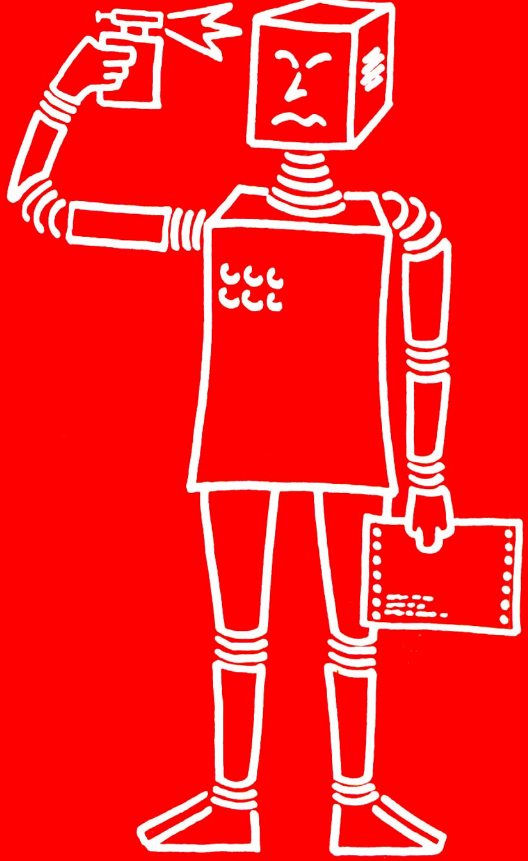
SEKI-PROJEKT

SEKI MEMO

Universität Kaiserslautern
Fachbereich Informatik
Postfach 3049
D-6750 Kaiserslautern
West Germany

Universität Karlsruhe
Institut für Informatik I
Postfach 6380
D-7500 Karlsruhe
West Germany



UNIFICATION IN THE DATASTRUCTURE MULTISETS

Wolfram Büttner

MEMO SEKI-85-V-KL

Unification in the Datastructure Multisets

Wolfram BÜttner
Siemens AG
Corporate Laboratories for
Information Technology

D-8000 Munich 83

Acknowledgement

I have begun and completed this work during a stay as a visiting researcher at the "Sonderforschungsbereich Künstliche Intelligenz" at Kaiserslautern University during the springterm 1985.

For suggestions and comments I am particularly grateful to Th. Guggenbiehl, A. Herold and J. Siekmann. I further gratefully acknowledge Prof. Dr. J. Siekmann from the University of Kaiserslautern and Dr. H. Schwärtzel from the Siemens Corporate Laboratories for Information Technology for making this stay possible.

Unification in the Datastructure Multisets

Abstract:

In a forthcoming paper A. Herold and J. Siekmann generalize "pure" AC-unification ([ST 75], [LS 76]) to terms containing additional function symbols (see also [ST 81], [FA 84]). Generalized AC-unification thus attains practical relevance for a broad range of applications. Pure AC-unification is used as a basic mechanism and it is this key role that has motivated our research.

We have improved upon earlier approaches by basing (pure) AC-unification on a firm theoretical basis and presenting algorithms which fully exploit the properties of the underlying mathematical structure.

In particular, the high degree of parallelism for AC-unification will become apparent. Our algorithms have been designed for parallel hardware but still yield significant improvements over earlier algorithms when used in sequential mode.

Introduction

The unification of terms built from variables, constants and an associative and commutative function symbol (AC-unification), has been treated by Stickel [ST 74], [ST 75], Livesey and Siekmann [LS 76], [SI 78] in the year 1975, using different approaches.

Algorithmic improvements as well as theoretical completions have been reported in a number of articles [FT 83], [FT 85], [HS 85], [HU 79], [ST 81]. The theory has only recently been extended to terms containing additional function symbols [FA 84], [HS 85], [ST 81].

The notion of AC-unification and unification of multisets refer to equivalent structures. For in a free term algebra built from variables, constants and a two-place function symbol we consider the following congruence relation: Terms which differ in a number of associative and commutative manipulations, are collected in a congruence class. The multiset formed by the leaves of a term can be identified with the congruence class containing this term.

Apart from the well established Robinson unification [RO 65] AC-unification is presently certainly the most important special unification algorithm. Some of its applications are automatic theorem proving, rewrite systems, program verification, the theory of abstract data types and logic programming. Using AC-unification algorithms it is possible to extend PROLOG to statements having the form of relations between multisets. Using an explicit extraction of the mathematical structures upon which AC-unification is based, we shall provide a compact and solid theoretical foundation for the unification process. This theoretical framework will prove to be useful for the design of efficient algorithms.

In the same spirit - mutatis mutandis - other unification problems can be treated. In order to prove correctness and completeness of AC-unification we need to argue in a free commutative semigroup H. Arguing in this semigroup is sometimes rather clumsy. The reason is basically the same as the reason that the equation $y + 1 = 2$ has to be solved more cautiously over $\mathbb{N}$ than over $\mathbb{Z}$. This trivial example should motivate our approach: we embed H in a rational vectorspace V and think of substitutions not as endomorphisms of H, but as vectorspace endmorphisms of V. Now we can use the machinery of linear algebra. Since the unification of two terms is a local problem we confine ourselves to finite dimensional vectorspaces. Here linear mappings and in particular substitutions, can be described by matrices. The choice of a basis which is particularly appropriate for a given linear mapping is "visually" reflected by the form of the corresponding

matrix.

A critical comparison of the two approaches cited above, can be based upon this interpretation of unifiers as specific matrices. Furthermore without any difficulty we recognize the equivalence between the AC-unification problem and the following numbertheoretical problem: Decide whether certain homogenous and nonhomogeneous linear diophantine equations can be solved over $\mathbb{N}$ and determine "minimal" solution sets.

