Syntactic theories and the algebra of record terms

Didier Rémy

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél.:(1)39 63 55 11

Rapports de Recherche

N°1869

Programme 2

Calcul symbolique, Programmation et Génie logiciel

SYNTACTIC THEORIES AND THE ALGEBRA OF RECORD TERMS

Didier Rémy

Mars 93

Syntactic Theories and the Algebra of Record Terms

Didier Rémy
INRIA-Rocquencourt*

*Author's address: INRIA, B.P. 105, F-78153 Le Chesnay Cedex. Email: `Didier.Remy@inria.fr`

Syntactic Theories and the Algebra of Record Terms

Abstract

Recently, many type systems for records have been proposed. For most of them, the types cannot be described as the terms of an algebra. In this case, type checking, or type inference in the case of first order type systems, cannot be derived from existing algorithms.

We define record terms as the terms of an equational algebra. We prove decidability of the unification problem for records terms by showing that its equational theory is syntactic. We derive a complete algorithm and prove its termination. We define a notion of canonical terms and approximations of record terms by canonical terms, and show that approximations commute with unification. We also study generic record terms, which extend record terms to model a form of sharing between terms. We prove that the equational theory of generic record terms and that the corresponding unification algorithm always terminates.

Théories syntaxiques et Algébres d'enregistrements

Résumé

De nombreux systèmes de types pour les enregistrements ont été proposés récemment. Pour la plupart d'entre eux les types ne peuvent pas être décrits comme les termes d'une algèbre. La vérification de types, ou la synthèse de type dans le cas des systèmes de types d'ordre un, ne peuvent alors plus être déduits d'algorithmes connus.

Nous définissons les termes à enregistrements comme les termes d'une algèbre équationnelle. Nous prouvons la décidabilité du problème d'unification pour les termes d'enregistrements en montrant que sa théorie est syntaxique. Nous en déduisons un algorithme complet et prouvons sa terminaison. Nous définissons une notion de termes canoniques et l'approximation d'un terme d'enregistrement par des formes canoniques, et nous montrons que les approximations commutent avec l'unification. Nous étudions les termes d'enregistrements génériques qui étendent les termes d'enregistrements pour modéliser une forme de partage entre les termes. Nous montrons que la théorie équationnelle des termes d'enregistrements générique est syntaxique ainsi que la terminaison de l'algorithme correspondant.

Introduction

Type systems for records have been studied extensively in recent years. For most of them the types of records are no longer terms of an algebra. Consequently, results about unification, which yield, for example, algorithms for type inference in the case of first order languages, cannot be reused, and the problems have to be studied again from the beginning. Type inference for ML can be decomposed into two steps. The first step decomposes ML programs into unification constraints; it needs to know very little about the structure of types, as opposed to the second step which resolves the unification constrains. The author described a type system for polymorphic extensible records that uses terms of an algebra modulo equations for its types [Rém93]. Type inference for ML with sorted equational theory on types has been studied in [Rém92a]. The construction of the record algebra and its properties have been studied in [Rém90] but had not been published in English yet. Here, the results of [Rém90] are reviewed and their presentation is improved — the study is also extended to generic record terms.

Here, record terms are studied for themselves, independently any particular use. However, we motivate some of the constructions by their use as the types of a functional language with record objects. In this context we temporarily call them "record types". They enable a natural extension of functional languages with records. One interest of extending functional languages with records is to have functions that can operate polymorphically on records with different sets of fields. Records are products of variable size with labeled components. The key idea is that types must reflect the structure of values and therefore record types must be products of different sizes with named components. Types of records carry information on all labels, but only a finite number of labels have different types, so type information can be finitely represented. More motivation can be found in [Rém93]. Types of records are often defined by constructing all fields at once. This requires the introduction of a collection of symbols $\{a_1 : \_ ; \ldots a_p : \_\}$ for all subsets of the set of labels (labels form a countable set). Record types with different sets of labels are completely incompatible types, which makes their unification quite difficult. On the contrary, record types are introduced with a symbol $\Pi(\_)$ and built field by field with symbols $(a : \_ ; \_)$, ending with constant symbol abs when there are no more fields to define. Two records types with some common set of fields, for instance $(a : \tau ; (b : \sigma ; \alpha))$, and $(a : \tau ; abs)$ can share some structure, namely the skeleton $\Pi(a : \_ ; \_)$. The fields of record types may be defined in any order, and commutativity equations are used to re-order fields. Some other kind of equations are also be needed to expand rows, for instance abs into $(b : abs ; abs)$, so that the two record types may be unified.

Record types are thus types of an algebra taken modulo equations. Type inference in ML reduces to unification problems. Unification in the empty theory is well known, but there is no general algorithm for unification in an arbitrary equational theory. For some theories, there may not even exist such algorithm, or the algorithm may be inherently inefficient. Section 1 recalls definitions and a few results about equational theories. Since the basic problem is rewriting proofs of equality between terms to proofs of a certain shape, we introduce a notation of equality relation that allows the manipulation of sets of proofs that share the same pattern.

Fortunately, there is a class of equational theories, called syntactic theories, for which there exists a unification algorithm quite similar to the free unification algorithm of Martelli-Montanari [MM82]. There is no known algorithm to decide whether a theory is syntactic. The usual methods for proving syntacticness do not apply either. Section 2 briefly introduces syntacticness, and develops a framework for studying syntacticness properties of equational theories. The main result of this section is theorem 1; it gives a sufficient condition for syntacticness and is used in section 3 to prove that the theory of record terms is syntactic.

The theory of record terms is described in section 3. It is shown to be syntactic; this automatically provides a complete unification semi-algorithm, and we prove its termination. In fact record terms have more structure than required by their use as types of record objects. In section 4 we introduce canonical forms and show that record terms can be approximated by canonical forms. Approximations commute with unification. In section 5 we study an extension of record terms with more structure, the generic record terms, which are used as the types of projections in the language Projective ML [Rém92b].

The main results of this article are theorem 6 in section 3 and theorem 11 in section 5, which states the decidability of unification in the theories of record terms and generic record terms, respectively. A secondary result is the existence of principal approximations for record terms and their commutation with unification, stated by theorem 9. The method used for proving theorem 1 and lemma 35 is also interesting and can probably be reused in other situations.

1 Equational theories

This section introduces the main definitions and some known results about unification in equational theories.

1.1 Sorted Free Algebras

We are given a set $\mathcal{K}$ of atomic sorts, written ι. Signatures are non-empty tuples of sorts, written ι for a one-element signature or $\iota_1 \otimes \ldots \iota_p \Rightarrow \iota_0$ for longer ones. The integer p (zero for a one element signature) is the arity of the signature. We are given a set of symbols $\mathcal{C}$ and a mapping $\mathcal{S}$ from $\mathcal{C}$ to the set of signatures. The arity of a symbol f is the arity of its signature, written $\varrho(f)$. The sort of a symbol is the right-most sort of its signature. Finally, we are given a set of variables $\mathcal{V}$ with infinitely many variables of every sort $(\mathcal{V}^\iota)_{\iota\in\mathcal{K}}$. The set of terms is the sorted free algebra $\mathcal{T}(\mathcal{V}, \mathcal{S})$. The mapping $\mathcal{S}$ is also called the signature of the algebra. Variables are written with the letters α, β, γ, and δ, and terms with the letters τ, σ, and ρ.

The set of variables appearing in a term τ is written $\mathcal{V}(\tau)$. We implicitly coerce a term τ into the set of its variables $\mathcal{V}(\tau)$ when a set of variables is required. For instance, two terms are said to be disjoint if their sets of variables are disjoint. The top symbol of a non-variable term τ, written $\mathrm{Top}(\tau)$, is the symbol at the empty occurrence in τ. For any symbol f of arity p, we write $f(\mathcal{T}^p)$ for the set of terms whose top symbol is f. The sort of a term is the sort of its top symbol. Two terms of the same sorts are said *homogeneous*.

Substitutions are sort-respecting mappings of finite domains from the set of variables to the set of terms. They naturally extend to mappings from terms to terms by compatibility with the structure of algebra. Substitutions are written with the letters μ, ν, and ξ. The domain of a substitution μ, written $dom(\mu)$, is the set of variables that are not their own images and the range of μ, written $im(\mu)$, is $\mu(dom(\mu))$. We say that a set of variables is disjoint from a substitution if it is disjoint from both the domain and the range of the substitution. We write $\mu \restriction W$ for the restriction of substitution μ to the set W, and $\mu \setminus W$ for the restriction of μ to the set $\mathcal{V} \setminus W$.

1.2 Equational theories

It is important to distinguish between the presentation of an equational theory and the theory generated by the presentation. A *presentation* of an equational theory is a set E of homogeneous pairs of terms called axioms. A congruence is an equivalence that is compatible with the structure of algebra. The *E-equality* on $\mathcal{T}$ generated by the presentation E, written $=_E$, is the smallest congruence containing all possible substitutions of the axioms. The *equational theory* $\mathcal{T}/E$ is the quotient of $\mathcal{T}$ by E. Several presentations may define the same equational theory (for instance, if one presentation extends the other with a pair of terms that can be proved equal in the equational theory the first).

We write q for an arbitrary axiom and q^l and q^r for its first and second projection, respectively. We often assume that presentations are closed under symmetry, and we write q^{-1} for the axiom (q^r, q^l).

1.3 Equality relations

The basic notion in studying equational theories, and in particular syntacticness, is the transformation of proofs. Equality relations are a formal way of manipulation proofs matching a certain pattern. The correctness of a proof transformation can be formalized by the assertion that one equality relation is a sub-relation of another.

We write $\tau_{/u}$ for the sub-term of τ at the occurrence u. We write $\tau[\sigma/u]$ for the term obtained from τ by replacing the subterm at the occurrence u by σ. Two terms τ and σ can be proved equal in one step if there exists an axiom q, an occurrence u, and a substitution μ such that $\tau_{/u}$ is $\mu(q^l)$ and $\sigma_{/u}$ is $\mu(q^r)$. In this case, we write $\tau \longrightarrow \sigma$. When more information is needed, we may write $\tau \xrightarrow[q/u]{} \sigma$. The former relation is symmetric, but the latter is not. We formalize and extend this notation below, so that it can be rigorously used in proofs.

An *equality step* is any sub-relation of $\longrightarrow$. Arbitrary equality steps are written $\xrightarrow[X]{}$, $\xrightarrow[Y]{}$, and $\xrightarrow[Z]{}$. An *Equality relation* is any composition of equality steps. They are sub-relations of $=_E$. Arbitrary equality relations are written $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$.

We write for Δ the identity relation. We write $\mathcal{X}^\delta$ and $\xrightarrow[X]{\delta}$ for equality steps, the union of the relation $\mathcal{X}$ with the identity relation Δ. The *composition* of two equality relations $\mathcal{X}$ and $\mathcal{Y}$, written $\mathcal{X}\mathcal{Y}$, relates any two terms τ and σ such that there exists a term ρ satisfying $\tau \ \mathcal{X} \ \rho \wedge \rho \ \mathcal{Y} \ \sigma$. The *union* of two equality relations relates any two terms that can be proved equal by either of the relations.

If K is a totally ordered set and $(\mathcal{X}_k)_{k\in K}$ is a sequence of equality relations equal to the identity after some rank N, we write

$$(\mathcal{X}_k)^{k\in K} \qquad \text{and} \qquad \xrightarrow[X_k]{k \in K} \qquad \text{for equality steps,}$$

the composition of all relations taken in the increasing order. When X does not depend on k we write $\mathcal{X}^K$ (or $\xrightarrow[X]{K}$) for short. We also write

$$\mathcal{X}^* \qquad \text{for} \qquad \bigcup_{q\in N} \mathcal{X}^q \qquad \text{and} \qquad \xrightarrow[X]{*} \qquad \text{for equality steps.}$$

The E-equality is just $\xrightarrow{*}$. For all terms τ and σ that are E-equal, there exists a sequence of relations

$$\xrightarrow[X_i]{i \in [1,p]}$$

and a sequence of terms $(\rho_i)_{i\in[1,p-1]}$ such that

$$\tau \xrightarrow[X_1]{} \rho_1 \dots \xrightarrow[X_i]{} \rho_i \dots \xrightarrow[X_{p-1}]{} \rho_{p-1} \xrightarrow[X_p]{} \sigma$$

Exhibiting these relations and these terms realizes a proof of $\tau =_E \sigma$. We say that the proof matches the relation $\xrightarrow[X_i]{i \in [1,p]}$.

Given two equality relations $\mathcal{X}$ and $\mathcal{Y}$, we write $\mathcal{X} \subset \mathcal{Y}$ whenever $\mathcal{X}$ is a sub-relation of $\mathcal{Y}$. Being a sub-relation can also be understood as set inclusion, viewing relations as their graphs, as we often do.

1.4 Paths

Occurrences are not enough precise. They define a subterm of a term, but are unable to tell anything about the symbols encountered on the way from the root to the subterm. A *direction* is a pair (f, x), also written f_x, of a symbol f and an integer x. A *path* is a finite sequence of directions. We say that a path u equal to $(f_i, x_i)_{i \in [1,p]}$ is a path in τ, or that τ contains the path u, if

1. the occurrence $(x_i)_{i \in [1,p]}$ is an occurrence in τ, and
2. for all k in $[1, p]$, the symbol f^p is at the occurrence $(x_i)_{i \in [1,p-1]}$.

We write $\tau_{/u}$ for the subterm of τ at the occurrence $(x_i)_{i \in [1,p]}$. We write ϵ for the empty path. Two paths are *disjoint* if neither is a prefix of the other. For instance, the sequence $(f, 1)(g, 2)$ is a path, abbreviated as $f_1 g_2$. The associated occurrence is the sequence 12 of length 2.

If q is the axiom (q^l, q^r) and u is a path, we write $\xrightarrow[q/u]{}$ for the equality step that proves

$$\tau[\mu(q^l)/u] =_E \tau[\mu(q^r)/u]$$

for any term τ containing u and any substitution μ. If u is a path and $\tau \xrightarrow[q/u]{} \sigma$ a one step proof, we get a proof $\tau_{/u} \xrightarrow[q/\epsilon]{} \sigma_{/u}$, called the sub-proof at path u.

Paths are much more precise than occurrences; sometimes they are too precise. An occurrence u can be considered as the set of all paths whose occurrences are u. The union of two equality steps $\xrightarrow[q/u]{}$ and $\xrightarrow[q/v]{}$ is an equality step. We write it $\xrightarrow[q/u \cup v]{}$. More generally, we write $\xrightarrow[q/U]{}$ for

$$\bigcup_{u \in U} \xrightarrow[q/u]{}$$

when U is a set of paths. By default, U is the set of all paths and can be omitted.

Proposition 1 *For any disjoint sets of paths U and V, the two equality steps $\xrightarrow[U]{}$ and $\xrightarrow[V]{}$ commute in $\mathcal{T}$, that is*

$$\xrightarrow[U]{}\xrightarrow[V]{} = \xrightarrow[V]{}\xrightarrow[U]{}$$

The following notations help in manipulating sets of paths:

- We identify a path u with the singleton $\{u\}$.
- For any symbol f of non-zero arity p, we write for $\underline{f}$ for the set of paths $\{(f, x) \mid x \in [1, \varrho(f)]\}$.
- For any integer p, we write $\underline{p}$ for the set of paths $\{(f, p) \mid f \in \mathcal{C} \wedge \varrho(f) \geq p\}$. For large values of p, this set may be empty.

- We write $\bullet$ for the path of length 1 composed of the set of all directions.

