}$ is a singleton set.*

Definition 3.10 *Let t_I be a marked term with a minimal compatible indexation, and i be an index occurring in t_I. The* $*$-position *of i in t_I is equal to $\mathcal{P}_*(t_I, i)$. The* pseudo-position *of i in t_I is the pair $(\mathcal{P}_*(t_I, i), n)$, where n is the number of distinct indexes occurring on the left hand side of the minimal position of i (that is, at a smaller position with respect to the lexicographic ordering) and which have the same $*$-position than i.*

It should be noticed that two distinct indexes have distinct pseudo-positions.

Lemma 3.11 *Let t_I and s_J be two DI-equal marked terms with a minimal compatible indexation, and let i be an index which occurs in t_I (hence in s_J). Then the pseudo-positions of i in t_I and in s_J are the same.*

Proposition 3.12 *Let t_I be a marked term with a minimal compatible indexation. Then all subterms of t_I have a minimal compatible indexation.*

Proof : This proof is quite technical. Actually, we show by induction on the pair (n_*, n_+), where n_* (respectively n_+) is equal to the number of $*$ (respectively $+$) function symbols in a developed normal form of t_I, that $t_I|_1$ and $t_I|_2$ have minimal compatible indexations. The key point is to prove minimality since lemma 3.7 ensures compatibility. □

We can now state the

Proposition 3.13 *Let t_I be a marked term with a minimal compatible indexation. If t_I contains a subterm $t_I|_p$ of the form $(a_I *_i b_I) + (c_I *_i d_I)$, then this subterm can be factored out, that is*

$$\begin{array}{lcl} (a_I *_i b_I) + (c_I *_i d_I) & =_{DI} & \alpha_J *_i (\beta_J + \gamma_J) \\ a_I *_i b_I & =_{DI} & \alpha_J *_i \beta_J \\ c_I *_i d_I & =_{DI} & \alpha_J *_i \gamma_J \\ & or & \\ (a_I *_i b_I) + (c_I *_i d_I) & =_{DI} & (\alpha_J + \beta_J) *_i \gamma_J \\ a_I *_i b_I & =_{DI} & \alpha_J *_i \gamma_J \\ c_I *_i d_I & =_{DI} & \beta_J *_i \gamma_J \end{array}$$

Proof : According to proposition 3.12, $t_I|_p$ has a minimal compatible indexation. If this term cannot be factored out, all sequences of rewrite steps modulo DI starting from $t_I|_p$ do not contain any step at position Λ. This is in contradiction with the minimality of the indexation, since the marked term $(a_I *_{i_1} b_I) + (c_I *_{i_2} d_I)$, where i_1 and i_2 are two distinct copies of i, has a finer compatible indexation. □

3.2 Solutions of $*$-problems

It is very easy to decide whether a given unification $*$-problem P has a solution modulo D: one simply has to check the existence of a solution modulo the free theory which is decidable [3, 2]. Moreover if P is satisfiable, its most general unifier in the free theory, "combined" with appropriate $AC1$ unifiers can be used to represent "almost all" solutions.

Proposition 3.14 *Let P be a $*$-problem. P has a solution modulo D if and only if it has a solution modulo the empty equational theory.*

Proof : It is obvious that a substitution which is a solution for P modulo the free theory is also a solution for P modulo D. Hence we only have to show that if there exists a substitution τ which is a solution modulo D for P, there exists a substitution σ which is a solution for P modulo the free theory. We may assume without restriction that τ is in developed normal form. Let $s =_D^? t$ be an equation in P. The two terms $s\tau$ and $t\tau$ are equal modulo D, and so are their developed normal forms. Their left-most maximal $*$-subterms are equal since the recursively defined $\mathcal{S}_l$

$$\begin{array}{lcll} \mathcal{S}_l(x) & = & x & \text{if } x \in \mathcal{X} \\ \mathcal{S}_l(t_1 + t_2) & = & \mathcal{S}_l(t_1) & \\ \mathcal{S}_l(t_1 * t_2) & = & \mathcal{S}_l(t_1) * \mathcal{S}_l(t_2) & \end{array}$$

provides obviously the left-most maximal $*$-subterm for a term which is in developed normal form, and gives the same result for all D-equal terms. Moreover $\mathcal{S}_l(s\tau)$ (respectively $\mathcal{S}_l(t\tau)$) is equal to $s\{x \mapsto \mathcal{S}_l(x\tau) \mid x \in \mathcal{D}om(\tau)\}$ (respectively $t\{x \mapsto \mathcal{S}_l(x\tau) \mid x \in \mathcal{D}om(\tau)\}$) since s (respectively t) does not contain any