Linear algebra is particularly well suited for a proper discussion of the solutions of these equations. The solution sets we try to find, can be interpreted geometrically, which will be helpful in the design of fast algorithms.
In this paper we present a parallel algorithm which searches some finite dimensional grid. Improved sequential search strategies can easily be incorporated. Measurements which try to determine the computational effects of these algorithms are in progress. Additionally we discuss the decision problem for AC-unification. Two terms containing only variables can always be unified. In case both terms contain variables and (possibly) constants the decision problem can be solved computing the gcd of integers which depend on the given terms. Very efficient algorithms are available to handle this computation [Br 70]. In case one of the terms is variable-free the decision problem is NP-complete [KNB 85].

The terminology we employ is the usual. Apart from some basic knowledge of the rudiments of universal algebra the mathematical prerequisites are restricted to linear algebra.

1. Free abelian monoids, Q-vectorspaces and endomorphisms of these structures

For countable sets X (variables) and C (constants) with $X \cap C = \emptyset$ we consider the free abelian monoid $\mathfrak{F}$ generated by $X \cup C$. We denote the semigroup multiplication by +, and abbreviate $(\dots (f_1 + f_2) + f_3) \dots + f_n)$ $(f_i \in \mathfrak{F})$ by $\sum_{i=1}^{n} f_i$. O denotes the identity of $\mathfrak{F}$ which is assumed to be not an element of C. The theory [GR 79] ensures the existence of $\mathfrak{F}$ and shows that $\mathfrak{F}$ is determined up to isomorphisms by $X \cup C$. A typical element $f \in \mathfrak{F}$ has the form $f = f_x + f_c$ with $f_x = \sum_{i=1}^{n} a_i x_i$, $f_c = \sum_{j=1}^{m} n_j c_j$ and $x_i \in X$, $c_j \in C$, $a_i, n_j \in \mathbb{N}_0$. The sums f_x, f_c generate (free) submonoids $\mathfrak{F}_x$, $\mathfrak{F}_c$ of $\mathfrak{F}$ and $\mathfrak{F}$ is the direct sum of these submonoids.

A familiar example of such structures is the semigroup $\mathbb{N}_0 \times \mathbb{N}_0$. The generating set $X \cup C$ corresponds here to the set $\{ (1,0), (0,1) \}$.
$\mathbb{N}_0 \times \mathbb{N}_0$ is most naturally embedded into the rational vectorspace $\mathbb{Q} \times \mathbb{Q}$, consisting of the elements of the canonical grid in $\mathbb{Q}^2$ with nonnegative integers as components.

Analogously we may embed $\mathfrak{F}$ into a rational vectorspace $\mathcal{V}$.
$\mathcal{V}$ consists of all finite sums $\sum_{i=1}^{n} q_i x_i + \sum_{j=1}^{m} p_j c_j$ with $x_i \in X$, $c_j \in C$, $q_i, p_j \in \mathbb{Q}$.
Such sums are added and multiplied by rationals in the obvious way.
A basis of $\mathcal{V}$ is given by an ordering of the (countable) set $X \cup C$.
To the decomposition $\mathfrak{F} = \mathfrak{F}_x + \mathfrak{F}_c$ corresponds a decomposition $\mathcal{V} = \mathcal{V}_x + \mathcal{V}_c$, $\mathcal{V}_x$, $\mathcal{V}_c$ being the subspaces of $\mathcal{V}$ generated by X,C respectively.
Since $\mathfrak{F}$ is free, any mapping from $X \cup C$ to $\mathfrak{F}$ can be uniquely extended to a semigroup endomorphisms of $\mathfrak{F}$ and conversely any semigroup endomorphism of $\mathfrak{F}$ is determined by its effect on $X \cup C$.

1.1 Definition: A substitution σ of $\mathfrak{F}$ is an endomorphism of $\mathfrak{F}$ which fixes C pointwise and moves only finitly many elements of X.

$X \cup C$ (more precisely, an ordering of $X \cup C$) is a basis of $\mathcal{V}$; any mapping from $X \cup C$ to $\mathcal{V}$ can be extended uniquely to a vectorspace endomorphism of $\mathcal{V}$. This applies in particular to any substitution σ and yields an endomorphism $\hat{\sigma}$ of $\mathcal{V}$. Conversely, any vectorspace endomorphism of $\mathcal{V}$ which acts trivially on C and almost all elements of X and maps the remaining elements of X onto $\mathbb{N}_0$-linear combinations of variables and constants, is the extension of some substitution.

2. The AC - Unification Problem

2.1 Definition: For elements f_1, f_2 of $\mathcal{F}$ a unifier of $\{f_1, f_2\}$ is a substitution σ with $\sigma(f_1) = \sigma(f_2)$.

We can compare unifiers σ, τ of $\{f_1, f_2\}$ by setting $\sigma \leq \tau$ iff there exists a substitution λ such that equality $\sigma = \lambda\tau$ holds for all variables occuring in f_1 and f_2. On $U\Sigma(\{f_1, f_2\})$, the set of all unifiers of $\{f_1, f_2\}$, we have an equivalence relation $\approx$ given by $\sigma \approx \tau$ iff $\sigma \leq \tau$ and $\tau \leq \sigma$. The relation $\leq$ induces a partial ordering in $U\Sigma(\{f_1, f_2\})/\approx$ and here we can talk about maximal elements.

2.2 Definition:

i) A most general unifier σ of $\{f_1, f_2\}$ is a unifier, which induces in $U\Sigma(\{f_1, f_2\})/\approx$ a maximal element.

ii) $\mu U\Sigma(\{f_1, f_2\})$ is a minimal set of unifiers of $\{f_1, f_2\}$, which induces in $U\Sigma(\{f_1, f_2\})/\approx$ all maximal elements.

Note that the axiomatic definition of $\mu U\Sigma(\{f_1, f_2\})$ as in [PL 72], [SI 84] produces the same set of most general unifiers. $\mu U\Sigma(\{f_1, f_2\})$ is uniquely determined up to equivalence.

Now the AC-unification problem poses three questions:

A: Given elements f_1, f_2 in $\mathcal{F}$, can the existence of a unifier of $\{f_1, f_2\}$ be decided?

B: Can the cardinality of $\mu U\Sigma(\{f_1, f_2\})$ be calculated?