- If U and V are sets of paths, we write UV for the concatenation of U and V, composed of all the concatenations of any path of U with any path of V.

- We write U^δ for the union $(\epsilon \cup U)$.

- If K is a totally ordered set and $(U_k)_{k\in K}$ is a sequence of path sets equal to ϵ after some rank N, we write $(U_k)^{k\in K}$ for the concatenation of all path sets (distinct from ϵ) and taken in the increasing order. When U does not depend on k and K is finite, we write U^K for short. We write U^* for the union $\bigcup_{k\in N} U^k$ and ∞ for $(\bullet^*)$.

For instance, $(\bullet\infty)$ denotes the set of all paths of length strictly greater than one.

The union of the equality steps $\xrightarrow[q/u]{}$ and $\xrightarrow[s/u]{}$ is an equality step; it is written $\xrightarrow[q\cup s/u]{}$. That is, axioms are replaced by sets of axioms: If R is a subset of E, we write $\xrightarrow[R/u]{}$ for the equality step

$$\bigcup_{q\in R}\left(\xrightarrow[q/u]{}\right)$$

By default, R is E and can be omitted. For example, the expression $\xrightarrow[\bullet^k\infty]{*}\xrightarrow[\bullet]{\delta}$ relates any two terms that are provably equal with any number of steps at a path of length at least k followed eventually by one step at the empty path.

The expression $\xrightarrow[\delta]{}\xrightarrow[\bullet\infty]{*}\xrightarrow[2\infty]{}$ relates any two terms that are provably equal with one step at the empty path followed by any number of steps at non-empty paths and one step in the second direct subterm.

1.5 Restriction of equality relations

In the following, we will be interested in transforming proofs inside a subset of $\mathcal{T}$. This is formalized by restriction of equality relations. The *restriction of the equality step* $\xrightarrow[X]{}$ to a subset H of $\mathcal{T}$, noted $\xrightarrow[X\restriction H]{}$ is the relation $\xrightarrow[X]{} \cap H^2$ in H. The *restriction of the equality relation* $\mathcal{X}$ to a subset H of $\mathcal{T}$, noted $\mathcal{X}\restriction H$, is the equality relation $\mathcal{X} \cap \xrightarrow[\restriction H]{*}$ in H; it always relates terms that can be proved equal in H. If H is a subset of $\mathcal{T}$, if τ and σ are in $\mathcal{T}$, and if all auxiliary terms are in H, we say that the proof is a *proof in* H.

The restriction to a set of terms H that is not closed under E-equality might be dangerous, in the sense that certain obvious properties in $\mathcal{T}$ might not hold in H. The reason is that restriction of an equality relation as we defined it is not the restriction of the relation in the usual meaning. If the set H is closed under E-equality, then the two notions coincide.

Definition 1 A subset H of $\mathcal{T}$ is *closed* if it contains all its subterms and satisfies

$$\forall \tau, \sigma \in H,\ \forall u \in dom(\tau),\ (\tau_{/u} \rightarrow \sigma) \Longrightarrow \tau[\sigma/u] \in H.$$

$\square$

In the rest of the article, we only allow the restriction of relations by closed subsets of $\mathcal{T}$.

This does not imply that H is closed by E-equality (if τ is not in H, then $\sigma[\tau_{/u}]$ may not be in H), but at least it has two interesting properties:

- A subproof of a proof in H at a defined occurrence is a proof in H.

- Two steps at disjoint paths commute.

These properties are highly desirable before any serious surgery on proofs can be done.

We write $\mathcal{X} \subset_H \mathcal{Y}$ for $\mathcal{X} \restriction H \subset \mathcal{Y} \restriction H$ and we say that $\mathcal{X}$ is a sub-relation of $\mathcal{Y}$ in H. The relation $\subset_H$ is reflexive and transitive.

1.6 Sorts and equational theories

Sorting terms is a common way of restricting the set of admissible terms. The addition of sorts to an equational theory can have two effects on unification:

- Since fewer terms can meet, difficult cases may sometimes disappear.
- There are fewer proofs, which can be either helpful or harmful. In some particular cases, all proofs between valid terms are valid proofs.

The first situation is unavoidable, because sorts are aimed at restricting the set of terms. It can only make unification easier. The second situation may have no effect, if all proofs between homogeneous terms are still permitted.

Let $\mathcal{T}$ be a set of terms, $\mathcal{S}$ a sort signature, and E an equational theory. We note $\mathcal{T} \restriction \mathcal{S}$ the restriction of $\mathcal{T}$ by the signature $\mathcal{S}$. In general, axioms of E may violate the sorts. We say that an axiom has sort ι if both left and right hand sides of the axiom have sort ι. We write $E \restriction \mathcal{S}$ the set of well-sorted axioms for $\mathcal{S}$.

We are interesting in the comparison of equality in the two theories $(\mathcal{T} \restriction \mathcal{S})/(E \restriction \mathcal{S})$ and $(\mathcal{T}/E)\restriction\mathcal{S}$, that is, the comparison of the equality relations $\left(\rightarrow \cap (\mathcal{T} \restriction \mathcal{S})^2\right)^*$ and $\xrightarrow{*} \cap (\mathcal{T}\restriction\mathcal{S})^2$. In particular, when these relations are equal, any unification algorithm for $\mathcal{T}/E$ also solves unification in $(\mathcal{T} \restriction \mathcal{S})/(E \restriction \mathcal{S})$.

Definition 2 We say that a signature Σ and a presentation E are *compatible* if, for all axioms q and all sorts ι,

$$\Sigma \vdash q^l :: \iota \iff \Sigma \vdash q^r :: \iota.$$

□

Proposition 2 *If the presentation E is regular and compatible with the signature Σ, then the algebras $(\mathcal{T}/E) \restriction \Sigma$ and $(\mathcal{T} \restriction \Sigma)/(E \restriction \Sigma)$ are isomorphic.*

Proof: We show that $\left(\rightarrow \cap (\mathcal{T} \restriction \Sigma)^2\right)^*$ and $\xrightarrow{*} \cap (\mathcal{T} \restriction \Sigma)^2$ are equal. That is, any proof in $\mathcal{T}$ of equality between two terms in $\mathcal{T} \restriction \Sigma$ has all intermediate subterms in $\mathcal{T} \restriction \Sigma$, which is a straight-forward consequence of the property

$$\forall \tau \in \mathcal{T} \restriction \Sigma,\ \forall \sigma \in \mathcal{T},\ \tau \rightarrow \sigma \Longrightarrow \sigma \in \mathcal{T} \restriction \Sigma$$

For some axiom q and some occurrence u, τ/u matches q^l, that is, there exists some substitution μ such that $\tau_{/u}$ is equal to $\mu(q^l)$. Since τ is in $\mathcal{T} \restriction \Sigma$, both q^l and the restriction of substitution μ to $\mathcal{V}(q^l)$ are well sorted. The restriction of μ to $\mathcal{V}(q^r)$ is identical to its restriction to $\mathcal{V}(q^l)$ because the theory is regular. Because of compatibility, q^r has the same sort as q^l. Thus $\tau[\mu(q^r)/u]$, that is, σ is well sorted. ■

The property is often used in one of the following two cases:

- The presentation E is well-sorted.
- The signature removes symbols in a way that is compatible with E.

2 Syntactic theories

Syntactic theories were introduced by C. Kirchner [Kir85].

Assumption In the rest of this article, we assume that the theory is *collapse free*, that is, there is no axiom composed of a variable on one side.

For any presentation E and any symbols f and g of respective arities p and q, we write $E(f,g)$ the set $E \cap f(\mathcal{T}^p) \times g(\mathcal{T}^q)$ of axioms whose top symbols are f and g, respectively.

Definition 3 A pair of symbols (f,g) is *syntactic* for the presentation E if, for all equal pairs of terms of $f(\mathcal{T}) \times g(\mathcal{T})$, there is a proof of their equality that uses at most one axiom at the empty path. A presentation is *syntactic* if all pairs of symbols are. A theory is *syntactic* if there exists a syntactic presentation of this theory such that all sets $E(f,g)$ are finite for all pairs of symbols (f,g). □

For a presentation to be syntactic is equivalent to

$$\xrightarrow{*} \subset \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*} .$$

Let H be a closed subset of $\mathcal{T}$. We say that a presentation is *syntactic in H* if

$$\xrightarrow{*} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*} .$$

We write *Synt*(H) when the presentation is syntactic in H.

The presentation $\{f =_E g, g =_E h\}$ is not syntactic, since the pair (f,h) is not (but the pair (f,g) is). However, the theory generated by the presentation is syntactic since the addition of the axiom $f =_E h$ does not modify the theory and makes all pairs of symbols syntactic. The empty theory and the theory $\{f(g(\alpha)) =_E g(\alpha)\}$ are syntactic, but the theory $\{f(g(\alpha)) =_E f(\alpha)\}$ is not. Many examples can be found in [Kir85] and [KK89].

Some interesting questions are:

- Is a presentation syntactic?
- Is a theory syntactic?
- How to find a syntactic presentation of a syntactic theory.
- Minimalizing the set of axioms of a syntactic presentation.

The definition of syntactic theories is purely "syntactic," but a semantic characterization was later found by C. Kirchner and F. Klay [KK89]. T. Nipkow showed that some proof transformations by rewriting techniques were closely related to syntactic presentations [Nip89]. All these studies help in understanding the second problem. A common instance of the third problem is the completion of a presentation into one that is syntactic; it has been thoroughly addressed by C. Kirchner. Our interest is only the first problem.

2.1 Sufficient conditions for commutativity

Proving that a presentation is syntactic usually requires some rearrangement of equality relations. The most frequent one is the commutation of two equality steps. The sufficient conditions fiven in section 2.2 require commutativity of consecutive equality steps of a certain shape. In general, two equality steps at disjoint paths commute. Here, we study the case when one occurrence is prefix of the other below.

Lemma 3 *Let H be a closed subset of $\mathcal{T}$. Let q be an axiom such that there is a variable α appearing exactly once in q^r at the occurrence u and appearing k times in q^l at occurrences $(u_j)_{j\in[1,k]}$. Then for any axiom s and any occurrence v,*

$$\xrightarrow[q/\epsilon]{}\xrightarrow[s/uv]{} \subset_H \xrightarrow[s/u_jv]{j\in[1,k]}\xrightarrow[q/\epsilon]{} .$$

<u>Proof</u>: A proof matching $\xrightarrow[q/\epsilon]{}\xrightarrow[s/uv]{}$ is of the form

$$\mu(q^l) \xrightarrow[q/\epsilon]{} \mu(q^r) \xrightarrow[s/uv]{} \mu(q^r)[\tau/uv].$$

Let α be $q^r/_u$. From the proof

$$\mu(q^r) \xrightarrow[s/uv]{} \mu(q^r)[\tau/uv],$$

we can extract the proof at u,

$$\mu(\alpha) \xrightarrow[s/v]{} \mu(\alpha)[\tau/v],$$

and apply it to all disjoint occurrences $(u_j)_{j\in[1,q]}$ of α in q^l. Abbreviating $\mu(\alpha)[\tau/v]$ by σ,

$$\mu(q^l) \xrightarrow[s/u_jv]{j\in[1,k]} \mu(q^l)[\sigma/u_j]_{j\in[1,k]}.$$

The right hand side of the above proof is equal to $(\mu\setminus\{\alpha\}+(\alpha\mapsto\sigma))(q^l)$. Applying axiom q at occurrence ϵ gives $(\mu\setminus\{\alpha\}+(\alpha\mapsto\sigma))(q^r)$, which is $(\mu\setminus\{\alpha\})(q^r)[\tau/u]$ since u is the only occurrence of α in q^r (the linearity condition of q^r in α is needed here). By expansion of σ, this simplifies into $\mu(q^r)[\tau/uv]$. ■

Corollary 4 *Let H be a closed subset of $\mathcal{T}$. If q is a collapse-free axiom* [1] *such that there exists one variable appearing exactly once in q^r at the occurrence u, then for any axiom s,*

$$\xrightarrow[q/\epsilon]{}\xrightarrow[s/u\infty]{} \subset_H \xrightarrow[s/\bullet\infty]{*}\xrightarrow[q/\epsilon]{}$$

<u>Proof</u>: It suffices to notice that the occurrences $(u_j)_{j\in[1,q]}$ cannot be empty. ■

Corollary 5 *Let H be a closed subset of $\mathcal{T}$. If the presentation is collapse-free and all axioms are linear (regular and such that a variable occurs at most once in each side of axioms) of depth less or equal to n, then*

$$\xrightarrow[q/\epsilon]{}\xrightarrow[s/\bullet^n\infty]{} \subset_H \xrightarrow[s/\bullet\infty]{}\xrightarrow[q/\epsilon]{} .$$

2.2 Sufficient conditions for syntacticness

Lemma 6 *The property $Synt(H)$ is equivalent to*

$$\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*} .$$

[1] This condition is important here. Though it is assumed to be true throughout this section, it was not needed for the preceding lemma.

Proof: The above condition is needed, since it is an instance of the condition *Synt*. We now assume the condition and prove *Synt*(H), that is,

$$\xrightarrow{*} \subset \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*} .$$

The equality relation $\xrightarrow{*}$ can be decomposed into

$$\bigcup_{k \in N} \left(\xrightarrow[\bullet\infty]{*} \left(\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{*}\right)^k \right) .$$

The inclusion

$$\xrightarrow[\bullet\infty]{*} \left(\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{*}\right)^k \subset \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

is shown for all k by an easy induction on k. ■

Proposition 7 *A sufficient condition for syntacticness in H is*

$$\wedge \begin{cases} \xrightarrow[\epsilon]{}\xrightarrow[\epsilon]{} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}, \\ \xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta} . \end{cases}$$

This condition was named ϵ-confluence by Claude Kirchner and proved in [Kir85]. The proof is very simple.

Proof: Under the given assumptions, the inclusion

$$\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{k}\xrightarrow[\epsilon]{} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

for all k is easily shown by induction on k. By taking the union over all integers, we get

$$\bigcup_{k \in N} \xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{k}\xrightarrow[\epsilon]{} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

That is,

$$\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

We conclude by the lemma 6. ■

This proposition is composed of two conditions. The first is usually checked for all possible coverings of axioms. The second is commutativity, and might be deduced from lemma 3. However, the lemma only applies if axioms are of depth at most one. For instance, if one member of an axiom is of depth two, commutativity might not hold when one occurrence is prefix of the other. If axioms are collapse-free, linear, and of depth at most two, then the commutation

$$\xrightarrow[\epsilon]{}\xrightarrow[\bullet\bullet\infty]{} \subset \xrightarrow[\bullet\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}$$

holds. Theorem 1 below generalizes proposition 7 to this case.

Note For any increasing sequence $(H_n)_{n \in N}$ of closed subsets of $\mathcal{T}$,

$$(\forall n \in I\!N,\ \mathcal{X} \subset_{H_n} \mathcal{Y}) \qquad \text{implies} \qquad \mathcal{X} \subset_H \mathcal{Y} \qquad \text{where} \qquad H = \bigcup_{n \in N} H_n$$

In particular this applies to the property *Synt*.

Definition 4 Let H be a closed subset of $\mathcal{T}$. An ordering $<$ is compatible with the equality relation on H if,

- for any term τ in H and any subterm σ of τ we have $\sigma < \tau$,
- for all terms τ, σ, τ' and σ' in H, we have

$$\left(\tau \xrightarrow{\delta} \tau' \wedge \sigma \xrightarrow{\delta} \sigma' \wedge \tau < \sigma\right) \Longrightarrow \tau' < \sigma'.$$

□

An equational theory is *strict* if a term is never E-equal to one of its subterms. If there exists a compatible ordering, the theory is necessarily strict. The last condition is in fact equivalent to

$$(\tau =_E \tau' \wedge \sigma =_E \sigma') \Longrightarrow (\tau < \sigma \Leftrightarrow \tau; < \sigma')$$

In the following, we will always define compatible orderings by the means of a function from terms to an ordered set, namely $\mathbb{N}$; these functions will always be constant on E-equality classes.