$+$ function symbols. Hence the following equality holds:

$$s\{x \mapsto \mathcal{S}_l(x\tau) \mid x \in \mathcal{D}om(\tau)\} = t\{x \mapsto \mathcal{S}_l(x\tau) \mid x \in \mathcal{D}om(\tau)\}$$

The substitution $\{x \mapsto \mathcal{S}_l(x\tau) \mid x \in \mathcal{D}om(\tau)\}$ is a solution for P in the free theory. □

We now proceed with the reduction of $*$-problems to well-balanced $*$-problems.

Definition 3.15 (Linear substitution) *A substitution σ is* linear *if any variable x in its codomain occurs exactly once in the multiset $\{y\sigma \mid y \in \mathcal{D}om(\sigma)\}$.*

Lemma 3.16 *Any substitution σ can be decomposed into a linear substitution λ and a variable identification ι: $\sigma = \lambda\iota$*

When P has no solution, there is no well-balanced instance of P. The following proposition shows that solving P boils down to solving its most general well-balanced instance.

Proposition 3.17 *Let P be a solvable $*$-problem, $\sigma_0 = \lambda_0\iota_0$ its most general unifier, where λ_0 is a linear substitution and ι_0 a variable identification. Let σ be a solution of P modulo D. Then $\sigma =_D \lambda_0\rho$ for some ρ.*

Sketch of the proof : The proof is by induction on n_*, the number of $*$ function symbols of λ_0. Let σ be a solution of P, let $s =^?_D t$ be an equation of P which is not well-balanced and p be a position such that $t(p)$ is a $*$ function symbol and $s(p)$ is a variable x. $s\sigma$ can be marked with a minimal compatible indexation, yielding a minimal compatible indexation over $t\sigma$ (lemmas 3.5 and 3.6). Let us denote by i the index occurring at position p in the marked version of $t\sigma$. i does not occur anywhere else in the marked version of $t\sigma$ since all $*$-positions of $*$ function symbols are not equal to the $*$-position of i in $t\sigma$. Thanks to properties of pseudo-positions, it is then possible to show that the marked version of $(s\sigma)|_p$ is the only subterm in $s\sigma$ which contains $*_i$. Moreover $*_i$ occurs along all paths from the root to the leaves in $(s\sigma)|_p$ and it is the first encountered indexed $*$ symbol. Thanks to the factorization proposition, we then know that $(s\sigma)|_p$ is D-equal to a term $u_1 * u_2$. The problem $P\{x \mapsto x_1 * x_2\}$, where x_1 and x_2 are some fresh variables, has a solution that has less $*$ function symbols than λ_0 and the induction hypothesis applies. Since there is a one-to-one mapping from the solutions of P to the solutions of P', we can conclude that σ is D-equal to an instance of λ_0. □

4 Well-balanced $*$-Problems

Well-balanced $*$-problems always admit solutions. The most trivial one is a substitution which maps all variables on a single variable. In this case, two terms occurring in the same equation become syntactically equal. Hence, the question is not the solvability of such problems, but a full description of their solutions.

In [6], Kirchner and Klay have proved that distributivity is not syntactic [5, 6], by giving an infinite set of uncomparable solutions for the very simple equation $x * y =_D^? u * v$. We introduce a representation of the solutions which enables us to represent such an infinite set by a single scheme. Actually, we cannot represent all solutions, but only the "upper part" of each solution, that is the part without any $*$ function symbol. Such an upper part is also a solution for a well-balanced $*$-problem. The schemata we use in order to represent the solutions of a well-balanced $*$-problem can also help for the general case, see [1]. This section is devoted to the study of these schemata, and emphasizes the strong and astonishing relationship between both-sided distributivity and the theory of associativity and commutativity with a unit element.