C: Can $\mu U\Sigma(\{f_1, f_2\})$ be computed?

A positive answer to these questions has (among others) the following implication:

A theorem prover, which works on terms, containing a two-place associative and commutative functionsymbol, is refutation complete [PL 72].

3. Problem Reduction

The following manipulations of f_1, f_2 do not change $\mu U\Sigma$. However they simplify the problem, hence in the sequel we will always assume that these simplifications have been done already.

Let $f_1 = \sum_{i=1}^{n} a_i x_i + \sum_{j=1}^{m} n_j c_j$, $f_2 = \sum_{k=1}^{s} b_k y_k + \sum_{e=1}^{w} m_e d_e$ be elements of $\mathfrak{F}$.

We shall consider three types of problem reductions:

a) We compute $d_x := \gcd \{ a_i, b_k \mid 1 \le i \le n, \ 1 \le k \le s \}$ for $d_x = 1$, the elements f_1, f_2 remain unchanged. For $d_x \neq 1$ we compute $d_c := \gcd \{ n_j, m_e \mid 1 \le j \le m, \ 1 \le e \le w \}$, $d := \gcd \{ d_x, d_c \}$ and replace f_i by $1/d \ f_i \ (1 \le i \le 2)$.

Of course σ is a unifier of $\{ f_1, f_2 \}$ if and only if σ unifies $\{ 1/d \ f_1, \ 1/d \ f_2 \}$.

(Note: In order to solve problem A of the AC-unification problems we have to compute d_x. Hence this work has to be done anyway.)

A reduction like in a) restricts searchspaces to be defined lateron. Passing from f_i to $1/d \ f_i$ is a trivial operation in the vectorspace $\mathfrak{V}$ but needs arguments if done in the semigroup $\mathfrak{F}$.

b) Assume that the decomposition of f_1, f_2 respectively contains variables x_{i0} and y_{k0} with $x_{i0} - y_{k0}$. Then we substract a maximal multiple rx_{i0} from f_1 and f_2 such that $f_1 - rx_{i0}$, $f_2 - rx_{i0}$ still do not have negative coefficients. Treating constants occuring in f_1, f_2, the same way we may assume f_1, f_2 to have no common factors.

c) Without loss of generality we may assume that for a unifier σ of $\{ f_1, f_2 \}$ the term $\sigma (f_1)$ contains only constants from $\{ c_j, d_e \mid 1 \le j \le m, 1 \le e \le w \}$ and σ is the identity on $X \setminus \{ x_i, y_k \mid 1 \le i \le n, 1 \le k \le s \}$.

For otherwise there exists some x_{i0} $(1 \le i_0 \le n)$ such that $\sigma(x_{i0})$ contains a constant c from $C \setminus \{c_j, d_e \mid 1 \le j \le m, 1 \le e \le w\}$. Now consider the unifier σ'

$$\text{with } \sigma'(x) = \begin{cases} \sigma(x) \text{ for } x \neq x_{io} \\ t \quad \text{for } x = x_{io} \end{cases}$$

where t is the element of $\mathcal{F}$ which is derived from $\sigma(x_{io})$ by replacing all occurences of c in $\sigma(x_{io})$ by a variable u (u being a variable not contained in f_1).

Let λ be the substitution

$$\text{with } \lambda(x) = \begin{cases} x \quad \text{for } x \neq u \\ c \quad \text{for } x = u \end{cases}$$

We have $\sigma = \lambda\,\sigma'$ and therefore we may replace σ in $\mu U\Sigma$ ({ f_1, f_2 }) by σ'. Similarly we can justify the second assumption about σ.

4. Localization and Transformation of the AC - Unification Problem into Linear Algebra

In this chapter we shall associate with any unifier σ of { f_1, f_2 } - normed according to 3 c) - a matrix R(σ). σ and R(σ) mutually determine each other and all information needed for the solution of the AC - unification problem can readily be obtained from R(σ).

Let σ be a unifier of { f_1, f_2 } and assume f_1, f_2 to be given by some linear combination of variables and constants as in 3. We define a set W as follows:
If $\sigma(f_1)$ does not contain variables, we set $W = \emptyset$. Otherwise let $W := \{ u_1 ,..., u_r \}$ be a finite subset of X containing all variables occuring in $\sigma(f_1)$ (but possibly more variables). Let $A := \{ a_1 ,..., a_t \}$ be the constants occuring in f_1 and f_2 and let $\{ x_1 ,..., x_n \}$, $\{ y_1 ,..., y_s \}$ be the variables of f_1, f_2 respectively.
$\mathcal{V}_e$ denotes the (finitedimensional) subspace of $\mathcal{V}$ with basis $\{ x_1 ,..., x_n , y_1 ,..., y_s , a_1 ,..., a_t \}$ and $\mathcal{U}$ is the subspace of $\mathcal{V}$ with basis

$\{ u_1 ,\ldots, u_r , a_1 ,\ldots, a_t \}$. $\hat{\sigma}(\mathcal{V}_e) \subseteq \mathcal{U}$ and $\hat{\sigma}$ acts (see 3.c) trivially on $X \setminus X \cap \mathcal{V}_e$. Hence $\hat{\sigma}|_{\mathcal{V}_e} : \mathcal{V}_e \to \mathcal{U}$ determines $\hat{\sigma}$ (and σ) completely.

4.1 Remark:

According to its definition W may contain variables not occuring in $\sigma(f_1)$. Hence we may always assume that for unifiers σ_1 , σ_2 of $\{f_1, f_2\}$, $\sigma_1 | \mathcal{V}_e$, $\sigma_2 | \mathcal{V}_e$ are both mappings from $\mathcal{V}_e$ to $\mathcal{U}$.