Theorem 1 *Let H be a closed subset of $\mathcal{T}$ with a well founded compatible ordering, and such that*

$$\underset{\epsilon}{\longrightarrow}\underset{\bullet\bullet\infty}{\longrightarrow} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta} \qquad (h_1)$$

$$\underset{\epsilon}{\longrightarrow}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\underset{\epsilon}{\longrightarrow} \subset_H \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*} \qquad (h_2)$$

Then Synt(H).

<u>Proof</u>: For any integer n, let H_n be the subset of H composed of all terms that do not start any decreasing sequence of length greater than n. All these sets are closed under $\xrightarrow[\restriction H]{}$. In particular, they are closed subsets of $\mathcal{T}$, so properties (h_1) and (h_2) are valid in any (H_n). Any term smaller than a term in H_{n+1} is in H_n. The sequence is increasing and its limit is H, thus $Synt(H)$ holds if $Synt(H_n)$ holds for any n. We show $Synt(H_n)$ by induction on n; in fact, by lemma 6, it is enough to show that $\underset{\epsilon}{\longrightarrow}\xrightarrow[\bullet\infty]{*}\underset{\epsilon}{\longrightarrow} \subset_{H_n} \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$ holds.

The set H_0 is composed of variables and constant symbols. Thus, the only instance of the premise is $\underset{\epsilon}{\longrightarrow}\underset{\epsilon}{\longrightarrow}$, for which the inclusion is satisfied by the hypothesis (h_2).

Let us assume the property $Synt(H_n)$ and prove $Synt(H_{n+1})$. Let $\mathcal{D}$ be the set of all directions. A relation $\xrightarrow[\bullet\infty]{*}$ in H_{n+1} can be written as the composition

$$\left(\xrightarrow[k\infty]{*}\right)^{k\in\mathcal{D}}$$

since disjoint occurrences commute. For each direction k in $\mathcal{D}$, the subproof at k is in H_n. Since $Synt(H_n)$, it can be rewritten in H_n so that it matches

$$\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}.$$

Re-assembling all subproofs, we get a proof matching

$$\left(\xrightarrow[k\bullet\infty]{*}\xrightarrow[k\epsilon]{\delta}\xrightarrow[k\bullet\infty]{*}\right)_,^{k\in\mathcal{D}}$$

which is a proof in H_{n+1}. It can be reordered as

$$\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}.$$

We have shown that

$$\xrightarrow[\bullet\infty]{*}\subset_{H_{n+1}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}.$$

That is,

$$\xrightarrow[\bullet\infty]{*}\subset_{H_{n+1}}\left(\xrightarrow[\bullet\bullet\infty]{*}\right)\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[\bullet\bullet\infty]{*}\right).$$

Composing the step at the root on both sides, and then using (h_1) and its symmetric image, we get

$$\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{}\subset_{H_{n+1}}\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

We conclude by a simple case analysis, using (h_2). ■

Proposition 7 and theorem 1 have two conditions. The first is a commutativity condition, and in general could be deduced from corollary 4. The second has to be proved by hand. However if the number of axioms is finite, the number of possible combinations of equality steps matching the premise is also finite. Thus it is possible to study each of them separately.

2.3 Example

The theory Cg where the only axiom is left commutativity is well known to be syntactic. We give a very short proof below. The axiom is:

$$x @ (y @ z) = y @ (x @ z)$$

The size of terms is unchanged by E-equality, which also defines a compatible ordering on $\mathcal{T}$. The condition h_1 is satisfied since the axiom is linear and the occurrence of non-variable terms is at most 1 (corollary 5). Two successive applications of the axiom at the empty occurrence annihilate each other. Since $\xrightarrow[\epsilon]{}\xrightarrow[1]{}\subset\xrightarrow[21]{}\xrightarrow[\epsilon]{}$, the only remaining case to consider in order to show (h_2) is $\xrightarrow[\epsilon]{}\xrightarrow[2]{}\xrightarrow[\epsilon]{}$, which is equal to $\xrightarrow[2]{}\xrightarrow[\epsilon]{}\xrightarrow[2]{}$. The proof schema for this relation and its reduction is shown by the diagram below (any cycle is a subset of the identity relation).

$$\begin{tikzcd}
& \beta @ (\alpha @ (\gamma @ \delta)) \arrow[r, "2"] & \beta @ (\gamma @ (\alpha @ \delta)) \arrow[dr, "\epsilon"] & \\
\alpha @ (\beta @ (\gamma @ \delta)) \arrow[ur, "\epsilon"] & & & \gamma @ (\beta @ (\alpha @ \delta)) \arrow[dl, "2"] \\
& \alpha @ (\gamma @ (\beta @ \delta)) \arrow[ul, "2"] & \gamma @ (\alpha @ (\beta @ \delta)) \arrow[l, "\epsilon"'] &
\end{tikzcd}$$

2.4 Unification in equational theories

In this section we describe how unification can be solved in syntactic theories. The problem of unification is, given a set of terms, to find a most general substitution that identifies all terms of the set. In fact, it is simpler to manipulate multi-sets of terms, called multi-equations. Since the satisfiability of a multi-equation is often reduced to the satisfiability of several multi-equations, it is also convenient to generalize unification problems to conjunctions of multi-equations called unificands.

Unificands are either multi-equations or conjunction or disjunction of unificands. They are written with letters U and V. Conjunction and disjunction of unificands are written $U \wedge V$ and $U \vee V$, respectively. The set of solutions of a conjunction (respectively disjunction) of unificands is the intersection (respectively union) of the sets of solutions of the unificands. Thus it is possible to consider unificands equal modulo associativity and commutativity of $\wedge$ and $\vee$ and modulo distributivity of one over the other. Then unificands can always be written as disjunctions of conjunctions of multi-equations.

It is also very convenient to restrict the set of solutions of a unificand U by some set of variables W. We write $\exists\, W \cdot U$ for the unificand composed of these restrictions. The $\exists$ acts as a binder, and we consider unificands equal modulo renaming of variables bound by $\exists$'s, exchange of consecutive $\exists$'s, and removal of vacuous $\exists$'s. It is convenient to add a unificand $\perp$ that has no solution and which is used to represent failure. Unificands were first introduced by C Kirchner [Kir85] and existential unificands were used later by J.-P. Jouannaud and C. Kirchner [KJ90]. A more abstract presentation of unificands can be found in [Rém92a].

A *complete set of unifiers* for a unificand U is a set $\mathcal{S}$ of solutions of U such that any other solution is an instance of at least one solution of $\mathcal{S}$. An equational theory is *unitary unifying* if any solvable conjunction of multi-equations admits a complete set of solutions composed of a unique substitution. A general method for finding complete set of unifiers of a unificand U is to transform U into a simpler unificand that has exactly the same solutions. Thus, solving a unification problem is done by building a chain of equivalent unificands, each being obtained by rewriting the preceding one with a very simple rule.

Unification in the empty theory may be described by the rules of figure 1. *Fusion* merges

$$\frac{\alpha \doteq e \wedge \alpha \doteq e'}{\alpha \doteq e \doteq e'} \leadsto \quad (\textsc{Fuse}) \qquad \frac{f(\tau_1, \ldots \tau_p) \doteq f(\beta_1, \ldots \beta_p) \wedge e}{\tau_1 \doteq \sigma_1 \wedge \ldots \tau_p \doteq \sigma_p \wedge f(\beta_1, \ldots \beta_p) \doteq e} \leadsto \quad (\textsc{Decompose})$$

$$\text{if } f \neq g, \qquad \frac{f(\tau_1, \ldots \tau_p) \doteq g(\sigma_1, \ldots \sigma_q) \doteq e}{\perp} \leadsto \quad (\textsc{Fail})$$

$$\text{if } \alpha \in \mathcal{V}(e) \setminus e \setminus \mathcal{V}(\tau) \wedge \tau \notin \mathcal{V}, \qquad \frac{(\alpha \mapsto \tau)(e)}{\exists\, \alpha \cdot (e \wedge \alpha \doteq \tau)} \leadsto \quad (\textsc{Generalize})$$

Figure 1: Rules for unification in the empty theory

two multi-equations that share a common term variable into a single multi-equation. *Collision* reduces a multi-equation that contains two terms with different top symbols to the empty unificand $\perp$. *Decomposition* splits a multi-equation that contains two non-variable terms with the same top symbol into a the conjunction of multi-equations composed of equations between the corresponding subterms. *Generalization* replaces a non-empty occurrence of a non-variable term τ in e by a variable α and adds the equation $\alpha \doteq \tau$. Generalization is used to reduce the height of terms in a unificand and prevent duplication of terms by other rules. For instance, if the right terms premise in rule DECOMPOSE were not variables, the conclusion would duplicate terms, which could prevent termination. It would be also possible to factor rules DECOMPOSE and GENERALIZE into a single rule that would not require any term of the premise to be small. Below, we use an unrestricted form of generalization where τ may be a variable, called U-GENERALIZE.

The four rules above applied in any order reduce any system of multi-equation to a completely decomposed one. A *completely decomposed unificand* is one for which no multi-equation has more than one non-variable term. An equational theory is *strict* if a term can never be a sub-term of an E-equal term. In a strict theory, it is immediate to write an algorithm that tells whether a completely decomposed unificand is solvable, and that returns a principal unifier if one exists. See [Rém92a] for more details or [Kir85, Rém90] for a more thorough but slightly different presentation.

In equational theories, decomposition only fires for some pairs of symbols, called decomposable symbols. Collision fires only when the top symbols are incompatible. To be complete, there must be other transformations, called *mutations*, that together with previous rules reduce any unificand into a completely decomposed unificand. Mutation may not exist if there is no complete set of rules that can be added to fusion and decomposition and that terminate on any input with a completely decomposed system. Mutation often introduces disjunction of unificands.

Syntactic theories have a very simple mutation that, moreover, can be automatically deduced from a syntactic presentation of the theory. Multi-equations can always be reduced by looking at the top symbol of terms, which leads to efficient algorithms that are similar to the Martelli-Montanari unification algorithm in the empty theory [MM82].

Mutation in syntactic theories

In syntactic theories, the mutation is a generalization of decomposition and is derived from the form of the axioms.

Definition 5 Let E be a syntactic theory. Let τ and σ be two non-variable terms. We define *decomposition* of the equation $\tau \doteq \sigma$ to be the following system, written $\mathrm{Dec}\,(\tau \doteq \sigma)$:

$$\bigwedge_{i\in[1,p]} \left(\tau_{/i} \doteq \sigma_{/i}\right)$$

We define *generalization* of the equation $\tau \doteq \sigma$ to be the disjunction of systems

$$\bigvee_{q\in E(f,g)} \left(\exists\, \mathcal{V}(q) \cdot \bigwedge \begin{cases} \tau_i \doteq {q^l}_{/i} & i \in [1,p] \\ \sigma_i \doteq {q^r}_{/j} & j \in [1,q] \end{cases}\right)$$

written $\mathrm{Gen}\,(\tau \doteq \sigma)$, where the axioms of $E(f,g)$ have been renamed so that they do not share any variable with the terms τ and σ. We define mutation of the equation $\tau \doteq \sigma$ to be the disjunction of systems

$$\bigvee \begin{cases} \mathrm{Gen}\,(\tau \doteq \sigma) \vee \mathrm{Dec}\,(\tau \doteq \sigma) & \text{if } \mathrm{Top}\,(\tau) = \mathrm{Top}\,(\sigma) \\ \mathrm{Gen}\,(\tau \doteq \sigma) & \text{otherwise} \end{cases}$$

This system is written $\mathrm{Mut}\,(\tau \doteq \sigma)$. □

Theorem 2 *If E is a syntactic presentation and τ and σ are two non variable terms, then*

$$\frac{\tau \doteq \sigma}{\mathrm{Mut}\,(\tau \doteq \sigma)} \rightsquigarrow$$

Proof: The following is proved is [Kir85]. We first show that the reduction is correct. Let τ and σ be two terms and μ a solution of the equation $\tau \doteq \sigma$. We show that μ satisfies Mut $(\tau \doteq \sigma)$. We use the following remark

$$\forall \tau, \sigma \in \mathcal{T},\ \tau \xrightarrow[\bullet\infty]{*} \sigma \Longrightarrow \bigwedge \begin{cases} \mathrm{Top}\,(\tau) = \mathrm{Top}\,(\sigma) \\ \forall i \in [1,\ \varrho(\mathrm{Top}\,(\tau))], \tau_{/i} =_E \sigma_{/i} \end{cases}$$

Since the theory is syntactic, there exists a proof of the form

$$\tau\mu \xrightarrow[\bullet\infty]{*} \tau' \xrightarrow[\epsilon]{\delta} \sigma' \xrightarrow[\bullet\infty]{*} \sigma\mu.$$

If there is no step at the empty occurrence, then the top symbols of τ and σ must be equal, and the system Dec $(\tau \doteq \sigma)$ is satisfied by μ. If one axiom q is used at the empty occurrence, it must be in $E(\mathrm{Top}\,(\tau'), \mathrm{Top}\,(\sigma'))$, and there must be a substitution ν of domain $\mathcal{V}(q)$ for which

$$\tau' = \nu(q^l) \qquad \text{and} \qquad \sigma' = \nu(q^r)$$

The substitution $\mu \restriction \mathcal{V}(q) + \nu$ is a solution of the system

$$\bigwedge \begin{cases} \tau_i \doteq {q^l}_{/i} & i \in [1, p] \\ \sigma_i \doteq {q^r}_{/j} & j \in [1, q] \end{cases}$$

It follows from the remark that q is in $E(\mathrm{Top}\,(\tau), \mathrm{Top}\,(\sigma))$ and μ is a solution of Gen $(\tau \doteq \sigma)$.

Conversely, let us assume that μ is a solution of Mut $(\tau \doteq \sigma)$. If it is a solution of Gen $(\tau \doteq \sigma)$, then there is an axiom q of $E(\mathrm{Top}\,(\tau), \mathrm{Top}\,(\sigma))$ and a substitution ν such that the substitution $\mu \restriction \mathcal{V}(q) + \nu$ is a solution of the system

$$\bigwedge \begin{cases} \tau_i \doteq {q^l}_{/i}\ i \in [1, p] \\ \sigma_i \doteq {q^r}_{/j} j \in [1, q] \end{cases}$$

By composing the proofs of these equalities we get

$$(\mu \restriction \mathcal{V}(q) + \nu)(\tau) \xrightarrow{*} (\mu \restriction \mathcal{V}(q) + \nu)(q^l)$$
$$(\mu \restriction \mathcal{V}(q) + \nu)(\sigma) \xrightarrow{*} (\mu \restriction \mathcal{V}(q) + \nu)(q^r)$$

which simplifies to

$$\mu(\tau) \xrightarrow{*} \nu(q^l) \qquad \text{and} \qquad \mu(\sigma) \xrightarrow{*} \nu(q^r)$$

Thus $\mu(\tau) \xrightarrow{*} \mu(\sigma)$. The case where the top symbols are equal and μ is a solution of Dec $(\tau \doteq \sigma)$ is immediate. ■

To get an algorithm for solving unification, the existence of mutation is not sufficient; the termination of mutation together with the other rules must also be proved.