First, we formally define the "upper part" of a term, which belongs to a particular term algebra called the algebra of structures. This notion is quite natural for a certain class of terms, but in the general case, we need some lemmas in order to prove the uniqueness of this upper part.

4.1 Upper part of a term

Definition 4.1 (Algebra of structures) $\mathcal{T}(\{+, \square\})$ *is the* algebra of structures.

Definition 4.2 (Upper part of a term in normal form) *Let t be a term in $\mathcal{T}(\{+, *\}, \mathcal{X})$ in developed normal form. The* upper part *of t is the unique maximal term (with respect to the size) u of the algebra of structures, such that there exist a term s and a substitution σ satisfying*

$$t \equiv s\sigma \qquad\qquad u \equiv s\{x \mapsto \square \mid x \in \mathcal{V}ar(s)\}$$

The upper part of t is denoted by $\widehat{t}$.

Remark : $\{x \mapsto \square \mid x \in \mathcal{V}ar(s)\}$ is not exactly a substitution since it does not define an endomorphism. This "pseudo-substitution" defines an homomorphism from $\mathcal{T}(\{+, *\}, \mathcal{X})$ to $\mathcal{T}(\{+, \square\})$.

The formal definition is quite complicated, but the computation of an upper part is rather simple: since t is in developed normal form, all $+$ function symbols are above the $*$ symbols. Hence, the upper part of a term in developed normal form can be obtained by replacing all $*$-headed terms and variables by the $\square$ symbol.

Example 4.3 *Consider the term $(x + (x * z)) + (x * y)$. Its upper part is equal to $(\square + \square) + \square$.*

When orienting the two axioms of distributivity as in definition 2.2, a critical pair appears between the obtained rules. Hence a term can have several developed normal forms. If we extend our notion of upper part to terms which are not in developed normal form, we cannot ensure the uniqueness property. This problem is solved by considering classes of structures modulo an equivalence relation $\sim_D$ derived from the equational theory D.

Definition 4.4 (D-equivalence) *Two structures s and t are D-equivalent if*

$$s\{\Box \mapsto x * y\} =_D t\{\Box \mapsto x * y\}$$

This is denoted by $s \sim_D t$.

An important remark is that if t_1 and t_2 are in developed normal form, there is a normal from of $t_1 * t_2$ the upper part of which is equal to $\widehat{t_1}\{\Box \mapsto \widehat{t_2}\}$. This normal form is obtained by applying D_r as far as possible and then D_l.

Hence we introduce the following:

Definition 4.5 (Composition of structures) *Let t_1 and t_2 be two structures. The composition of t_1 and t_2 is the structure denoted $t_1@t_2$ obtained by replacing all leaves of t_1 by t_2: $t_1@t_2 = t_1\{\Box \mapsto t_2\}$*

Lemma 4.6 *$\mathcal{T}(\{+, \Box\})$ and $\mathcal{T}(\{+, \Box\})/\sim_D$ are semi-groups with respect to composition.*

This lemma shows a relationship between D and $A1$, but it is actually a relationship between D and $AC1$.

Lemma 4.7 *The composition of structures is commutative in $\mathcal{T}(\{+, \Box\})/\sim_D$.*

Proof : Let s and t two structures, we will show that $s@t$ and $t@s$ are D-equivalent. By applying D_r as far as possible to all subterms $t\{\Box \mapsto x * y\}$ occurring in $s@t\{\Box \mapsto x*y\}$, one gets $s\{\Box \mapsto (t\{\Box \mapsto x\})*y\}$. Then apply D_l to the whole term as far as possible, in order to obtain $(s\{\Box \mapsto x\}) * (t\{\Box \mapsto y\})$. A developed normal form of this term can now be computed by applying first D_r as far as possible and then D_l as far as possible. This normal form is equal to $t@s\{\Box \mapsto x * y\}$. $\Box$

Corollary 4.8 *$\mathcal{T}(\{+, \Box\})/\sim_D$ is a commutative semi-group with respect to @.*

Remark : Sometimes, we will use the notation $\underset{i \in I}{@} s_i$ instead of $s_{i_1}@\ldots@s_{i_n}$ if $I = \{i_1, \ldots, i_n\}$. The composition has these key properties: it is assocative, has a unit element equal to $\Box$, and is compatible with D-equivalence.