With respect to the given basis of $\mathcal{V}_e$ and $\mathcal{U}$ the linear map $\sigma | \mathcal{V}_e$ is described by some matrix $R(\sigma)$ with nonnegative integers as entries. We recall:
the image of the i - th basis vector of $\mathcal{V}_e$ is a linear combination of the basis of $\mathcal{U}$. The j - th coefficient of this linear combination defines the entry at position (i , j) of $R(\sigma)$. Therefore we have

4.2 Lemma:

With respect to the given basis of $\mathcal{V}_e$, $\mathcal{U}$ respectively the matrix $R(\sigma)$ has the form

$$R(\sigma) = \begin{bmatrix} R_x(\sigma) & 0 \\ \underbrace{R_c(\sigma)}_{n+s} & \underbrace{I}_{t} \end{bmatrix} \begin{matrix} \}\,r \\ \}\,t \end{matrix}$$

with submatrices $R_x(\sigma)$, $R_c(\sigma)$ of the indicated size. 0 , I denote suitable all zero - and identity matrices.

Note: All matrices used in the following have only nonnegative integers as entries.

The matrices $R_x(\sigma)$, $R_c(\sigma)$ describe special AC - unification problems.

a) By definition a substitution τ fixes C pointwise. Hence τ induces an endomorphism $\overline{\tau}$ on $\overline{\mathcal{V}} := \mathcal{V}/\mathcal{V}_c \simeq \mathcal{V}_x$. $\{f_1, f_2\}$ induces in $\mathcal{V}_x$ (identified with $\mathcal{V}$) an AC - unification problem $\{\overline{f_1}, \overline{f_2}\}$. We have

$\overline{\mu U\Sigma(\{f_1, f_2\})} \subseteq \mu U\Sigma(\{\overline{f}_1, \overline{f}_2\})$.
If σ has been associated the matrix $R(\sigma)$, $\overline{\sigma}$ will be associated the matrix $R_X(\sigma)$.

b) Let $\sigma'(x_i)$, $\sigma'(y_j)$ $(1 \leq i \leq n, 1 \leq j \leq s)$ be the terms constructed from $\sigma(x_i)$, $\sigma(y_j)$ by eliminating all variables. The substitution σ' defined by

$$\sigma'(x) = \begin{cases} \sigma'(x_i) & \text{for } x = x_i \\ \sigma'(y_j) & \text{for } x = y_j \\ x & \text{otherwise} \end{cases}$$

unifies $\{f_1, f_2\}$ and will be associated to the matrix $R_c(\sigma)$.

Hence σ is characterized by two special unifiers. As we shall see later on these unifiers cannot be chosen independently.

By definition f_1, f_2 and
$f_1 - f_2 = (a_1, \ldots, a_n, -b_1, \ldots, -b_s, n_1, \ldots, n_m, -m_1, \ldots, -m_w)^T$ (*) are vectors in $\mathcal{V}_e$. We set $a := (a_1, \ldots, a_n, -b_1, \ldots, -b_s)^T$ and formulate the following useful lemma:

4.3 Lemma:

The following statements are equivalent:

i) $\sigma(f_1) = \sigma(f_2)$ i.e. $\sigma \in \mu U\Sigma(\{f_1, f_2\})$

ii) $f_1 - f_2 \in \text{Kern } \hat{\sigma} | \mathcal{V}_e$

iii) The rows of the matrix $R_x(\sigma)$ are orthogonal to a,
the scalarproduct of the first row of $R_c(\sigma)$ with a equals $-n_1$ and ...
the scalarproduct of the last row of $R_c(\sigma)$ with a equals m_w.

(*) T denotes the transposition of matrices.

Proof:

We only have to show the equivalence of ii) and iii).

$$\hat{\sigma}\,|\,\mathcal{V}_e\,(f_1-f_2)=R(\sigma)\,(f_1-f_2)=\begin{bmatrix} R_x(\sigma) & 0\\ & \\ R_c(\sigma) & I\end{bmatrix}(a_1,\dots,-b_s,n_1,\dots,-m_w)^T$$

Now simple matrix multiplication shows the stated equivalence.

We now have to translate extremallity of σ (with respect to $\le$) into an equivalent notion in linear algebra.

5. Transfer of $\le$

Let σ_1 be another unifier of $\{f_1, f_2\}$. By 3 c) and remark 4.1 we may assume, that $\hat{\sigma}\,|\,\mathcal{V}_e$ is a linear mapping from $\mathcal{V}_e$ to $\mathcal{U}$. By lemma 4.2 $\hat{\sigma}_1\,|\,\mathcal{V}_e$ is described by a matrix $R(\sigma_1)$, whose structure is restricted by Lemma 4.3 iii). Let $\sigma_1 \le \sigma$ i.e. $\sigma_1=\lambda\sigma$ with a suitable substitution λ. But then we also have $\hat{\sigma}_1\,|\,\mathcal{V}_e=\hat{\lambda}\,|\,\mathcal{V}_e\cdot\hat{\sigma}\,|\,\mathcal{V}_e$.

With respect to the given basis of $\mathcal{U}$ we associate with $\lambda\,|\,\mathcal{V}_e$ the matrix

$$R(\lambda)=\begin{array}{l}\left[\begin{array}{cc} R_x(\lambda) & 0\\ & \\ R_c(\lambda) & I\end{array}\right]\begin{array}{l}\}\,r\\ \\ \}\,t\end{array}\\ \quad\underbrace{\qquad}_{r}\quad\;\underbrace{\quad}_{t}\end{array}$$

with suitably chosen submatrices.
Multiplying the matrices in $R(\sigma_1)=R(\lambda)\cdot R(\sigma)$ blockwise we obtain:

5.1 Lemma:

For unifiers σ_1, σ of $\{f_1, f_2\}$ and a substitution λ we have $\sigma_1=\lambda\sigma$ if and only if the submatrices of the matrices $R(\sigma_1)$, $R(\sigma)$, $R(\lambda)$ are connected by the following two equations:

$R_x(\sigma_1)=R_x(\lambda)\cdot R_x(\sigma)$ and $R_c(\lambda)\cdot R_x(\sigma)=-R_c(\sigma)+R_c(\sigma_1)$.