Note The presentation of left commutativity in the previous section is syntactic, but the corresponding mutation does not provide a unification algorithm since the mutation itself might not terminate. A solution is proposed in [Kir85].

3 Record terms

Record terms, used as types of record objects, enable a natural extension of functional languages with records. Records are products of variable size with labeled components. The key idea is that types must reflect the structure of values and therefore record types must be products of different sizes with named components. We first study a simplification of record terms obtained by forgetting labels and accessing components by position instead of by name: these structures are infinitary tuples.

3.1 Infinitary tuples

Let $\mathcal{C}$ be a set of symbols given with their arities $(\mathcal{C}_n)_{n\in N}$. Let $\mathcal{K}$ be composed of

- a sort $Type$ and
- a countable collection of sorts $(Row(n))_{n\in N}$.

Let Σ be the signature composed of the following symbols, given with their sorts:

$$\begin{array}{ll} \Sigma \vdash \Pi :: Row(0) \Rightarrow Type & \\ \Sigma \vdash f^{\iota} :: \iota^{\varrho(f)} \Rightarrow \iota & f \in \mathcal{C}, \iota \in \mathcal{K} \\ \Sigma \vdash @^n :: Type \otimes Row(n+1) \Rightarrow Row(n) & n \in I\!N \end{array}$$

Let E be the set of axioms

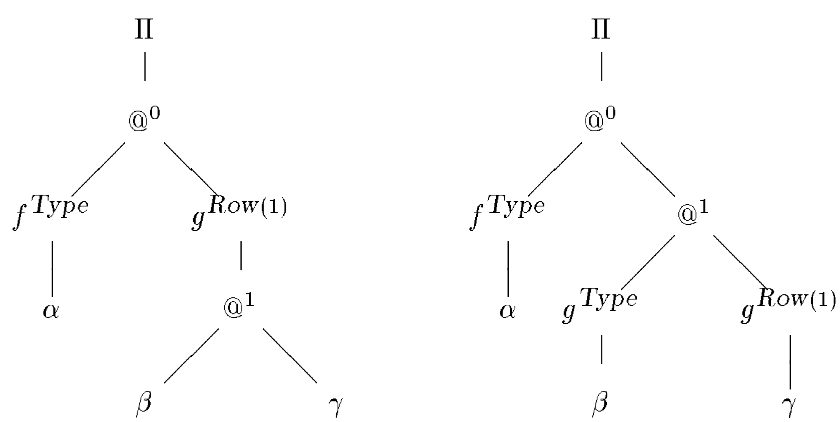
$$f^{Row(n)}(\alpha_1 @^n \beta_1, \ldots \alpha_p @^n \beta_p) = f^{Type}(\alpha_1, \ldots \alpha_p) @^n f^{Row(n+1)(\beta_1,\ldots\beta_p)} \qquad (f \triangleright n)$$

All axioms are collapse-free, regular and linear.

Let $\mathcal{V}$ be a denumerable set of variables with infinitely many variables of every sort.

Definition 6 The algebra of infinitary tuples is the equational theory $\mathcal{T}(\Sigma, \mathcal{V})/E$. □

The following two infinitary tuples are E-equal



Theorem 3 *The presentation E is syntactic.*

<u>Proof</u>: Lemma 7 is not sufficient because the commutativity condition is not satisfied; for instance the proof

$$\begin{tikzcd}[column sep=tiny] & f^{Type}(\alpha_i)_I @^n f^{Row(n)}\left((\beta_i)_I @^{n+1} (\gamma_i)_I\right) \arrow[dl, "\epsilon"'] \arrow[dr, "2"] & \\ f^{Row(n+1)}\left(\alpha_i @^n (\beta_i @^{n+1} \gamma_i)\right)_I & & f^{Type}(\alpha_i)_I @^n \left(f^{Type}(\beta_i)_I @^{n+1} f^{Row(n+1)}(\gamma_i)_I\right) \end{tikzcd}$$

where indexing expression $_{i\in I}$ is written $_I$ for short, cannot start with a step at the empty occurrence.

Of course, this proof is subsumed by the proof below for record terms. In particular, theorem 1 can be used, leading to a shorter proof, but it would require controlling terms with an ordering (as in the next section). In fact there would be no instance of the relation

$$\underset{\epsilon}{\longrightarrow}\left(\underset{k}{\overset{\delta}{\longrightarrow}}\right)^{k\in N}\underset{\epsilon}{\longrightarrow}$$

We give a direct proof. In fact we show that

$$\underset{\epsilon}{\longrightarrow}\xrightarrow[\bullet\infty]{n}\underset{\epsilon}{\longrightarrow}\subset\xrightarrow[\bullet\infty]{n}$$

by induction on n. We will use the following remarks, which are immediate consequences of the form of the axioms.

- Two successive steps at the empty occurrence must be inverse and they annihilate each other. This solves the initial case (n is zero).
- Two applications of axioms at disjoint occurrences, or at occurrences such that one is prefix of the other but at least too directions shorter, commute.
- A proof between two terms with the same top symbol cannot have exactly one step at the empty occurrence.

Let us assume that the property is true for n and consider a proof matching

$$\underset{\epsilon}{\longrightarrow}\xrightarrow[\bullet\infty]{n+1}\underset{\epsilon}{\longrightarrow}$$

We consider the subproof of $\xrightarrow[\bullet\infty]{n+1}$ at a direction i.

If it does not start or end with an application of an axiom at the empty occurrence, then this equality step commutes with one of the equality steps at the empty occurrence in the original proof. Thus the original proof is of the form $\underset{\epsilon}{\longrightarrow}\xrightarrow[\bullet\infty]{n}\underset{\epsilon}{\longrightarrow}\xrightarrow[\bullet\infty]{1}$, and by induction hypothesis it is also of the form $\xrightarrow[\bullet\infty]{n}\xrightarrow[\bullet\infty]{1}$.

Otherwise, the subproof at occurrence i has at least two steps at the empty occurrence; thus it matches

$$\underset{\epsilon}{\longrightarrow}\xrightarrow[\bullet\infty]{p}\underset{\epsilon}{\longrightarrow}\xrightarrow[\infty]{q}$$

for some p and q whose sum is strictly smaller than n. By the induction hypothesis, this relation is a sub-relation of

$$\xrightarrow[\bullet\infty]{p}\underset{\epsilon}{\longrightarrow}\xrightarrow[\infty]{q}$$

Re-composing this with the original proof, the problem reduces to previous case. ∎

The mutation for the infinitary tuple algebra:

$$\frac{f^{Row(n)}(\sigma_i)_{i\in[1,p]} \doteq (\alpha \mathbin{@^n} \gamma) \doteq e}{\exists\,(\alpha_i,\gamma_i)_{i\in[1,p]}\cdot\bigwedge\begin{cases}(\alpha \mathbin{@^n} \gamma) \doteq e\\ \alpha \doteq f^{Type}(\alpha_i)_{i\in[1,p]}\\ \gamma \doteq f^{Row(n+1)}(\gamma_i)_{i\in[1,p]}\\ \sigma_i \doteq \alpha_i \mathbin{@^n} \gamma_i \qquad \text{for} \qquad i\in[1,p]\end{cases}}\rightsquigarrow \qquad \text{MUTATE}$$

For all other pairs of terms (τ,σ), if they have identical top symbols, they are decomposable, otherwise they produce a collision. Mutation rules can be generalized as shown above.

Theorem 4 *Unification in the infinitary tuple algebra is decidable and unitary unifying.*

Proof: The theory is strict (we can find a compatible ordering as we will do below for record terms). Therefore, the rules of mutation, decomposition, collision, fusion and generalization applied any order form a complete semi-algorithm for unification. It is unitary since mutation does not introduce any disjunction. It remains to prove termination.

Let $\mathcal{T}^n$ be the subsets of $\mathcal{T}$ restricted to the sorts *Type* and $Row(k)$ for k smaller or equal to n. Generalized decompositions and fusion are stable on the sets $\mathcal{T}^n$, and for a system of unificands they decrease in the following lexicographic ordering:

- the number of symbols $f^{Row(n)}$ in the lexicographic order of increasing n,
- the number of all other symbols,
- the number of multi-equations,
- the sum of heights of terms of the unificand.

This guarantees the termination of the process. ■

Note The introduction of an infinite collection of copies of the original set of symbols $\mathcal{C}$ might be considered luxurious, while two copies *Type* and *Row* would seem sufficient. The main symbol @ would have signature $Type \otimes Row \Rightarrow Row$. The axioms would be

$$f^{Row}(\alpha_1 \mathbin{@^n} \beta_1, \ldots \alpha_p \mathbin{@^n} \beta_p) = f^{Type}(\alpha_1, \ldots \alpha_p) \mathbin{@} f^{Row(\beta_1, \ldots \beta_p)}$$

But even if the presentation remained syntactic, termination could not be guaranteed as above.

Note We can consider the set of raw terms, that is, the algebra $\mathcal{T}'$ built from the all symbols $\mathcal{C}$ extended with a binary symbol @ and a unary symbol Π. To any term τ in $\mathcal{T}$ there corresponds a raw term obtained from τ by removing all superscripts of symbols. Conversely, for any raw term τ' and any sort ι, there is at most one term of sort ι whose erasure is τ'. This allows us to define a term of $\mathcal{T}$ by giving its erasure and its sort.

3.2 Record terms

We generalize the theory of infinitary tuples to the theory of record terms, where components are named. We describe the theory and show that it is syntactic and that unification is decidable.

The algebra of record terms is define relatively to is a collection of symbols given with their arities $(\mathcal{C}_n)_{n \in N}$. Let $\mathcal{L}$ be a countable set of labels. Labels are written a, b, c and ℓ, finite subset of $\mathcal{L}$ are written L and K the set of all of them is written $\mathcal{P}_f(\mathcal{L})$. We also write $a.L$ for $a.L$.

Let $\mathcal{K}$ be the set composed of

- a sort *Type*, and
- a finite collection of sorts $(Row(L))_{L \in \mathcal{P}_f(\mathcal{L})}$.

Let Σ be the signature composed of the following symbols given with their sorts:

$$\begin{array}{ll}
\Sigma \vdash \Pi :: Row(\emptyset) \Rightarrow Type & \\
\Sigma \vdash f^\iota :: \iota^{\varrho(f)} \Rightarrow \iota & f \in \mathcal{C}, \iota \in \mathcal{K} \\
\Sigma \vdash (\ell^L : _ \,; _) :: Type \otimes Row(\ell.L) \Rightarrow Row(L) & \ell \in \mathcal{L}, L \in \mathcal{P}_f(\mathcal{L} \setminus \{\ell\})
\end{array}$$

We define *projection symbols* to be all symbols $(\ell^L : \_ \,; \_)$. We write $\mathcal{D}$ for the new set of symbols.

Let E be the set of axioms composed of:

- Left commutativity axioms. For any labels a and b and any finite subset of labels L that do not contain them,

$$a^L : \alpha \; ; \left(b^{a.L} : \beta \; ; \gamma\right) = b^L : \beta \; ; \left(a^{b.L} : \alpha \; ; \gamma\right) \qquad (a \triangleright b, L)$$

- Distributivity axioms. For any symbol f, any label a and any finite subset of labels L that do not contain a,

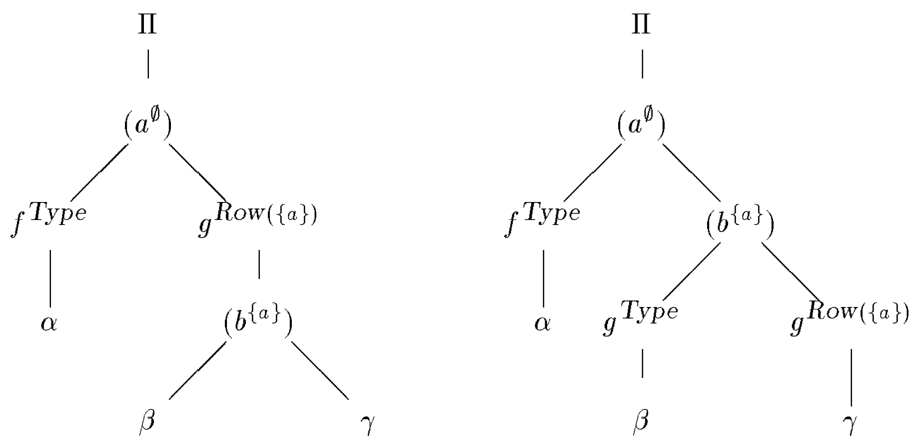
$$f^{Row(L)}(a^L : \alpha_1 \; ; \beta_1, \ldots a^L : \alpha_p \; ; \beta_p) = a^L : f^{Type}(\alpha_1, \ldots \alpha_p) \; ; f^{Row(a.L)}(\beta_1, \ldots \beta_p) \qquad (f \triangleright a, L)$$

All axioms are collapse-free, regular and linear.

Let $\mathcal{V}$ be a denumerable set of variables with infinitely many variables of every sort.

Definition 7 The algebra of record terms (also called the record algebra) is the equational theory $\mathcal{T}(\Sigma, \mathcal{V})/E$. □

Below are two E-equal record terms:



Theorem 5 *The presentation E is syntactic.*

Proof: Let $\mathcal{T}^n$ be the subset of terms that use only the sorts $Type$ or $Row(L)$, where $Card(L)$ is at most n. The sequence of these sets is increasing and its union is $\mathcal{T}$. Thus it is enough to show $Synt(\mathcal{T}_n)$ for any integer n. Let n be an integer. We show $Synt(\mathcal{T}^n)$ using theorem 1.

Let Θ_n be the usual size (sum of weights of symbols) where symbols are weighted as follows. Symbols $f^{Row(L)}$ of arity q have a weigh of $2 * (n - Card\, L) + q$. Symbols f^{Type} weigh their arity q augmented by 1, and all other symbols weight 1. The size of a term is strictly larger than the size of any of its subterms; the size is constant on E-equality classes. Thus it defines a compatible ordering in $\mathcal{T}^n$ by $\tau < \sigma$ if $\Theta_n(\tau) < \Theta_n(\sigma)$.

The condition (h_1) is always satisfied: left commutativity axioms are of depth greater than two, and so are distributivity axioms for symbols of non zero arity. Left distributivity axioms for constant symbols f are such that the equality relations $\xrightarrow[f \triangleright a, L/\epsilon]{} \xrightarrow[\bullet\bullet\infty]{}$ are empty.

The condition (h_2) is

$$\xrightarrow[\epsilon]{} \left(\xrightarrow[k]{\delta}\right)^{k \in \mathcal{D}} \xrightarrow[\epsilon]{} \subset_H \xrightarrow[\bullet\infty]{*} \xrightarrow[\epsilon]{\delta} \xrightarrow[\bullet\infty]{*}$$

We show it for all instances of the premise:

$$\xrightarrow[q/\epsilon]{} \left(\xrightarrow[s_k/k]{\delta}\right)^{k \in \mathcal{D}} \xrightarrow[t/\epsilon]{}$$

We write $\mathcal{Z}$ for the intermediate relation

$$\left(\xrightarrow[s_k/k]{\delta}\right)^{k\in\mathcal{D}}$$

Case $\mathcal{Z}$ is empty: The axioms q and t must be inverse and annihilate each other, that is, the equality relation is included in the identity relation Δ.