4.2 Decomposition of structures

The quotient algebra $\mathcal{T}(\{+, \Box\})/\sim_D$ has a unique decomposition property, similar to the decomposition property of natural numbers. This does not imply, however, that it is a factorial ring.

Definition 4.9 (Decomposition of structures) *Let s be a structure in $\mathcal{T}(\{+, \Box\})$. $s_1@s_2$ is a* decomposition *of s if s and $s_1@s_2$ are D-equivalent. If s_1 (or s_2) is equal to the empty structure $\Box$, the decomposition is* trivial. *A non empty structure is* prime *if all its decompositions are trivial. A decomposition is* maximal *if it is reduced to $\Box$ or if all its members are prime structures.*

The following property means that $\mathcal{T}(\{+,\Box\})/\sim_D$ is actually the free commutative semi-group generated by the set of "letters" identified as the equivalence classes of the prime structures.

Proposition 4.10 *Let r be a structure. r has exactly one maximal decomposition up to permutation.*

Sketch of the proof : Let $s_1@\ldots@s_k$ and $t_1@\ldots@t_l$ be two maximal decompositions of r. These two structures are D-equivalent, hence there is a sequence of rewrite steps modulo D between the two terms $s \equiv (s_1@\ldots@s_k)\{\Box \mapsto x * y\}$ and $t \equiv (t_1@\ldots@t_l)\{\Box \mapsto x * y\}$. We can always choose a sequence such as the one below:

$$s \overset{(\neq\Lambda)^*}{\underset{D}{\longleftrightarrow}} w \overset{\Lambda}{\underset{D}{\longleftarrow}} (u\{\Box \mapsto x\}) * (v\{\Box \mapsto y\}) \overset{*}{\underset{D}{\longleftrightarrow}} (u'\{\Box \mapsto x\}) * (v'\{\Box \mapsto y\}) \overset{\Lambda}{\underset{D}{\longrightarrow}} w' \overset{(\neq\Lambda)^*}{\underset{D}{\longleftrightarrow}} t$$

The proof of the proposition is by induction on the pair made of the number of $+$ function symbols occurring in r and the number of rewrite steps at position Λ in the middle part of the sequence

$(u\{\Box \mapsto x\}) * (v\{\Box \mapsto y\}) \overset{*}{\underset{D}{\longleftrightarrow}} (u'\{\Box \mapsto x\}) * (v'\{\Box \mapsto y\})$ $\Box$

4.3 Top of a term

Proposition 4.11 *Let s and t be two D-equal terms in $\mathcal{T}(\{+,*\},\mathcal{X})$, and s' and t' some developed normal forms of s and t. Then the upper parts of s' and t' are D-equivalent.*

Sketch of the proof : The proof is by induction on the triple $(n_+(s'), n_*(s'), n_\Lambda(s',t'))$, where $n_+(s')$ (respectively $n_*(s')$) is the number of $+$ (respectively $*$) function symbols in s' and $n_\Lambda(s',t')$ is the number of rewrite Λ-steps modulo D in a sequence between s' and t'. $\Box$

The intuitive meaning of this result is that every proof modulo distributivity between two terms in developed normal form can be performed by using only the top-most $*$-symbols. As a consequence, two distinct developed normal forms of a term t have the same upper part modulo $\sim_D$.

Definition 4.12 (Top of a term) *Let t a term of $\mathcal{T}(\{+,*\},\mathcal{X})$, and t' one of its developed normal forms. The* top *of t is the class of the upper part of t' modulo D-equivalence, denoted by $\widehat{t}$.*

In the following, for short, we will identify the top which is a set of structures with one (not anyone) of its members.