By lemma 4.3 iii) the rows of $R_x(\sigma_1)$, $R_x(\sigma)$ and $R_c(\sigma_1)$, $R_c(\sigma)$ are vectors of $\mathbb{N}^{m+s} \leq \mathbb{Q}^{m+s}$. These vectors lie on a hyperplane H_0 containing 0 or on parallel hyperplanes $H_0 + v$ with $v = v(n_i)$ or $v = v(m_j)$ $(1 \leq i \leq m,\ 1 \leq j \leq w)$. The vectors $v \times (n_i)$ are determined by $v(n_i).a = -n_i$ and $v(m_j)$ is determined by $v(m_j).a = m_j$.

The componentwise partial order of $\mathbb{Q}^{m+s}$ is inherited to $\mathbb{N}^{m+s} \cap H_0$, $N^{m+s} \cap (H_0 + v)$. We anticipate results of the next chapter which state that $N^{m+s} \cap H_0$ and $N^{m+s} \cap (H_0 + v)$ have only finitely many minimal elements with respect to componentwise order (s. [CP 61]). We let $\mathfrak{B} := \{h_1,...,h_k\}$, $\mathfrak{B}(v) := \{h_1(v),..., h_{k(v)}(v)\}$ respectively be these finite sets. Now the main result about AC-unification can be formulated.

5.2 Theorem:

For elements f_1, f_2 in $\mathfrak{F}$ the set $\mu U\Sigma(\{f_1, f_2\})$ is finite. Let σ be a most general unifier of $\{f_1, f_2\}$. Then - passing to an equivalent unifier if necessary - $R(\sigma)$ is determined as follows:

i) $R_x(\sigma)$ is the $k \times (n+s)$ matrix whose i - th row is the vector h_i from $\mathfrak{B}$ $(1 \leq i \leq k)$.

ii) $R_c(\sigma)$ is a $t \times (n+s)$ matrix whose j - th row $(1 \leq j \leq m)$ is a vector from $\mathfrak{B}(v(n_j))$ and whose e - th row $(1 \leq e \leq w)$ is a vector from $\mathfrak{B}(v(m_e))$.

Conversely any unifier σ with associated matrix $R(\sigma)$ which satisfies i), ii) is an element of $\mu U\Sigma(\{f_1, f_2\})$.

Proof:

The finiteness of $\mu U\Sigma(\{f_1, f_2\})$ is a consequence of i), ii) and the finiteness of the sets $\mathfrak{B}$, $\mathfrak{B}(v)$ respectively.

a) We show, that - up to equivalence - the matrix $R_x(\sigma)$ which belongs to σ is given by i).
By remark a) following lemma 4.1 it suffices to proof 5.2 i) for the related AC - unification problem $\{\overline{f}_1, \overline{f}_2\}$.
According to lemma 4.3 iii) the matrix given in 5.2 i) defines a unifier

σ_1 of $\{ \overline{f}_1 , \overline{f}_2 \}$. Let σ_2 be some most general unifier of $\{ \overline{f}_1 , \overline{f}_2 \}$ with associated matrix $R_x(\sigma_2)$. Again by 4.3 iii) the rows z_i $(1 \le i \le r)$ of $R_x(\sigma_2)$ have the form $z_i = \sum_{j=1}^{k} a_{ij} h_j$. (Here we use implicitly that the minimal elements of $N^{n+s} \cap H_0$ form a basis of the semigroup $\mathbb{N}^{n+s} \cap H_0$). The coefficients a_{ij} define an $r \times k$ - matrix L. But then for a suitable substitution λ we have $L = R_x(\lambda)$.
From the definition of L we derive $R_x(\sigma_2) = R_x(\lambda) \cdot R_x(\sigma_1)$ i.e. $\sigma_2 \le \sigma_1$. By hypothesis σ_2 is a most general unifier of $\{ f_1 , f_2 \}$ hence σ_1 is most general. The arbitrary choice of σ_2 implies a) and thus 5.2 i). Passing if necessary to an equivalent unifier we may assume for the rest of the proof that $r = k$ and $R_x(\sigma)$ is given by 5.2 i).

b) We show that - up to equivalence - the matrix $R_c(\sigma)$ associated with σ is one of the matrices given in 5.2 ii).
Let us assume the first row w_1 of $R_c(\sigma)$ is not contained in $\mathbb{B}(v(n_1))$. Then $w_1 = w_{min} + \sum_{j=1}^{k} a_{ij} h_j$ with some element w_{min} in $\mathbb{B}(v(n_1))$. We choose - according to 4.3 iii) - a unifier σ_1 of $\{ f_1, f_2 \}$ such that $R_x(\sigma_1) = R_x(\sigma)$ ($R_x(\sigma)$ given by a)). Furthermore we take w_{min} as the first row of $R_c(\sigma_1)$ and choose the remaining rows of $R_c(\sigma_1)$ equal to those of $R_c(\sigma)$. Consider the $t \times k$ - matrix

$$M := \begin{bmatrix} a_{11} & a_{1k} \\ 0 & 0 \\ & \\ 0 & 0 \end{bmatrix} .$$

We compute

$$\begin{bmatrix} \Sigma a_{ij} h_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} = M \cdot R_x(\sigma_1) = -R_c(\sigma_1) + R_c(\sigma).$$

Let the substitution λ be defined by $R(\lambda) := \begin{bmatrix} I & 0 \\ M & I \end{bmatrix}$.