Case q is $(f \triangleright a, L)$: If f were of arity zero, $\mathcal{S}$ would be empty (first case). The occurrence 1 in s is not possible, and s_2 must be another axiom $(f \triangleright b, a.L)$, where b is distinct from a. Then t must be the axiom $(a \triangleright b, L)$. We conclude with:

$$\xrightarrow[f \triangleright a, L/\epsilon]{}\xrightarrow[f \triangleright b, a.L/2]{}\xrightarrow[a \triangleright b, L/\epsilon]{} \subset \xrightarrow[a \triangleright b, L/i]{i \in [1, \varrho(f)]}\xrightarrow[f \triangleright b, L/\epsilon]{}\xrightarrow[f \triangleright a, b.L/2]{}$$

Case q is $(a \triangleright f, L)$: Then t must be $(f \triangleright b, L)$. If a and b were equal, $\mathcal{Z}$ would be the identity (first case). At least one application of an axiom $(a \triangleright b, L)$ at each occurrence i between 1 and the arity of f is needed to change the symbols $(a^L : \_ ; \_)$ into $(b^L : \_ ; \_)$. We conclude with:

$$\xrightarrow[a \triangleright f, L/\epsilon]{}\xrightarrow[a \triangleright b, L/i]{i \in [1, \varrho(f)]}\xrightarrow[f \triangleright b, L/\epsilon]{} \subset \xrightarrow[f \triangleright b, L/2]{}\xrightarrow[a \triangleright b, L/\epsilon]{}\xrightarrow[a \triangleright f, L/2]{}$$

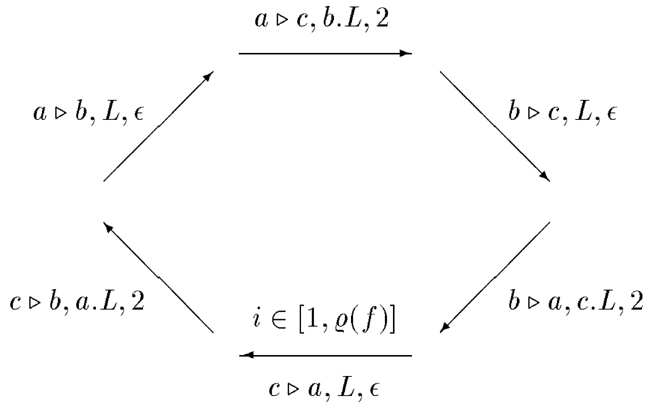
Case q is $(a \triangleright b, L)$: If t is a distributivity axiom, then by symmetry we fall into one of the preceding case; otherwise it is an axiom $(c \triangleright b, L)$. Since 1 is a variable occurrence of q, the rule $\xrightarrow[s_1,1]{}$ commutes with q, and we ignore this case. If c and a were equal, the sequence $\mathcal{Z}$ would be the identity (first case). So t must be the axiom $(a \triangleright c, b.L)$. We conclude with:

$$\xrightarrow[a \triangleright b, L/\epsilon]{}\xrightarrow[a \triangleright c, b.L/2]{}\xrightarrow[b \triangleright c, L/\epsilon]{} \subset \xrightarrow[b \triangleright c, a.L/2]{}\xrightarrow[a \triangleright c, L/\epsilon]{}\xrightarrow[a \triangleright b, c.L/2]{}$$

Note that the two first inclusions can be deduced from the following cycle (any composition of six of these rules is a subset of the identity relation):

$$\begin{tikzcd}
& {} \arrow[r, "a \triangleright b, L, \epsilon"] & {} \arrow[dr, "a \triangleright f, b.L, 2"] & \\
{} \arrow[ur, "f \triangleright b, a.L, 2"] & & & {} \arrow[dl, "b \triangleright f, L, \epsilon"] \\
& {} \arrow[ul, "f \triangleright a, L, \epsilon"] & {} \arrow[l, "i \in [1, \varrho(f)]", "b \triangleright a, L, i"'] &
\end{tikzcd}$$

The last inclusion rule forms the cycle:



■

Since the theory is syntactic, we automatically derive the mutation rule in the record term algebra:

$$\frac{f^{Row(L)}(\sigma_i)_{i\in[1,p]} \doteq a^L : \alpha \; ; \; \beta \doteq e}{\exists\, (\alpha_i, \beta_i)_{i\in[1,p]} \cdot \bigwedge \begin{cases} a^L : \alpha \; ; \; \beta \doteq e \\ \alpha \doteq f^{Type}(\alpha_i)_{i\in[1,p]} \\ \beta \doteq f^{Row(a.L)}(\beta_i)_{i\in[1,p]} \\ \sigma_i \doteq a^L : \alpha_i \; ; \; \beta_i \qquad \text{for} \qquad i \in [1,p] \end{cases}} \rightsquigarrow \quad \textsc{Mutate}(a \triangleright f)$$

$$\frac{a^L : \tau \; ; \; \sigma \doteq b^L : \alpha \; ; \; \beta \doteq e}{\exists\, \gamma \cdot \bigwedge \begin{cases} b^L : \alpha \; ; \; \beta \doteq e \\ \sigma \doteq b^{a.L} : \alpha \; ; \; \gamma \\ \beta \doteq a^{b.L} : \tau \; ; \; \gamma \end{cases}} \rightsquigarrow \qquad \textsc{Mutate}(a \triangleright b)$$

Theorem 6 *Unification in the record algebra is decidable and unitary unifying.*

Proof: The theory is strict (we exhibited a compatible ordering). Therefore, the rules for mutation, decomposition, collision, fusion and generalization applied any order a complete semi-algorithm for unification. The algorithm is unitary since mutation does not introduce any disjunction. It remains to prove termination.

All transformations are stable in the sets $\mathcal{T}^n$ and decrease in the following lexicographic ordering:

- the size Θ_n, that is the sum of the sizes of terms (decomposition and mutation),
- the number of multi-equations (fusion),
- the sum of heights of terms (generalization).

This guarantees the termination of the process. ■

3.3 Raw terms

Let $\mathcal{C}'$ be the set of symbols $\mathcal{C}$ extended with a symbol Π and the collection of symbols $(a : \_ ; \_)_{a \in \mathcal{L}}$. Terms of the algebra $\mathcal{T}(\mathcal{C}', \mathcal{V})$ are called raw terms. To any record term, we associate the raw term obtained by erasing all superscripts of symbols. Conversely, for any raw term τ' and any sort ι, there is at most one record term whose erasure is τ'. Thus any record term τ of sort ι is completely defined by its erasure τ' and the sort ι. In the rest of the paper we will mostly use this notation and often drop the sort whenever it is implicit from the context.

Projection symbols associate to the right; that is, $(a : \tau ; b : \sigma ; \rho)$ stands for $(a : \tau; b : \sigma ; \rho)$. In formulas we sometimes write $\big((a_i : \tau_i)_{i \in [1,p]} ; \sigma\big)$ for $(a_1 : \tau_1 ; \ldots a_p : \tau_p ; \sigma)$.

3.4 Examples of Record terms

When record terms are used as types of records in ML, the types of fields must first say whether the field is absent or present and in the last case whether it is an arrow type or some other structured type. For instance, a type could be $\alpha \to \Pi(a : \mathit{pre}\,(\alpha \to \alpha) ; \beta)$. However, types that would tell their structure before telling that the field is defined must be forbidden: $\alpha \to \Pi a : \alpha \to \alpha ; \beta$ is should not be a correct type. These constrains are, of course, realized using sorts. The properties of the section 1 allow record terms to be restricted by a signature compatible with its equations, and still use the same unification algorithm.

We give two examples of restricted record terms used as types in ML with records. The first instance distinguishes a constant symbol abs and a unary symbol pre in $\mathcal{C}$. The signature Σ' on the two sorts *type* and *field* is:

$$\begin{array}{ll}
\Sigma' \vdash \Pi :: \mathit{field} \Rightarrow \mathit{type} & \\
\Sigma' \vdash \mathit{abs}^{\,\iota} :: \mathit{field} & \iota \in \mathcal{K} \\
\Sigma' \vdash \mathit{pre} :: \mathit{type} \Rightarrow \mathit{field} & \\
\Sigma' \vdash f^{\,\mathit{Type}} :: \mathit{type}^{\,\varrho(f)} \Rightarrow \mathit{type} & f \in \mathcal{C} \setminus \{\mathit{abs}, \mathit{pre}\} \\
\Sigma' \vdash (\ell^L : _ ; _) :: \mathit{field} \otimes \mathit{field} \Rightarrow \mathit{field} & \ell \in \mathcal{L}, L \in \mathcal{P}_f(\mathcal{L} \setminus \{\ell\})
\end{array}$$

The signature Σ' is compatible with the equations of the record algebra. We define *simple record terms* as record terms that are well sorted for Σ'. They have a very simple record structure. Terms of the sort $\mathit{Row}(L)$ are either of depth 0 (reduced to a variable or a symbol) or are of the form $(a : \tau ; \rho)$. By induction, they are always of the form

$$(a_1 : \tau_1 ; \ldots a_p : \tau_p ; \sigma)$$

where σ is either abs or a variable, including the case where p is zero and the term is reduced to σ.

The generality of record term algebras is better justified by *complex records terms*. The problem with simple record terms is the inability to merge two records defined on different fields. For instance, the two record types $\Pi(a : \mathit{pre}\,(\alpha) ; b : \mathit{pre}\,(\sigma) ; \mathit{abs}\,)$ and $\Pi(a : \mathit{pre}\,(\tau); \mathit{abs}\,)$ cannot be unified, since on field b this would require $\mathit{pre}\,(\sigma)$ be unifiable with abs. A solution is to separate the access information from the structure information in fields. The two records could be typed with $\Pi(a : \mathit{pre}\,.\alpha ; b : \mathit{pre}\,.\sigma ; \mathit{abs}\,.\beta)$ and $\Pi(a : \mathit{pre}\,.\tau ; \mathit{abs}\,.\beta)$. They do not yet unify. But if we write instead $\Pi(a : \gamma_a.\alpha ; b : \gamma_b.\sigma ; \mathit{abs}\,.\beta)$ and $\Pi(a : \gamma'_a.\tau ; \mathit{abs}\,.\beta')$, then they are unified by the substitution:

$$\gamma_a \mapsto \gamma'_a \qquad \alpha \mapsto \tau \qquad \gamma_b \mapsto \mathit{abs} \qquad \beta' \mapsto b : \beta_b ; \beta$$

Generic record terms are well sorted records terms for the following signature Σ''. Distinguishing two constant symbols pre and abs and a binary symbol "." in $\mathcal{C}$, the signature Σ''

is defined on the three sorts *type* , *flag* , and *field* by:

$$
\begin{array}{ll}
\Sigma'' \vdash \Pi :: \mathit{field} \Rightarrow \mathit{type} & \\
\Sigma'' \vdash \mathit{abs}^{\,\iota} :: \mathit{flag} & \iota \in \mathcal{K} \\
\Sigma'' \vdash \mathit{pre}^{\,\iota} :: \mathit{flag} & \iota \in \mathcal{K} \\
\Sigma'' \vdash .^{\iota} :: \mathit{flag} \otimes \mathit{type} \Rightarrow \mathit{field} & \iota \in \mathcal{K} \\
\Sigma'' \vdash f^{\mathit{Type}} :: \mathit{type}^{\,\varrho(f)} \Rightarrow \mathit{type} & f \in \mathcal{C} \setminus \{\mathit{abs}\,, \mathit{pre}\,, .\} \\
\Sigma'' \vdash (\ell^{L} : \_ \; ; \_) :: \mathit{field} \otimes \mathit{field} \Rightarrow \mathit{field} & \ell \in \mathcal{L}, L \in \mathcal{P}_f(\mathcal{L} \setminus \{\ell\})
\end{array}
$$

The signature is still compatible with the equations. Terms of the sort $Row(L)$ can now have a more complex structure such as

$$(a : \tau \; ; \rho).(b : \sigma \; ; \rho')$$

If a and b were equal, this would simplify into

$$(a : \tau.\sigma \; ; \rho.\rho')$$

We say that the latter is a canonical form but the former is not. In the next section we study canonical forms of record terms in general.

4 Approximations of terms

The complex record type $\Pi(a : \mathit{pre}\,.\tau \; ; \sigma)$ can intuitively be understood as the type of some record whose field a is present with type τ and whose presence of other fields is defined by σ. However, it is harder to understand the objects that would have the record type τ equal to $\Pi(a : \mathit{pre}\,.\tau \; ; (b : \mathit{pre} \; ; \sigma).\gamma)$. Substituting γ by $(b : \gamma' \; ; \gamma'')$ would lead to the E-equal record type $\Pi(a : \mathit{pre}\,.\tau \; ; b : \mathit{pre}\,.\gamma' \; ; \sigma.\gamma'')$, which is less general but has a more intuitive meaning. The substitution $\gamma \mapsto (b : \gamma' \; ; \gamma'')$ does not impose any more structure on the type of τ. It just "reads" information form τ in the sense that any non-variable instance of γ is necessarily an instance of $(b : \gamma' \; ; \gamma'')$.

Canonical terms are record types in which a symbol $(a : \_ \; ; \_)$ can only occur below some other symbol but $(b : \_ \; ; \_)$ or the symbol Π, and E-canonical terms are those that are E-equal to canonical terms. There are terms that are not E-canonical. For instance, the term τ above is not. We first define a class of "reversible" substitutions called expansions. Then we show that any term can be transformed by expansion into a term that is E-canonical. There exist least E-canonical expansions; however, expansions do not commute with unification. Allowing the reverse of expansions, called contractions, leads to E-canonical approximations, which commute with unification.

4.1 Expansions

Definition 8 An *elementary substitution* of W is a substitution of the following form:

- $\alpha \mapsto a : \beta \; ; \gamma$ is an *elementary expansion* of W if α is in W and β and γ are not in W.
- $\alpha \mapsto \beta$ is an *elementary renaming* of W if α is in W and β is outside of W.
- $\alpha \mapsto \beta$ is an *elementary fusion* of W if α and β are in W.
- $\alpha \mapsto f(\beta_i)_{i \in [1,p]}$ is an *elementary structuration* of W if α is in W and β_i are all outside of W.

A *perfect composition* of W is any composition $\mu_1 \circ \ldots \mu_p$ such that there exists a sequence $(W_i)_{i \in [0,p]}$ of sets of variables satisfying:

1. $W_0 = W$;
2. μ_i is an elementary substitution of W_i;
3. W_i is equal to $\mathcal{V}(\mu(W_{i-1}))$.

An *expansion* is a perfect composition of elementary expansions. An *α-expansion* is a perfect composition of elementary expansions and renamings. □

Notation

We write expansions with letters φ, ψ, and χ. We indicate elementary substitutions by a “” accent: $\hat{\varphi}$, $\hat{\mu}$. The notation

$$\mu : V \longrightarrow T$$

is not well adapted to describe perfect compositions. We write

$$V \xrightarrow{\mu} W$$

for a perfect composition μ of V such that W is the set of variables of the image of V. We can compose them as follows:

$$V \xrightarrow{\mu} W \xrightarrow{\nu} W'$$

We also draw diagrams with the convention that continuous lines are universally quantified while dashed lines are existentially quantified. For instance

$$\begin{tikzcd} V \arrow[r, "\mu"] \arrow[d, "\varphi"'] & W \arrow[d, dotted, "\psi"] \\ W' \arrow[r, dotted, "\nu"'] & \cdot \end{tikzcd}$$

is read “*For any substitution μ from V to W and any expansion from V to W', there exist a substitution ν of W' and an expansion ψ of W such that $\nu \circ \varphi$ and $\psi \circ \mu$ are E-equal.*” All diagrams commutes modulo E-equality, except if explicitly mentioned otherwise.