4.4 Solutions of well-balanced $*$-problems

All previous definitions of upper part and top extend naturally to substitutions.

Lemma 4.13 *Let P be a well-balanced $*$-problem without equations between variables and σ b a solution of P. Then the top of σ (replacing $\Box$ by any variable in $\mathcal{X}$) is a solution of P.*

Proof : Let $t\iota =_D^? t\iota'$ be an equation of P, where t is a linear term and ι and ι' are two variable identifications. Since σ is a solution of P, $t\iota\sigma =_D t\iota'\sigma$ holds. Hence, according to proposition 4.11, $t\iota\sigma$ and $t\iota'\sigma$ have the same top:

$$\mathop{@}_{x\in\mathcal{V}ar(t)} \widehat{x\iota\sigma} \sim_D \widehat{t\iota\sigma} \sim_D \widehat{t\iota'\sigma} \sim_D \mathop{@}_{x\in\mathcal{V}ar(t)} \widehat{x\iota'\sigma}$$

P does not contain any equation between two variables, hence t is an instance of $x * y$. By definition of D-equivalence,

$$(\mathop{@}_{x\in\mathcal{V}ar(t)} \widehat{x\iota\sigma})\{\Box \mapsto t\}\{x \mapsto \Box \mid x \in \mathcal{V}ar(t)\} \underset{D}{=} (\mathop{@}_{x\in\mathcal{V}ar(t)} \widehat{x\iota'\sigma})\{\Box \mapsto t\}\{x \mapsto \Box \mid x \in \mathcal{V}ar(t)\}$$

This equality can be rephrased as $t\iota\widehat{\sigma} =_D t\iota'\widehat{\sigma}$. $\Box$

Corollary 4.14 *Let P be a well-balanced $*$-problem, σ a solution of P and $\widehat{\sigma}$ the top of σ. Then $\widehat{\sigma}\{\Box \mapsto x * y\}$ is a solution of P.*

Thanks to the property of uniqueness for the maximal decomposition of a structure, we can describe all tops of the solutions of a well-balanced $*$-problem as instances of the most general unifier of an $AC1$ problem. $AC1$ is unitary, and the most general unifier of a problem can be easily computed by solving systems of linear Diophantine equations [4].

Proposition 4.15 *Let P be a well-balanced $*$-problem, σ a solution of P and ρ_{AC1} the most general solution of P modulo the equational theory $AC1$ of $*$. Then there exists a semi-group homomorphism h from $\mathcal{T}(\{*\},\mathcal{X})/AC1(*)$ to $\mathcal{T}(\{+,\Box\})/\sim_D$ such that*

$$\widehat{\sigma} \equiv \rho_{AC1}h$$

Conversely, let h' be any semi-group homomorphism from $\mathcal{T}(\{*\},\mathcal{X})/AC1(*)$ to $\mathcal{T}(\{+,\Box\})/\sim_D$. Then $\rho_{AC1}h'\{\Box \mapsto x * y\}$ is a solution of P.

Proof : First, we prove that $\widehat{\sigma} \equiv \rho_{AC1}h$. Consider an equation of P, $t\iota =_D^? t\iota'$, where t is a linear term. σ is a solution of P, hence $t\iota\sigma =_D t\iota'\sigma$ holds. These two terms have the same top. On the other hand, we have two decompositions of this top, $@_{x\in\mathcal{V}ar(t)}(x\iota\widehat{\sigma})$ and $@_{x\in\mathcal{V}ar(t)}(x\iota'\widehat{\sigma})$. We can derive two maximal decompositions by decomposing the $x\iota\widehat{\sigma}$ and $x\iota'\widehat{\sigma}$ as far as possible. There are finitely many distinct (up to D-equivalence) prime structures occurring in the decompositions: $s_1, s_2, \ldots, s_n$. Let us associate a variable x_i with s_i for $1 \le i \le n$. For each variable y of P we can build an arbitrary term u_y in $\mathcal{T}(\{*\},\mathcal{X})$ whose multiset of variables is equal to the multiset of variables attached to the prime structures occurring in the maximal decomposition of $y\widehat{\sigma}$. According to proposition 4.10, the maximal decompositions derived from $@_{x\in\mathcal{V}ar(t)}(x\iota\widehat{\sigma})$ and $@_{x\in\mathcal{V}ar(t)}(x\iota'\widehat{\sigma})$ are equivalent up to permutation, that is, the multisets of their prime structures are the same. Hence the multisets of variables of $t\iota\{x\iota \mapsto u_{x\iota} \mid x \in \mathcal{V}ar(t)\}$ and $t\iota'\{x\iota' \mapsto u_{x\iota'} \mid x \in \mathcal{V}ar(t)\}$ are the same and these two terms are equal modulo the equational theory $AC1$.