Then we have $R(\sigma) = R(\lambda) \cdot R(\sigma_1)$ i.e. $\sigma \le \sigma_1$.
Applying this process to the remaining rows of $R_c(\sigma)$ we finally obtain a unifier σ_2 such that $R(\sigma_2)$ is given by 5.2 i), ii) and $\sigma \le \sigma_2$. By hypothesis σ is most general, hence σ_2 is most general and b) has been proved.

c) Any matrix $R(\sigma)$ given by 5.2 i), ii) defines a most general unifier of $\{f_1, f_2\}$.
For otherwise let τ be a most general unifier of $\{f_1, f_2\}$ with $\sigma \le \tau$. Using a) and b) we may replace τ by an equivalent unifier τ' such that $R(\tau')$ satisfies 5.2 i), ii). From $\sigma \le \tau'$ we derive the existence of a matrix

$$R(\lambda) = \begin{bmatrix} R_x(\lambda) & 0 \\ R_c(\lambda) & I \end{bmatrix} \text{ such that}$$

$R_x(\sigma) = R_x(\lambda) \cdot R_x(\tau') = R_x(\lambda) \cdot R_x(\sigma)$ and
$R_c(\lambda) \cdot R_x(\tau') = -R_c(\tau') + R_c(\sigma)$.
$R_c(\lambda) \cdot R_x(\tau')$ is a matrix with nonnegative integers as entries.
$-R_c(\tau') + R_c(\sigma)$ is a matrix with nonnegative integers as entries iff $R_c(\tau') = R_c(\sigma)$. (Here we use the minimality of the vectors in the sets $\mathfrak{B}(v)$ constituting the rows of $R_c(\tau')$). Hence $R(\tau') = R(\sigma)$ and c) has been proved. a), b) and c) prove the theorem.

At this point we can compare the approaches of Livesey and Siekmann [LS 76], on the one hand and Stickel [ST 73], [ST 81], HU 79], [FA 84] on the other hand.

The first method computes solutions of the homogeneous and nonhomogenous equations defined in 4.3 iii) (via scalar products). These solutions constitute the sets $\mathfrak{B}$, $\mathfrak{B}(v)$ respectively. By theorem 5.2 these computations are necessary and sufficient for a complete determination of $\mu U\Sigma(\{f_1 , f_2\})$. Hence the method cannot be improved theoretically. (of course efficient practical improvements are possible).
Stickel treats constants as variables being subject to later considerations. Now the matrix $R(\sigma)$ carries nontrivial matrices at the positions were in lemma 4.2 trivial matrices have been placed.
This redundancy has to be eliminated later on by costly matching operations. The fact, that two terms cannot be unified will be detected at the very end of the procedure. In comparison the splitting of the problem given by 5.2 i.), ii) allows an immediate and fast test for unifiability as we shall see in the next chapter.

6. Algorithmic Aspects of AC-unification

In this chapter we contribute to the computational problems related to AC-unification. We shall discuss the decision problem and search strategies for finding the minimal solutions of homogeneous (nonhomogeneous) linear diophantine equations (over N).

6.1 The decision problem

By theorem 5.2 the decision problem is equivalent to the problem of deciding whether a homogeneous (nonhomogenous) linear diophantine equation possesses solutions over N_0. For the homogeneous equations the problem is trivial; hence we may confine ourselves to the nonhomogeneous case.
We firstly assume that both of the elements f_1, f_2 to be unified contains variables i.e.

$f_1 = \sum_{i=1}^{n} a_i\ x_i + \sum_{j=1}^{m} n_j\ c_j$, $f_2 - \sum_{k=1}^{s} b_k\ y_k + \sum_{e=1}^{w} m_e\ d_e$ with natural numbers a_i, b_k, n_j, m_e.

With $a = (a_1,\dots,a_n, \ - b_1,\dots,-b_s)$ we have - according to 4.3 iii - to check, if the equations

(*) $a\,z = b$ $(b \in \{-n_i, m_k \mid 1 \leq i \leq m , 1 \leq k \leq w\})$

can be solved over $\mathbb{N}$.

By elementary number theoretical results these equations are solvable over $\mathbb{Z}$ iff $g = \gcd\{a_1,...,a_n, -b_1,...,-b_s\}$ divides the integers n_i, m_k $(1 \le i \le m, 1 \le k \le w)$. We show, that under our hypothesis these conditions are necessary and sufficient conditions for the decision problem. The euclidean algorithm, which computers g, produces a solution $z' = (z'_1,...,z'_{n+m})$ $(z'_i \in Z)$ of the equation
(**) $a\, z' = g$.
Multiplying z' with $-n_i/g$, m_j/g respectively yields integer valued solutions $z^{(i)}$, $z^{(j)}$ of (*). From these we construct solutions of (*) over $\mathbb{N}$. Let us assume $z^{(i)}_1$ (the first component of $z^{(i)}$) is negativ and consider the equation $a_1\, z_1 - b_1 z_{n+1} = 0$. A nontrivial solution (w, w') of this equation with nonnegative integers w, w' defines a solution $q_1 := (w, o,...,w', o,...,o)$ (w' as n+1-st component) of the homogeneous equation associated with (*). Let r be a natural number such that $r\, w \ge z^{(i)}_1$. Then $r\, q_1 + z^{(i)}$ is a solution of (*) (for $b = -n_i$) with nonnegative first component. Furthermore, for every nonnegative component of $z^{(i)}$, the corresponding component of $rq_1 + z^{(i)}$ is nonnegative. Iterating this procedure we obtain a solution of $a \cdot z = -n_i$ over $\mathbb{N}$. Analogously, the other equations of (*) can be solved over $\mathbb{N}$. Hence we have proved:

<u>6.1. Lemma:</u>
If both of the elements f_1, f_2 of $\mathcal{F}$ contain variables, then $\{f_1, f_2\}$ can be unified iff the gcd of the coefficients of the variables in $f_1 - f_2$ divides the coefficients of all constants in $f_1 - f_2$.

Computer algebra provides efficient methods for the computation of the gcd of a finite set of integers. A fast algorithm has been reported by Bradley [Br 70]. In case f_1 or f_2 are variable free the decision problem is NP-complete. The construction used in lemma 6.1 does not apply, since the special homogeneous equations we used there have only trivial solutions. Hence only search decides if f_1, f_2 can be unified. Soon, reasons for the exponential search effort will become apparent. We collect these observations as follows:

<u>6.2 Lemma:</u>
If f_1 or f_2 are variable free we have to search in order to solve the corresponding decision problem. In this case the decision problem is NP-complete.