Lemma 8 *Any substitution whose domain is in W can be written as a perfect composition of elementary substitutions of W.*

Proof: The lemma is first shown for the substitution of a variable by a term, by induction on the size of the term. The general case is then shown by induction on the size of the domain of the substitution. ■

Lemma 9 *A substitution μ is an expansion of W if and only if:*

- *all symbols of the image of μ are projection symbols;*
- *all variables of the image of μ are outside of W; and*
- *all terms of $\mu(W)$ are linear (a variable does notoccur twice) and disjoint (no variable is shared between two terms).*

Concretely, an expansion φ is sum $\oplus_{\alpha \in dom(\varphi)}(\varphi_\alpha)$ where φ_α are

$$\alpha \mapsto a_1 : \alpha_1 \ ; \ \ldots a_{p_\alpha} : \alpha_{p_\alpha} \ ; \ \beta_\alpha$$

and expansions φ_α are pairwise disjoint.

Proof: Any such substitution is trivially an expansion. Conversely, the set of such substitutions contains elementary expansions and is closed by composition with elementary expansions. ■

Note The set of expansions is closed by E-equality. Proving the E-equality of two expansions can always be done using left commutativity axioms only.

Lemma 10 *If θ is a renaming of W and φ is an expansion of $\theta(W)$, then there exists an expansion ψ of W and a renaming of $\psi(W)$ such that $\theta \circ \varphi$ and $\eta \circ \psi$ are equal, and conversely, inverting the roles of renamings and expansions.*

$$\begin{tikzcd} W \arrow[r, "\theta"] \arrow[d, dotted, "\psi"'] & \cdot \arrow[d, "\varphi"] \\ \cdot \arrow[r, dotted, "\eta"'] & \cdot \end{tikzcd} \qquad \begin{tikzcd} W \arrow[r, dotted, "\eta"] \arrow[d, "\varphi"'] & \cdot \arrow[d, dotted, "\psi"] \\ \cdot \arrow[r, "\theta"'] & \cdot \end{tikzcd}$$

Proof: For the first diagram, the renaming θ can be decomposed into the sum of two renamings θ' equal to $\theta \restriction W'$ and θ'' equal to $\theta \setminus W'$ where W' is $\theta^{-1}(dom(\varphi))$. Take $\theta'' + \omega^{-1}$ for η, and $\omega \circ \varphi \circ \theta'$ for ψ where ω renames variables of $im(\varphi)$ away from all other variable involved.

For the second diagram, the renaming θ can be decomposed into the sum of two renamings θ' equal to $\theta \restriction W'$ and θ'' equal to $\theta \setminus W'$, where W' is $im(\varphi)$. Take $(\theta' \circ \varphi) + \omega^{-1}$ for ψ and $\omega \circ \theta''$ for η, where ω renames variables of $im(\omega'')$ away form all other variables involved.

In both cases, it is easy to prove that ψ is an expansion, and so the diagram commutes, as required. ■

Lemma 11 *If a perfect composition of elementary substitutions is an α-expansion, then all elementary substitutions are renamings or expansions.*

Proof: Let μ be the perfect composition of W_0.

$$W_0 \xrightarrow{\hat{\mu}_1} W_1 \quad \cdots \quad \xrightarrow{\hat{\mu}_p} W_p$$

If the sequence contains a structuration $\alpha \mapsto f(\alpha_i)_{i \in [1,p]}$, then $im(\mu)$ contains the symbol f. Otherwise it contains a fusion $(\alpha \mapsto \beta)$, where α and β are in some W_i. If α and β are the images of distinct variables α' and β' of W_0, then the images by μ of α' and β' will not be disjoint. Otherwise, α and β are in the image of a common variable α' and $\mu_1 \circ \ldots \mu_i$ will not be linear. In all cases, μ cannot be an α-expansion. ■

Corollary 12 *Two substitutions that perfectly compose into an α-expansion are α-expansions.*

Lemma 13 *For any set of variables W, any elementary substitution $\hat{\mu}$ of W, and any expansion φ of W, there exists an expansion ψ of $\hat{\mu}(W)$ and a substitution ν such that $\psi \circ \hat{\mu}$ and $\nu \circ \varphi$ are E-equal on W, that is, the diagram on the right commutes modulo E-equality.*

$$\begin{tikzcd} W \arrow[r, "\hat{\mu}"] \arrow[d, "\varphi"'] & \cdot \arrow[d, dotted, "\psi"] \\ \cdot \arrow[r, dotted, "\nu"'] & \cdot \end{tikzcd}$$

<u>Proof</u>: It is enough to show the lemma for substitutions with disjoint images $\varphi(W)$ and $\hat{\mu}(W)$. The lemma then follows by introducing a renaming θ of the image of φ outside of the image of $\hat{\mu}$, applying the lemma to $\theta \circ \varphi$ instead of φ, and replacing the resulting substitution ν by $\nu \circ \theta^{-1}$. Similarly, the images can be assumed disjoint from the domains without loss of generality.

We first show the lemma for an expansion φ of the form

$$\alpha \mapsto a_1 : \alpha_1 \;;\; \ldots a_p : \alpha_p \;;\; \alpha_0$$

If the domain of $\hat{\mu}$ is not reduced to α, then the two substitutions commute, and ν is $\hat{\mu}$, and ψ is φ. Otherwise we reason by cases on $\hat{\mu}$:

- If $\hat{\mu}$ is a renaming $(\alpha \mapsto \beta)$, then $\varphi \circ \hat{\mu}$ for ψ and the identity for ν.
- If $\hat{\mu}$ is a fusion $(\alpha \mapsto \beta)$, then take $\beta \mapsto \varphi(\alpha)$ for both ν and ψ. Otherwise the image of $\hat{\mu}$ is outside of W.
- If $\hat{\mu}$ is an expansion $(\alpha \mapsto a_j : \beta_j ; \beta_0)$ for j in $[1, p_i]$, take the renaming $\beta_j \mapsto \alpha_j$ for ν and the expansion $\beta \mapsto (a_k : \alpha_k)_{k \in K} \;;\; \alpha_0$ where K is $[1, p_i] \setminus \{j\}$ for ψ.
- If $\hat{\mu}$ is an expansion $(\alpha \mapsto b : \beta_b ; \beta_0)$, then take $(\beta_0 \mapsto (a_k : \alpha_k)_{k \in [1,p_i]} \;;\; \gamma_0)$ for ψ and the expansion $(\alpha_0 \mapsto b : \beta_b ; \gamma_0)$ for ν, where γ_0 is a variable distinct from all others.
- If $\hat{\mu}$ is a structuration $(\alpha \mapsto f(\beta_j)_{j \in [1,q]})$, then take the expansion

  $$\Big(\beta_j \mapsto (a_k : \gamma_k)_{k \in [1,p_i]} \;;\; \gamma_0\Big)_{j \in [1,q]}$$

  for ψ and the structuration

  $$\Big(\alpha_k \mapsto f(\gamma_k)_{i \in [1,p]}\Big)_{k \in [0,p_i]}$$

  for ν, where all variables γ_k are distinct and distinct from all others.

A general expansion φ is the disjoint sum $\oplus_{i \in [1,p]}(\varphi_i)$ of simple expansions. If it is not disjoint from $\hat{\mu}$, there exists one expansion (φ_1 for instance) that has the same domain as $\hat{\mu}$. Applying the lemma with φ_1 gives ψ_1 and ν. Take $\psi_1 + \oplus_{i \in [2,p]}(\varphi_i)$ for ψ. ∎

Note If $\hat{\mu}$ is an expansion, then $\hat{\nu}$ is an α-expansion.

Proposition 14 *For any substitution μ of W and any expansion φ of W, there exists an expansion ψ of $\mu(W)$ and a substitution ν of $\varphi(W)$ such that $\psi \circ \mu$ and $\nu \circ \varphi$ are E-equal on W, that is, the diagram on the right commutes modulo E-equality.*

$$\begin{tikzcd} W \arrow[r, dashed, "\mu"] \arrow[d, "\varphi"'] & \cdot \arrow[d, dotted, "\psi"] \\ \cdot \arrow[r, dotted, "\nu"'] & \cdot \end{tikzcd}$$

<u>Proof</u>: By induction on the number of elementary substitutions that compose μ, we obtain the following diagram:

$$\begin{tikzcd} W \arrow[r, dashed, "\hat{\mu}_1"] \arrow[d, "\varphi_0"'] & \cdot \arrow[r, phantom, "\cdots\cdots"] \arrow[d, dotted, "\varphi_1"] & \cdot \arrow[r, "\hat{\mu}_n"] \arrow[d, dotted, "\varphi_{n-1}"] & \cdot \arrow[d, dotted, "\varphi_n"] \\ \cdot \arrow[r, dotted, "\nu_1"'] & \cdot \arrow[r, phantom, "\cdots\cdots"] & \cdot \arrow[r, dotted, "\nu_n"'] & \cdot \end{tikzcd}$$

∎

Corollary 15 *If a unificand U admits a solution μ outside of $\mathcal{V}(U)$, and φ is an expansion of U, then there exists a solution ν of $\varphi(U)$ and an expansion ψ of $\mu(U)$ such that $\psi \circ \mu$ and $\nu \circ \varphi$ are E-equal.*

Corollary 16 *For any expansions φ and ψ of W, there exist two expansions φ' of $\varphi(W)$ and ψ' of $\psi(W)$ and a renaming θ such that $\theta \circ \varphi' \circ \varphi$ and $\psi' \circ \psi$ are E-equal on W, that is, the adjoining diagram commutes modulo E-equality.*

$$\begin{tikzcd}
W \arrow[rr, "\psi"] \arrow[d, "\varphi"'] & & \cdot \arrow[d, dotted, "\psi'"] \\
\cdot \arrow[r, dotted, "\varphi'"'] & \cdot \arrow[r, dotted, "\theta"'] & \cdot
\end{tikzcd}$$

<u>Proof</u>: By proposition 14, where μ is an expansion, and then by corollary 12, ν must be an α-expansion. ■

Theorem 7 *Let U be a unificand and φ an expansion of U into V. If μ and ν are principal unifiers of U and V, then there exists an α-expansion $\eta \circ \psi$ of $\mu(U)$ into $\nu(V)$ such that $\nu \circ \varphi$ and $\eta \circ \psi \circ \mu$ are equal, that is, the adjoining diagram commutes modulo E-equality.*

$$\begin{tikzcd}
U \arrow[r, dashed, "\mu"] \arrow[d, "\varphi"'] & \cdot \arrow[d, dotted, "\eta \circ \psi"] \\
V \arrow[r, dashed, "\nu"'] & \cdot
\end{tikzcd}$$

<u>Proof</u>: The proof is sketched in the adjoining diagram, taken modulo E-equality. The existence of ξ and ψ follows from corollary 15. Since $\nu \circ \varphi$ is a solution of U and μ is the principal solution of U, the substitution ξ' must exist. Finally, the existence of ξ'' is a consequence of the fact that ξ is a solution of V, since $\xi(V)$ is equal to $\psi \circ \nu(U)$, which is satisfied, while ν is a principal solution of V. Since $\xi'' \circ \xi'$ is an expansion, ξ'' and, in particular ξ'' are α-expansions. ■

$$\begin{tikzcd}
U \arrow[rr, "\mu"] \arrow[dd, "\varphi"'] \arrow[dr, phantom, "(1)"] & & \cdot \arrow[dl, dotted, "\psi"'] \arrow[dd, dotted, "\xi'"] \\
& \cdot \arrow[r, phantom, "(2)"] \arrow[d, phantom, "(3)"] & \\
V \arrow[ur, dotted, "\xi"] \arrow[rr, "\nu"'] & & \cdot \arrow[ul, dotted, "\xi''"]
\end{tikzcd}$$

4.2 Canonical terms

In this section we define canonical terms and show that any term can be expanded into a canonical term. We assume we are given an ordering on $\mathcal{L}$ that extends naturally to projection symbols.

Definition 9 A record term τ is *canonical* if projection symbols can only occur below a symbol Π or some other smaller projection symbol.

$$\forall ux \in dom(\tau),\ \mathrm{Top}\,(\tau_{/ux}) = (a : \_ \ ; \_) \Longrightarrow \mathrm{Top}\,(\tau_{/u}) \in \{\Pi\} \cup \{(a : \_ \ ; \_) \mid a < b\}$$

A term is *E-canonical* if it is E-equal to a canonical term.

A *canonical expansion* of τ is a canonical term obtained by an expansion of τ. □

For instance, the term $(a : \Pi(a : \alpha \ ; \alpha') \ ; b : \beta' \ ; \gamma)$ is canonical, but the term

$$f((a : \alpha \ ; \alpha')\,, (a : \beta \ ; b : \beta' \ ; \gamma))$$

is not. The term $(a : \tau \ ; \rho).(b : \sigma \ ; \rho')$ of section 3.4 is canonical.

Notation

In general if $\mathcal{Q}$ is a set of terms (respectively a set of symbols) and ι is a sort, we write $\mathcal{Q}^\iota$ for the subset of $\mathcal{Q}$ of terms (respectively symbols) of the sort ι.

If L is a finite subset of labels, we write $\mathcal{V}_L$ for the union $\mathcal{V}^{Type} \cup \mathcal{V}^{Row(L)}$ and $\mathcal{V}_{\subset L}$ for the union of all $\mathcal{V}_K$ for all subsets K of L, $\mathcal{C}_{\subset L}$ for the union of $\mathcal{C}^{Type}$ and all $\mathcal{C}^{Row(K)}$ for all subsets K of L, and $\Sigma_{\subset L}$ for the restriction of the signature $\Sigma \restriction \mathcal{C}_{\subset L}$. We write $\mathcal{T}_L$ for the algebra $\mathcal{T}_L(\mathcal{V}_L, \Sigma_{\subset L})$ and $\mathcal{T}_{\subset L}$ the algebra $\mathcal{T}(\mathcal{V}_{\subset L}, \Sigma_{\subset L})$.

For instance, the term τ equal to $\{a : \alpha \ ; \alpha'\}$ is in $\mathcal{T}_{\{a\}}$. The term σ equal to $(b : \beta \ ; \beta')$ is in $\mathcal{T}'_{\{b\}}$, and the term $\tau \rightarrow \sigma$ is in $\mathcal{T}_{\subset\{a,b\}}$.

Definition 10 For any finite set of labels L we write $\vec{E}_L$ for the rewriting system obtained by orienting the equations $(a \triangleright b, K)$ if a and b are ordered and $(f \triangleright a, K)$ for all subsets K of L. □

Lemma 17 *For any finite set of labels L, the rewriting system $\vec{E}_L$ is stable and terminates in the set $\mathcal{T}_{\subset L}$.*

Proof: We extend the total order on projection symbols to a partial order on all symbols by placing projection symbols before all other symbols. An ordered element in τ is a pair (u, v) of two non-variable occurrences of τ such that u is a prefix of v and the symbol at the occurrence u is smaller than the symbol at occurrence v. Any step increases the number of ordered elements. The size $\Theta_{Card\,(L)}$ is constant, which provides a bound on the number of occurrences of symbols of the term, and thus on the number of pairs of non-variable prefix occurrences. ■

Note The rewriting system $\vec{E}_\infty$ would not terminate, since constant symbols could be rewritten forever in terms of $\mathcal{T}_L$ for increasing L.

Lemma 18 *For any finite set of labels L, all terms of $\vec{E}_L(\mathcal{T}_L)$ are canonical.*

Proof: A term of $\vec{E}_L(\mathcal{T}_L)$ reduces by $\vec{E}_L$ to some term where non-projection symbols can only have sorts $Row(L)$. Such terms are trivially canonical. ■

Corollary 19 *For any finite set of labels L, all terms of $\mathcal{T}_L$ are E-canonical.*

This gives us a means of computing the canonical form of a term in $\mathcal{T}_L$.