Consider the substitution τ defined in $\mathcal{T}(\{*\}, \mathcal{X})$ by $\tau \equiv \{y \mapsto u_y \mid y \in \mathcal{V}ar(P)\}$. $t\iota\tau =_{AC1} t\iota'\tau$ holds for all equations $t\iota =^?_D t\iota'$ of P. Hence τ is a solution of P modulo $AC1$. Since $AC1$ is unitary when there is no constant, τ is an instance of the most general solution of P, $\tau \equiv \rho_{AC1}\tau'$. Let g be the semi-group homomorphism from $\mathcal{T}(\{*, 1\}, \mathcal{X})/AC1$ to $\mathcal{T}(\{+, \square\})/\sim_D$ such that $x_i g = s_i$ for all i, $1 \leq i \leq n$, and $yg = \square$ if $y \notin \{x_1, \ldots, x_n\}$. $\widehat{\sigma}$ is equal to $\underbrace{(\rho_{AC1}\tau')}_{\tau} g \equiv \rho_{AC1} \underbrace{(\tau' g)}_{h}$.

Now, we prove that $\rho_{AC1}h'\{\square \mapsto x * y\}$ is a solution of P for all semi-group homomorphisms h' from $\mathcal{T}(\{*, 1\}, \mathcal{X})/AC1$ to $\mathcal{T}(\{+, \square\})/\sim_D$. Let $t\iota =^?_D t\iota'$ be an equation of P. ρ_{AC1} is the minimal solution of P modulo the equational theory $AC1$, hence the two terms $t\iota\rho_{AC1}$ and $t\iota'\rho_{AC1}$ are equal modulo $AC1$. Since h' is a semi-group homomorphism, $(t\iota\rho_{AC1})h'$ and $(t\iota'\rho_{AC1})h'$ are D-equivalent and by definition $(t\iota\rho_{AC1})h'\{\square \mapsto x * y\} =_D (t\iota'\rho_{AC1})h'\{\square \mapsto x * y\}$. Here, it should be noticed that h' maps the whole terms $t\iota\rho_{AC1}$ and $t\iota'\rho_{AC1}$ to structures, and that all $*$ function symbols of t and ρ_{AC1} become @. By $(\rho_{AC1}h'\{\square \mapsto x * y\})$, we mean a substitution in $\mathcal{T}(\{+, *\}, \mathcal{X})$, such that only the $*$ function symbols of ρ_{AC1} are replaced by @, hence the following equalities hold:

$$t\iota(\rho_{AC1}h'\{\square \mapsto x * y\}) =_D (t\iota\rho_{AC1})h'\{\square \mapsto (t\{u \mapsto x * y \mid u \in \mathcal{V}ar(t)\})\}$$
$$t\iota'(\rho_{AC1}h'\{\square \mapsto x * y\}) =_D (t\iota'\rho_{AC1})h'\{\square \mapsto (t\{u \mapsto x * y \mid u \in \mathcal{V}ar(t)\})\}$$

since t contains only $*$ function symbols, $(t\{u \mapsto x * y \mid u \in \mathcal{V}ar(t)\})$ is an instance of $x * y$, hence $t\iota(\rho_{AC1}h'\{\square \mapsto x * y\}) =_D t\iota'(\rho_{AC1}h'\{\square \mapsto x * y\})$. $\square$

There are infinitely many semi-group homomorphisms from $\mathcal{T}(\{*\}, \mathcal{X})/AC1(*)$ to $\mathcal{T}(\{+, \square\})/\sim_D$, but ρ_{AC1} is unique and can be used to schematize all tops of solutions of P.