6.2 Parallel Search for Minimal Solutions of Linear Homogenous Diophantine Equations over N

By theorem 5.2 i) an AC-unification algorithm has to produce the minimal solutions (over $\mathbb{N}$) of some linear homogeneous diophantine equation. The exponential search effort (s. Lemma 6.2) forms a natural barrier for sequential strategies. We therefore propose parallel procedures with substantial speedup under the realistic assumption of limited 'termdepth'. The equation to be discussed is given by $\sum_{i=1}^{n} a_i x_i - \sum_{j=1}^{m} b_j y_j = 0$ (trivial cases are not considered.)

Huet [HT 78] has shown, that all minimal solutions of this equation are contained in a square [*] $Q := [0,b]^n \times [0,a]^m \subseteq Q^{n+m}$ with $a = \max\{a_i \mid 1 \le i \le n\}$ and $b = \max\{b_j \mid 1 \le j \le m\}$. Huet provides an algorithm which searches lexicographically in Q for minimal solutions.

Other algorithms have been reported [FT 83]. We believe that modifications of the Smith-Normalform (a construction which simulates the Gauß-Algorithm for rings) eventually will result in algebraic solution strategies.

In the following we assume the existence of an algorithm α which searches lexicographically in Q for minimal solutions. This algorithm will be applied simultaneously to certain subsquares of Q.

The following preparation will prove to be useful:

Let $S=\{1,\dots,n\} \cup \{1,\dots,m\}$ [**] and $\widetilde{\mathbb{P}}(S) = \{T \subset S \mid T \cap \{1,\dots,n\} \neq \emptyset$ and $T \cap \{1,\dots,m\} \neq \emptyset\}$.

We have $|\widetilde{\mathbb{P}}(S)| = 2^{n+m} - 2^n - 2^m + 1$.

A set $T \in \widetilde{\mathbb{P}}(S)$ splits into sets T_1, T_2 with $T_1 = T \cap \{1,\dots,n\}$ and $T_2 = T \cap \{1,\dots,m\}$.

By (S) we abbreviate the equation $\sum_{i=1}^{n} a_i x_i - \sum_{j=1}^{m} b_j y_j = 0$, more general for $T \in \widetilde{\mathbb{P}}(S)$ (T) denotes the equation $\sum_{i \in T_1} a_i x_i - \sum_{i \in T_2} b_j y_j = 0$.

As for (S) any of these equations (T) determines a square as the set of all solutions of (T). Let Q_T denote the embedding of this square into $Q := Q_S$.

[*] From now on by a square Q we mean the set of vectors in Q with nonnegative integers as components.

[**] The numbers 1,...,n ,and 1,..., m stand for variables in a linear equation, with positive, respectively negative coefficients. In order to avoid overloaded notation we agree to consider the two sets defining S as disjoint.

$\widetilde{P}(S)$ is a subset of the boolean lattice $P(S)$ (the powerset of S). This lattice admits layers. The zeroth layer is the empty set, the first layer are the one-element subsets of S etc. This structure inherits to $\widetilde{P}(S) \subset P(S)$ and allows parallel search in Q as follows.

Let R,T be elements of the i-th layer of $\widetilde{P}(S)$ and let u,v be vectors in Q_R, Q_T respectively which are neither elements of $Q_{R'}$ nor $Q_{T'}$ with $R' \subset R$ and $T' \subset T$. We show that u,v do not compare with respect to the componentwise ordering of Q^{n+m}. From this we conclude that the algorithms α, applied simultaneously to $Q_R \setminus \bigcup_{R' \subset R} Q_{R'}$, $Q_T \setminus \bigcup_{T' \subset T} Q_{T'}$, respectively does not produce a minimal solution of S twice.

By reasons of symmetry we may assume R_1 and T_1 as different sets. Let $R \setminus R \cap T = \{i_1, ..., i_r\}$, $T \setminus R \cap T = \{j_1, ..., j_r\}$. The components of u are nonzero in positions $i_1, ..., i_r$ and equal zero in positions $j_1, ..., j_r$. A dualized statement holds for v. Hence u and v do not compare.

Where should α start while searching Q_R (R in the i-th layer of $\widetilde{P}(S)$)?

If we assume search in layers α has to search only the portion $Q_R \setminus \bigcup_{R' \subset R} Q_{R'}$ of Q_R. The lexicographically smallest vector in this set is the vector with 1 in all positions corresponding to R and zero elsewhere. This vector is the startingpoint for α while completing search in Q_R.

If d processors are available for search, we can apply these processors in parallel to d squares Q_{R_k} $(1 \leq k \leq d)$. The effort needed by one of these processors while searching in Q_{R_k} is given by the measurements a_k, b_k of the square Q_{R_k}. In order to distribute work uniformly over all d processors, the R_k should be chosen such that the size of the squares R_k differ as little as possible. This can be done by sorting the Q_{R_k}. We just put a suitable ordering upon the set of all pairs (a,b) where (a,b) are the measurements of a square Q_R with an element R of the i-th layer of $\widetilde{P}(S)$.

If α finds a solution of (S) in $Q_R \setminus \bigcup_{R' \subset R} Q_{R'}$ this solution need not be considered in case smaller solutions have been found previously. Smaller solutions are to be found exclusively in squares $Q_{R'}$ with $R' \subset R$. Hence in addition to parallelity in search we have parallelity in comparison with previously found solutions.

The speedup by parallel use of α can roughly be estimated as follows:

Using d processors for parallel search the minimal solutions of Q will be

found approximately in the same time as sequential use of α needs to investigate the d-th part of Q.

We have mentioned the exponential character of the search. Since all subsquares Q_R have to be visited by α, the size of $\widetilde{P(S)}$ is a measure for the complexity of the search. Since $|\widetilde{P(S)}| = 2^{n+m} - 2^n - 2^m + 1$, $\widetilde{P(S)}$ grows exponentially with the number of variables in the equation (S).