Definition 11 Let τ be a term, and L be the smallest K such that τ is in $\mathcal{T}_{\subset K}$. If φ is an expansion of row variables of τ into $\mathcal{T}_L$, then $\varphi(\tau)$ is called a *canonical expansion* of τ. Let μ be a substitution and L the smallest K such that $im\,(\mu)$ is in $\mathcal{T}_{\subset K}$. If φ is an expansion of variables of $im\,(\mu)$ into $\mathcal{T}_L$, then $\varphi \circ \mu$ is called a *canonical expansion* of μ. □

It is clear that canonical expansions are E-canonical. They are defined modulo a renaming of variables that are introduced by the expansion. In general, they are not the smallest E-canonical expansions, even for a single term τ of $\mathcal{T}'$. For instance, if τ is in $\mathcal{T}_L$, α is in $\mathcal{V}_L(\tau)$, and the label a is not in L, then the canonical expansion of $(\alpha \mapsto a : f \ ; f)(\tau)$ is in $\mathcal{T}_{a.L}$, but there the E-canonical term $(\alpha \mapsto f)(\tau)$ in $\mathcal{T}_L$ obtained by the empty expansion.

The existence of E-canonical forms is not enough. As computed above, canonical forms may be very large. Are there smallest E-canonical forms? Do they commute with unification? That is, given a unificand W, an α-expansion φ that maps W to a the smallest canonical form W', and two principal unifiers μ and μ' of W and W', are the smallest α-expansions $\psi(\mu(W))$ and $sx1'(\mu'(W'))$ E-equal modulo renaming?

$$
\begin{tikzcd}
W \arrow[r, "\mu"] \arrow[d, "\varphi"'] & \cdot \arrow[r, "\psi"] & \cdot \arrow[d, dotted, "\theta?"] \\
W' \arrow[r, "\nu"] & \cdot \arrow[r, "\chi"] & \cdot
\end{tikzcd}
$$

The following example shows that this cannot be the case.

Example

$$
\begin{tikzcd}
f((a : g \; ; \; \alpha), \beta) \doteq f(g, \beta) \arrow[r, "(\alpha \mapsto g)", "\mathcal{T}'_{\emptyset}"'] \arrow[d, "(\beta \mapsto (a : \gamma \; ; \; \delta))"', "\mathcal{T}'_{\{a\}}"] & f(g, \beta) \doteq f(g, \beta) \arrow[d, dotted, "\emptyset"] \\
f((a : g \; ; \; \alpha), (a : \gamma \; ; \; \delta)) \doteq f(g, (a : \gamma \; ; \; \delta)) \arrow[r, "(\alpha \mapsto g)", "\mathcal{T}'_{\{a\}}"'] & f(g, (a : \gamma \; ; \; \delta)) \doteq f(g, (a : \gamma \; ; \; \delta))
\end{tikzcd}
$$

This counterexample removed any hope of providing the set of E-canonical forms with a "sup" operation that extends the "sup" in $\mathcal{T}$. The preceding diagrams can be easily extended to show that this operation would not be associative. The problem comes from axioms $(f \triangleright a, K)$ for constant symbols f. One solution to is to allow contraction — the reverse of expansion — while computing canonical forms. These more general canonical forms are called approximations and are studied in the rest of this section.

Definition 12 An *approximation* of a substitution μ (respectively of a term τ) is an E-canonical substitution ν (respectively an E-canonical term σ) such that there exist two expansions φ and ψ such that $\psi \circ \nu = \varphi \circ \mu$ (respectively $\varphi(\tau) = \psi(\sigma)$).

Approximations are called *contractions* when φ is the identity. □

We first show that computing approximations in $\mathcal{T}$ can be reduced to computing approximations in $\mathcal{T}'$.

4.3 Approximations in $\mathcal{T}$

Recall that a substitution μ is *p-potent* if μ^p is equal to μ. It is potent if it is potent for some p. Its composition for high enough powers is written μ^∞. A 1-potent substitution is also said *idempotent*. An idempotent substitution is characterized by having disjoint domain and image. We say that μ is potent on W if the restriction of μ to W is potent.

A substitution μ is linear in a variable α if α appears at most once in the image of at most one variable of the domain of μ. By extension a substitution is linear in a set of variables W if it is linear in all variables of W.

Definition 13 A decomposition of a substitution μ is a pair, written $\operatorname{dec} W \cdot \nu$, of a set of variables W disjoint from μ and a substitution ν potent on W such that $(\nu \upharpoonright W)^\infty \circ (\nu \setminus W)$ is equal to μ. A Π-decomposition of μ is a decomposition $\operatorname{dec} W \cdot \nu$ such that ν sends W to $\Pi(\mathcal{T}')$ and $\mathcal{V} \setminus W$ to $\mathcal{T}'$. It is *linear* if ν is linear on W. If ν is potent on W, we also write $\operatorname{dec} W \cdot \nu$ for the substitution $(\nu \upharpoonright W)^\infty \circ (\nu \setminus W)$ itself. □

For instance, the substitution $\alpha \mapsto \{a : \{b : \alpha \ ; \beta\} \ ; abs\} \rightarrow \{b : \alpha \ ; \beta\}$ can be decomposed into

$$\operatorname{dec} \gamma \cdot \wedge \begin{cases} \alpha \mapsto \gamma \rightarrow \gamma \\ \gamma \mapsto \{a : \delta \ ; abs\} \\ \delta \mapsto \{b : \alpha \ ; \beta\} \end{cases} \qquad \text{or} \qquad \operatorname{dec} \gamma, \delta, \delta' \cdot \wedge \begin{cases} \alpha \mapsto \gamma \rightarrow \delta' \\ \gamma \mapsto \{a : \delta \ ; abs\} \\ \delta \mapsto \{b : \alpha \ ; \beta\} \\ \delta' \mapsto \{b : \alpha \ ; \beta\} \end{cases}$$

and the latter decomposition is linear.

Lemma 20 *Each term has a linear* Π*-decomposition.*

Proof: We can build a linear Π-decomposition for any term by induction on the number of Π symbols. If μ has no Π symbol, then ν is μ and W is empty. Otherwise there exists a variable α in the domain of μ whose image is $\tau[\Pi(\sigma)/\beta]$, where σ has no Π symbol and the variable β is taken out of μ and appears exactly once in τ. By induction, the substitution μ' equal to $\mu \setminus \{\alpha\} + \alpha \mapsto \tau$ has a linear Π-decomposition $\operatorname{dec} W' \cdot \nu'$. Then $\operatorname{dec} W \cup \{\beta\} \cdot \nu' + \beta \mapsto \sigma$ is a linear Π-decomposition of μ. ■

The decomposition $\operatorname{dec} W \cdot \nu$ can be represented by annotating all occurrences of Π symbols in the image of μ with distinct variables of W.

Lemma 21 *If μ is has a* Π*-decomposition then it has a linear* Π*-decomposition.*

Proof: Let $\operatorname{dec} W \cdot \nu$ be a Π-decomposition of μ.

- If ν sends two variables α and β two terms τ and σ that share a variable γ of W, then $\operatorname{dec} W \cup \{\gamma'\} \cdot \nu \setminus \{\beta\} + \beta \mapsto \sigma[\gamma'/\gamma]$, where γ' is disjoint from ν, is a decomposition of μ.
- If ν sends a variable α to a term τ that contains two occurrences u and v of the same variable β in W, then $\operatorname{dec} W \cup \{\beta'\} \cdot \nu \setminus \{\alpha\} + \alpha \mapsto \sigma[\beta'/v]$ where β' is disjoint from ν is a decomposition of μ.

Each transformation increases the number of variables in W, which is bound by the number of Π symbols in μ. When no more transformations are possible, the decomposition is linear. ■

Lemma 22 *Let* $\operatorname{dec} W \cdot \nu$ *and* $\operatorname{dec} W' \cdot \nu'$ *be two* Π*-decompositions of μ, and μ'. If there exists a renaming θ of W into W' and if ν' and $\theta \circ \nu \circ \theta^{-1}$ are E-equal, then μ' and μ are E-equal.*

Proof: For a high enough p, the substitution μ' is $(\nu' \upharpoonright W')^p \circ (\nu \setminus W')$, that is

$$(\theta \circ \nu \circ \theta^{-1} \upharpoonright \theta(W))^p \circ (\theta \circ \nu \circ \theta^{-1} \setminus \theta(W)) \qquad \text{or} \qquad (\theta \circ (\nu \upharpoonright W) \circ \theta^{-1})^p \circ (\theta \circ (\nu \setminus W) \circ \theta^{-1})$$

After removing intermediate θ^{-1} and θ, we get $\theta \circ (\nu \upharpoonright W)^p \circ (\nu \setminus W) \circ \theta^{-1}$ which is $\theta \circ \mu \circ \theta^{-1}$. Since μ is disjoint from μ, it is equal to μ. ■

The converse is not true in general. But it is true if the decompositions are linear. Thus, all linear Π-decompositions are equal up to renaming of intermediate variables.

Lemma 23 *Let* $\operatorname{dec} W \cdot \nu$ *and* $\operatorname{dec} W' \cdot \nu'$ *be two linear* Π*-decompositions of μ and μ'. If μ' and μ are E-equal, then there exists a renaming θ of W into W' such that ν' and $\theta \circ \nu \circ \theta^{-1}$ are E-equal.*

Proof: We reason by induction on the number of Π-symbols in μ. If μ contains no Π symbol, then neither does μ' and both W and W' are empty; consequently ν' and ν are E-equal.

Otherwise, there exists a variable α in the domain of μ whose image by μ is $\tau[\Pi(\sigma_i)/\beta_i]$, where τ has no Π symbol and variables β_i are taken away from μ and appear exactly once in τ. The term $\mu'(\alpha)$ is of the form $\tau'[\Pi(\sigma'_i)/\beta_i]$, where τ and τ' on one hand and σ_i and σ'_i on the other hand are E-equal. The substitutions $\mu \setminus \{\alpha\} + \beta_i \mapsto \sigma_i$ and $\mu' \setminus \{\alpha\} + \beta_i \mapsto \sigma'_i$ are E-equal. We name them respectively μ_α and μ'_α. For each i there must be a variable in W whose image under ν is $\Pi(\sigma_i)$. We can assume that it is β_i without lost of generality. It is easy to check that $\operatorname{dec} W \setminus \{\beta_i\} \cdot \nu \setminus \{\alpha\}$ is a decomposition of μ_α. Similarly for each i there is a variable β'_i in W' whose image under ν' is σ'_i, and such that $\operatorname{dec} W' \setminus \{\beta'_i\} \cdot \nu \setminus \{\alpha\}$ is a decomposition of μ'_α. By the induction hypothesis, there exists a substitution θ_0 such that the substitutions $\nu \setminus \{\alpha\}$ and $\theta_0 \circ (\nu' \setminus \{\alpha\}) \circ {\theta_0}^{-1}$ are E-equal. The substitution θ equal to $\theta + (\beta \mapsto \beta_i)$ satisfies the lemma. ■

Lemma 24 *For any linear* Π*-decomposition* $\operatorname{dec} W \cdot \nu$ *of a substitution* μ*,* μ *is* E*-canonical if and only if* ν *is.*

Proof: For any substitution ν and any substitution ξ of into $\Pi(\mathcal{T}')$, then $\xi \circ \nu$ is E-canonical if and only if ν and ξ are E-canonical, since $\mathcal{S}(\xi \circ \nu)$ and $\mathcal{S}(\xi) \cup \mathcal{S}(\nu)$ are equal, where $\mathcal{S}(\mu_0)$ is the set of all pairs $(f, (a : \_ \,;\, \_))$ such that $(a : \_ \,;\, \_)$ occurs directly below some occurrence of f in μ_0. This shows the lemma for canonicity instead of E-canonicity. The general case follows from lemmas 22 and 23. ■

Definition 14 The composition of two Π-decompositions $\operatorname{dec} W_1 \cdot \nu_1$ and $\operatorname{dec} W_2 \cdot \nu_2$ such that W_1 is disjoint from ν_2 and W_2 is disjoint from ν_1 is the Π-decomposition $\operatorname{dec} W_1, W_2 \cdot \nu_1 \circ \nu_2$. □

Lemma 25 *The composition of the* Π*-decompositions of two substitutions is a* Π*-decomposition of their composition.*

Proof: We must check that for some high enough p,

$$(\mu \restriction U)^p \circ (\mu \setminus U) \circ (\nu \restriction V)^p \circ (\nu \setminus V) = (\mu \circ \nu \restriction U \cup V)^p \circ \underline{(\mu \circ \nu \setminus U \setminus V)}$$

where μ is disjoint from V and ν is disjoint from U. Since $im(\nu)$ is disjoint from U, the substitution $\mu \circ \nu \setminus U \setminus V$ is equal to $(\mu \setminus U) \circ (\nu \setminus V)$ and we are left with:

$$(\mu \restriction U)^p \circ \underline{(\mu \setminus U) \circ (\nu \restriction V)^p} = (\mu \circ \nu \restriction U \cup V)^p \circ (\mu \setminus U)$$

Since $dom(\mu \setminus U)$ is disjoint from $dom(\nu \restriction V)$, the substitution $(\mu \setminus U) \circ (\nu \restriction V)$ is equal to $(\mu \setminus U) + ((\mu \setminus U) \circ (\nu \restriction V)) \restriction V$, that is, $(\mu \circ \nu \restriction V) \circ (\mu \setminus U)$. By induction, we find that the substitution $(\mu \setminus U) \circ (\nu \restriction V)^p$ is E-equal to $(\mu \circ \nu \restriction V)^p \circ (\mu \setminus U)$. Thus, we are left with:

$$(\mu \restriction U)^p \circ (\mu \circ \nu \restriction V)^p = (\mu \circ \nu \restriction U \cup V)^p$$

which holds since $\mu \circ \nu \restriction U \cup V$ is equal to $(\mu \circ \nu \restriction V) + (\mu \restriction U)$. ■

The lemma is true even for non-linear Π-decompositions. The composition of linear Π-decompositions may not be linear. If $\operatorname{dec} W \cdot \nu$ is a linear expansion of μ and φ is an expansion disjoint form W, then $\operatorname{dec} W \cdot \varphi \circ \nu$ is a linear Π-decomposition of $\varphi \circ \mu$.

Lemma 26 *Let* $\operatorname{dec} W \cdot \nu$ *be a linear* Π*-decomposition of* μ*. A substitution* μ' *is an approximation of* μ *if and only if there exists a linear* Π*-decomposition of* $\operatorname{dec} W \cdot \nu'$ *of* μ' *such that* ν' *is an approximation of* ν*.*

Proof: Assume that there exists a Π-decomposition $\operatorname{dec} W \cdot \nu'$ of μ' such that ν' is an approximation of ν. There exist two substitutions φ and φ' such that $\varphi \circ \nu$ and $\varphi' \circ \nu'$ are E-equal. The substitution $\varphi \circ \mu$ is equal to $\varphi \circ (\nu \upharpoonright W)^p \circ (\nu \setminus W)$. By sort considerations, the substitution φ is necessarily disjoint from W, and it is potent. Thus the above substitution is also $((\varphi \circ \nu) \upharpoonright W)^p \circ ((\varphi \circ \nu) \setminus W)$. We get an E-equal substitution replacing $\varphi \circ \nu$ by $\varphi' \circ \nu'$ which is E-equal to $\varphi' \circ \mu'$ by the same reasoning.