By putting together proposition 4.15 and proposition 3.17, we get a description of the solutions of any $*$-problem P.

Corollary 4.16 *Let P be a solvable $*$-problem, λ_0 the linearization of its most general unifier in the free theory, and ρ_{AC1} the most general unifier of $P\lambda_0$ modulo $AC1$. Let σ be a solution of P modulo D. There exist a substitution ρ of $\mathcal{T}(\{+, *\}, \mathcal{X})$ and a semi-group homomorphism h from $\mathcal{T}(\{*\}, \mathcal{X})/AC1(*)$ to $\mathcal{T}(\{+, \square\})/\sim_D$ such that $\sigma =_D \lambda_0\rho$ and $\widehat{\rho} = \rho_{AC1}h$.*

Conversely, let h' be a semi-group homomorphism from $\mathcal{T}(\{*\}, \mathcal{X})/AC1(*)$ to $\mathcal{T}(\{+, \square\})/\sim_D$. Then the substitution $\lambda_0(\rho_{AC1}h'\{\square \mapsto x * y\})$ is a solution of P.

5 Conclusion

We have introduced new tools (indexed distributivity and the algebra of structures) which enable us to check very easily the solvability modulo D of a

$*$-problem and to represent its infinite set of solutions with a single scheme. The most surprising result is that D-unification of $*$-problems boils down to $AC1$-unification. The question of course, is whether this can be generalized to arbitrary problems. We believe it can, and based on the previous tools, we have solved the general problem under the technical assumption that there exists no cycle of some special form called "cycles of non-null weight" [1]. We therefore conjecture that unification modulo distributivity is decidable, but the algorithm appears to be at least as complex as Makanin's algorithm [8]. Moreover, it is interesting to notice that both of them use the solutions of Diophantine equations, for solving positions equations in Makanin's algorithm, and for computing $AC1$-solutions in our algorithm.

References

1. Evelyne Contejean. *Eléments pour la Décidabilité de l'Unification modulo la Distributivité*. Thèse de doctorat, Université Paris-Sud, Orsay, France, April 1992.
2. Nachum Dershowitz and Jean-Pierre Jouannaud. Rewrite systems. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 243–309. North-Holland, 1990.
3. Jacques Herbrand. Recherches sur la théorie de la démonstration. Thèse d'Etat, Univ. Paris, 1930. Also in: Ecrits logiques de Jacques Herbrand, PUF, Paris, 1968.
4. Alexander Herold and Jörg H. Siekmann. Unification in abelian semi-groups. *Journal of Automated Reasoning*, 3(3):247–283, 1987.
5. Claude Kirchner. Méthodes et outils de conception systématique d'algorithmes d'unification dans les théories equationnelles. Thèse d'Etat, Univ. Nancy, France, 1985.
6. Claude Kirchner and Francis Klay. Syntactic theories and unification. In *Proc. 5th IEEE Symp. Logic in Computer Science, Philadelphia*, June 1990.
7. Donald E. Knuth and Peter B. Bendix. Simple word problems in universal algebras. In J. Leech, editor, *Computational Problems in Abstract Algebra*, pages 263–297. Pergamon Press, 1970.
8. G.S. Makanin. The problem of solvability of equations in a free semigroup. *Akad. Nauk. SSSR*, 233(2), 1977.
9. J. V. Matijasevic. Enumerable sets are diophantine. *Soviet Mathematics (Doklady)*, 11(2):354–357, 1970.
10. J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12(1):23–41, 1965.
11. P. Szabo. Unifikationstheorie erster ordnung. Technical Report Thesis, Fakultat fur Informatik, University Karlsruhe, Karlsruhe, West Germany, 1982.
12. Erik Tiden and Stefan Arnborg. Unification problems with one-sided distributivity. *Journal of Symbolic Computation*, 3:183–202, 1987.

This article was processed using the LaTeX macro package with LLNCS style