6.3 Parallel Search for Minimal Solutions of Linear Nonhomogeneous Diophantine Equations over N

Similarly as in b) we can parallelize search for minimal solutions of

$$(S)\begin{cases} \Sigma\, a_i x_i - \Sigma\, b_j y_j = c & (c \in \mathbb{Z} \setminus \{0\}) \\ \text{or} & \\ \Sigma\, a_i x_i = c & \end{cases}$$

The rational solutions of (S) form an affine hyperplane in $\mathbb{Q}^{n+m}$. As in 6.2 we assume the existence of an algorithm β which searches lexicographically in squares for minimal solutions of (S).

Such an algorithm can be constructed from the previously used algorithm α as follows:

We turn the equation (S) $\sum_{i=1}^{n} a_i x_i - \sum_{j=1}^{m} b_j y_j = c$ into the homogeneous equation (S') $\sum_{i=1}^{n} a_i x_i - \sum_{j=1}^{m} b_j y_j = c\, z$ with $z := x_{n+1}$ for $c < o$ and $z := y_{m+1}$ for $c > o$. (Similarly we proceed for (S): $\sum_{i=1}^{n} a_i x_i = c$).

Without loss of generality we may assume $z := x_{n+1}$. Minimal solutions of S' are to be found in the square $Q_{S'} := [0,b]^{n+1} \times [0,a]^m$ with $a := \max\{a_i, c \mid 1 \le i \le n\}$ $b := \max\{b_j \mid 1 \le j \le m\}$. If $(x_1, \ldots, x_n, 1, y_1, \ldots, y_m)$ is a minimal solution of (S') then $(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is a minimal solution of (S). Conversely let $u = (x_1, \ldots, x_n, y_1, \ldots, y_m)$ be a minimal solution of (S) and assume $u' = (x_1, \ldots, x_n, 1, y_1, \ldots, y_m)$ is not a minimal solution of (S'). But then there exists a solution $v = (x_1', \ldots, x_n', 0, y_1', \ldots, y_m')$ with $x_i' \le x_i$ and $y_j' \le y_j$ $(1 \le i \le n,$ $1 \le j \le m)$. The vector $v' = (x_1', \ldots, x_n\,, y_1', \ldots, y_m)$ solves the homogeneous equation associated with (S). Hence u-v' is a solution of (S) with nonnegative integer components and $u - v' < u$, a contradiction. Therefore u' is a minimal solution of (S') and we have:

The minimal solutions of (S) correspond bijectively to the minimal solutions of (S') with 1 as n+1-st component. Searching lexicographically with α the portion of $Q_{S'}$, where the n+1-st component equals 1, we obtain an algorithm β with the desired properties.

Now we proceed as in the homogeneous case. (Again we may assume $c < o$) Let $S := \{1,...,n\} \dot\cup \{1,...,m\}$ and $\overline{\mathbb{P}(S)} := \widetilde{\mathbb{P}(S)} \dot\cup \mathbb{P}(\{1,...,n\})$. For $T \in \overline{\mathbb{P}(S)}$ we denote by (T) the corresponding nonhomogeneous equation and by (T ') the associated "homogenized" equation. Let $Q_{T'}^{(1)}$ be the points of $Q_{T'}$, carrying a 1 as n+1-st component. As in the homogeneous case using the layers or $\overline{P(S)}$ we may parallelize the search for minimal solutions in $Q_S^{(1)}$.
The set $\overline{\mathbb{P}(S)}$ corresponds to the set of subsquares which are scanned by the algorithm. By the following argument this set still can be pruned without affecting the completeness of the algorithm.
Let T be in $\overline{\mathbb{P}(S)} \setminus \mathbb{P}(\{1,...,n\})$ then with a slight modification of lemma 6.1 we have:
$Q_{T'}^{(1)}$ is empty, iff the gcd of the components of the homogeneous equation (T') does not divide c. In this case T can be eliminated from $\overline{\mathbb{P}(S)}$ without affecting the completeness of the algorithm.

7. References

[Br 70] G. Bradley, "Algorithm and bound for the gcd of n integers"
Comm. of the ACM, vol.13, no.7, 1970

[CP 61] A. Clifford, G. Preston, "The Algebraic Theory of Semigroups"
Math. Surveys, vol1, no.7, 1961

[FA 84] F. Fages, "Associative-Commutative Unification"
7th Int. Conf. on Automated Deduction LNCS
Springer Verlag, Vol. 170, 1984

[FT 83] A. Fortenbacher, "Algebraische Unifikation"
Diplomarbeit, Institut für Informatik, Univ. of Karlsruhe, 1983

[FT 84] A. Fortenbacher, "An Algebraic Approach to Unification under Associativity and Commutativity"
1st Intern. Conf. on Rewriting Techn. and Applic., Dijon 1985

[HS 85] A. Herold, J. Siekmann, "Unification in Abelian Semigroups"
SEKI-Research-Report, Univ. of Kaiserslautern, 1985

[HT 78] G. Huet, "An algorithm to generate the basis of solutions to homogeneous linear diophantine equations"
Information Processing Letters, vol.7, no.3, 1978

[HU 79] J.M. Hullot, "Associative-Commutative Pattern Matching"
IJCAI 79, Tokyo, 1979

[KNB 85] D. Kapur, P. Narendran, D. Bennunav, "Complexity of Matching Problems"
1st Intern. Conf. on Rewriting Techn. and Applic., Dijon 1985

[LS 76] M. Livesey, J. Siekmann, "Unification of Sets and Multisets".
Univ. of Karlsruhe, Techn. Report, 1976

[PL 72] G. Plotkin, "Building in Equational Theories"
Machine Intelligence, vol.7, 1972

[RO 65] J. A. Robinson, "A Machine Oriented Logic based on the Resolution Principle"
JACM 12, 1965

[SI 78] J. Siekmann, "Unification and Matching Problems"
Ph.D. thesis, University of Essex, 1978

[ST 75] M Stickel, "A complete unification algorithm for
associative-commutative functions"

[ST 76] M. Stickel, "Unification Algorithms for Artificial Intelligence
Languages"
Ph.D. thesis, Carnegie-Mellon-University, 1976

[ST 81] M. Stickel, "A Unification Algorithm for Assoc. Comm.
Functions"
JACM 28, no.3, 1981