Conversely, assume that μ' is an approximation of μ. Then there exists a linear approximation $\operatorname{dec} W \cdot \nu'$ of μ'. A similar calculus to the one above shows that substitution $\operatorname{dec} W \cdot \varphi \circ \nu$ is a linear Π-decomposition of $\varphi \circ \mu$ and similarly $\operatorname{dec} W \cdot \varphi' \circ \nu'$ is a linear Π-decomposition of $\varphi' \circ \mu'$. The E-equality of $\varphi \circ \mu$ and $\varphi' \circ \mu'$ and lemma 23 show that $\varphi \circ \nu$ and $\varphi' \circ \nu'$ are E-equal. Thus, φ' is an approximation of φ. ■

The lemma is not true if the decompositions are not linear. For instance, take

$$\operatorname{dec} \alpha_1, \beta_2 \cdot \begin{cases} \alpha \mapsto \alpha_1 \to \alpha_2 \\ \alpha_1 \mapsto \Pi\,(a : \beta \;;\; \gamma_1) \\ \alpha_2 \mapsto \Pi\,(\gamma_2) \end{cases} \qquad \text{and} \qquad \operatorname{dec} \alpha_1 \cdot \begin{cases} \alpha \mapsto \alpha_1 \to \alpha_1 \\ \alpha_1 \mapsto \Pi\,(a : \beta \;;\; \gamma_1) \end{cases}$$

for μ and μ'.

We are led to study approximations for substitutions in $\mathcal{T}' \cup \Pi(\mathcal{T}')$.

4.4 Approximations in $\mathcal{T}'$

For simplicity of exposition, we study approximations of substitutions into $\mathcal{T}'$, but all results straightforwardly extend to approximations of substitutions into $\mathcal{T}' \cup \Pi(\mathcal{T}')$.

Definition 15 The *connexe components* of a substitution μ in $\mathcal{T}'$ is the partition of the domain of μ by the smallest equivalence $\sim_\mu$ that contains all pairs $\alpha \sim_\mu \beta$ such that images of α and β by μ share at least one row variable, i.e. $\mathcal{V}(\mu(\alpha)) \cap \mathcal{V}(\mu(\beta)) \setminus \mathcal{V}^{\mathrm{Type}} \neq \emptyset$.

A substitution is said to be *connexe* if $\sim_\mu$ has only one connexe component □

The connexe components of a substitution are preserved by expansion. Let $(W_i)_{i \in [1,p]}$ be the connexe components of μ. We write μ_i the restrictions of μ to W_i. Then μ is equal to $\oplus_{i \in [1,p]} \mu_i$. If μ is idempotent, it is also the composition $\mu_1 \circ \ldots \mu_p$ in any order. If φ is an expansion of μ, then $\varphi \circ \mu$ has the same connexe components as μ and the substitution φ_i perfectly composes with μ_i, where φ_i is the restriction φ to $im\,(\mu_i)$. Conversely, if φ_i are expansions that perfectly compose with μ_i and are disjoint, then $\oplus_{i \in [1,p]} \varphi_i$ is an expansion that perfectly composes with μ.

Lemma 27 *Let μ and ν be two substitutions in $\mathcal{T}'$. The substitution ν is an approximation of μ if and only if they have the same connexe components and $\mu \upharpoonright W$ is an approximation of $\nu \upharpoonright W$ on each connexe component W.*

Proof: We first assume the existence of two expansions φ and ψ such that $\varphi \circ \mu$ and $\psi \circ \nu$ are equal and ν is E-canonical. Since expansion does not change connexe components, both μ and ν have the same connexe components $(W_i)_{i \in [1,p]}$. The substitution ν_i is of course E-canonical since the restriction of a E-canonical substitution and such that $\psi \upharpoonright \mu(W_i) \circ \mu \upharpoonright W_i$ and $\varphi \upharpoonright \nu(W_i) \circ \nu \upharpoonright (W_i)$ are E-equal for each connexe component W_i.

Conversely, we assume that μ and ν have the same components $(W_i)_{i \in [1,p]}$, and that $\nu \upharpoonright W_i$ is an approximation of $\mu \upharpoonright W_i$ on each connexe component W_i. Therefore, there exists expansions φ_i and ψ_i such that $\psi_i \circ \nu \upharpoonright W_i$ and $\varphi_i \circ \mu \upharpoonright W_i$ are E-equal. We can always choose them such that they are disjoint form each other and from the common domain of μ and ν. Then their respective sums φ and ψ perfectly composes. The substitution $\psi \circ \nu$ and $\varphi \circ \mu$ are E-equal. ■

We are left with the computation of approximations on connexe components.

4.5 Approximations on connexe components in $\mathcal{T}'$

By corollary 19, we know that substitutions in ${\mathcal{T}'}_L$ are E-canonical. Conversely, a connexe E-canonical substitution μ in $\mathcal{T}'$ is in fact always in a set ${\mathcal{T}'}_L$ for some L. Connectivity is preserved by expansion. Therefore, any approximation ν of μ is also a connexe E-canonical substitution, say ν. It can be looked for among the substitutions whose range is in ${\mathcal{T}'}_{L'}$ for some L'. When the approximation ν is a contraction of μ, there exist an expansion φ that sends variables of $\mathcal{V}^{Row(L')}$ to terms of $\mathcal{T}_L$. The set L' is a subset of L. The smallest contractions are *contractions on* a when L' is $L \setminus \{a\}$ for some label a. We show that any contraction of a connexe E-canonical substitution can be obtained by successive contractions on labels, independently of the order in which labels are chosen. The process will end with a minimal approximation.

Definition 16 If a term of $\mathcal{T}'$ is E-equal to $(a : \tau \,;\, \sigma)$ we define its *projection on* a and its *residual on* a as the terms τ and σ written $\tau_{/a}$ and $\tau_{\backslash a}$, respectively. We recursively define the *template* of a term τ, written $\tau_{/\infty}$, as $(\tau_{\backslash a})_{/\infty}$ if τ is equal to $(a : \sigma \,;\, \rho)$ and τ otherwise.

The projection (respectively residual, template) of a substitution are raw substitutions of the same domain that maps variables to the projection (respectively residual, template) of their substitution, when defined. □

Lemma 28 *Let μ be a connexe substitution into $\mathcal{T}_L$. It contracts on label a if and only if $\mu_{/a}$ is defined and is raw-isomorphic to $\mu_{/\infty}$.*

Let ϱ be a raw-renaming from $\mathcal{V}^{Row(L)}(\mathrm{im}\,(\mu))$ to $\mathcal{V}^{Row(L\setminus\{a\})}$. The contraction of μ (up to renaming and E-equality) is the substitution $\varrho^{-1} \circ (\mu_{\backslash a})$, which is equal to μ by the expansion $(a : \varrho_a \circ \varrho^{-1} \,;\, \varrho^{-1})$, where ϱ_a is a raw-renaming that maps $\mu_{/\infty}$ to $\mu_{/a}$ disjoint from μ.

Proof: Both sides are immediate. ■

Corollary 29 *Let μ be a connexe E-canonical substitution. If it contracts on label a and on label b separately, then its contraction on a contracts on b and its contraction on b contracts on a, both ways ending with the same substitution, up to renaming and E-equality.*

$$\begin{tikzcd}
Z \arrow[dr, dotted, "\xi'"'] \arrow[drr, "\mu"] \arrow[ddr, "\nu"'] & & \\
& W' \arrow[r, dotted, "\theta \circ \varphi'_b"] \arrow[d, dotted, "\psi'_a"] \arrow[dr, "\xi"] & U \arrow[d, "\varphi_a"] \\
& V \arrow[r, "\psi_b"'] & W
\end{tikzcd}$$

Proof: The proof follows immediately from the fact that $\mu_{\backslash a \backslash b}$ and $\mu_{\backslash b \backslash a}$ are defined simultaneously and then are equal. ■

Corollary 30 *Let ξ be an connexe E-canonical substitution. If μ and ν are two contractions of ξ, then there is an approximation ξ' and two expansions φ' and ψ' such that μ is E-equal to $\varphi' \circ \xi'$ and ν is E-equal to $\psi' \circ \xi'$.*

$$\begin{tikzcd}
Z \arrow[dr, dotted, "\xi'"'] \arrow[drr, "\mu"] \arrow[ddr, "\nu"'] & & \\
& W' \arrow[r, dotted, "\theta \circ \varphi'"] \arrow[d, dotted, "\psi'"] \arrow[dr, "\xi"] & U \arrow[d, "\varphi"] \\
& V \arrow[r, "\psi"'] & W
\end{tikzcd}$$

Proof: By lemma 9, any contraction is composed of contractions on labels. Then repeatedly apply the previous lemma. ■

Corollary 31 *Any connexe E-canonical substitution has a minimal contraction.*

Lemma 32 *Any connexe substitution has a minimal approximation, which is the minimal contraction of a E-canonical expansion.*

Proof: If a substitution has two approximations, there exist expansions of the two approximations that are E-equal up to renaming. Thus both expansions are contraction of a same connexe E-canonical form, and are thus E-equal up to renaming. ■

4.6 Minimal approximations in $\mathcal{T}$

Theorem 8 *Any substitution μ has a minimal approximation μ' in $\mathcal{T}$, which can be computed as follows:*

1. *Find a linear Π-decomposition $\mathrm{dec}\, W \cdot \nu$ of a μ.*
2. *Decompose μ into connexe components, $\oplus_{i \in [1,p]} \mu \upharpoonright W_i$.*
3. *For each connexe component:*
   (a) *find an E-canonical expansion ξ_i of $\mu \upharpoonright W_i$,*
   (b) *find a minimal contraction ν_i of ξ_i, by contracting on all labels for which the projection of ξ_i is raw-isomorphic to the template of ξ_i.*

Take the substitution $\mathrm{dec}\, W \cdot \oplus_{i \in [1,p]}(\nu_i)$ for μ'.

Proof: The algorithm and the theorem is a combination of lemmas 32, 27, and 26. ■

In fact in step 1, the Π-decomposition $\mathrm{dec}\, W \cdot \nu$ of μ need not be linear. In this case, the approximations of μ do not correspond to approximations of ν, but it can be shown that minimal approximations do.

Finding a Π-decomposition and the connexe components can always be done in linear time. The expensive part of the algorithm is step 3b, which looks for possible isomorphisms. This is inherent to contraction. Removing this step (that is, taking ν_i for ν_i) will compute a good approximation of μ (it is a small canonical expansion of μ). This is sufficient in practice.

4.7 Minimal approximations and unification

The following theorem shows that unification commutes with approximations. Therefore unification can be done modulo approximations.

Theorem 9 *If U' is a minimal approximation of a unificand U and μ is a principal unifier of U, then U' has a principal unifier μ' and μ and μ' have the same minimal approximation modulo renaming and E-equality.*

Proof: The following diagram commutes: the existence of μ'' follows from corollary 15. Then the existence of μ' follows from the unifier $\mu'' \circ \varphi'$ of U'. Theorem 7 is applied twice to get the existence of ψ and ψ'. The renamings can in fact be included in the principal unifiers μ' and μ''. The unificand V'' and V' may not be E-canonical. However, V and V' have an identical E-canonical form, and thus they have the same principal E-canonical approximation modulo renaming. ∎

$$\begin{tikzcd}
U \arrow[rr, "\mu"] \arrow[dd, "\varphi"'] & & V \arrow[dd, dotted, "\psi"] & \\
& U' \arrow[dl, "\varphi'"] \arrow[rr, dotted, "\mu'"] & & V' \arrow[dl, dotted, "\psi'"] \\
U'' \arrow[rr, dotted, "\mu''"] & & V'' &
\end{tikzcd}$$

There are cases where the structure of unificands ensures that the principal unifier is E-canonical. For instance, with the signature of record terms Σ' of section 3.4, all terms are always E-canonical. This is no longer the case with the signature Σ'', as shown by the examples at the beginning of this section. However, for some input unificands to the type inference algorithm, it is guaranteed that the output unificands will be in canonical form. Obviously, it is not sufficient that the input unificand is E-canonical.

Definition 17 A unificand is *separated* if it can be written (by rearranging the multi-equations) in the form $\exists W \cdot U \wedge V$ such that

1. all multi-equations of U are of the sort $Type$ and those of V are of a row sort,
2. all row subterms of U and all type subterms of V are variables and their union is W,
3. row variables of W are only variable terms in V,
4. any pair of terms of V have disjoint or equal sets of row variables, and
5. all terms are E-canonical.

A unificand is *separable* if it is equivalent to a separated unificand using only the U-Generalize rule. □

Lemma 33 (Separate) *Let $\exists W \cdot U \wedge \exists \mathcal{V}(e) \cdot (e \wedge V)$ be a separated canonical unificand and let μ be a principal E-unifier of e such that its image is outside all bound or free variables of the input unificand. Then, $\exists W \cdot \mu(U) \wedge \mu(V)$ is an equivalent unificand in separated E-canonical form, up to reordering the multi-equations.*

Proof: We first show that $\mu(e)$ is E-canonical. Each term σ of E is in some set $\mathcal{T}_{L_\sigma}$ for some set of labels L_σ. Let L be the union of all sets L_σ. It is enough to show that $\mu(e)$ is in $\mathcal{T}_L$. Let α be a variable of $\mu(e)$. For each σ, the variable α is in $\mu(\sigma)$; thus there exists a variable α_σ such that α is in $\mu(\alpha_\sigma)$. Since there is no Π symbol in e, the variable α_σ is of sort L_σ; therefore α has a sort $Row(L)$ for some L greater than L_σ. The set L contains L_σ, for any

σ. We know that there is a solution of the multi-equation in $\mathcal{T}_{\subset L}$; thus L cannot be greater than L, otherwise μ would not be principal.

Each term τ of V either shares no row variable with e or has exactly the same row variables as one term of e. In the former case the image by μ is unchanged. In the latter case, the term $\mu(\tau)$ has exactly the same row variables as $\mu(e)$, i.e., it is in $\mathcal{T}_{\subset L}$ and is thus E-canonical. The equivalence of the two unificands is obvious. ■

Corollary 34 (SEPARATE-BIS) *Let $\exists\, W, \alpha \cdot U \wedge \exists\, \mathcal{V}(e) \cdot (\alpha = e \wedge V)$ be a separated canonical unificand and let μ be a principal E-unifier of e such that its image is outside all bound or free variables of the input unificand. Then $\exists\, W \cdot \mu(U) \wedge (\alpha = \mu(e) \wedge \mu(V))$ is an equivalent unificand in separated E-canonical form.*

Proof: Generalize the unificand into $\exists\, W, \alpha \cdot U \wedge \exists\, \beta, \mathcal{V}(e) \cdot (\beta = e \wedge \alpha = \beta \wedge U)$ and apply the previous lemma. ■

Theorem 10 *A completely decomposed form of a separable solvable unificand is separable.*

Proof: Let U be the input unificand and V a completely decomposed form. Let U' be a separated unificand equivalent to U. Let V' be a separated, completely decomposed unificand obtained by the algorithm we describe below. The unificands V and V' are equivalent and completely decomposed. They are necessarily equal modulo unrestricted generalization.

To obtain an algorithm that solves separated unificands, we restrict the rule SEPARATE-BIS to the case where e contains more than one term. Applying DECOMPOSE for symbols of the sort *Type* and rules SEPARATE, SEPARATE-BIS, GENERALIZE and FUSE is stable on separated unificands and terminates. Each rule decrease in the following lexicographic order:

1. The number of symbols of the sort *Type* (decomposition).
2. The sum of heights of row terms (generalization).
3. The number of multi-equations (fusion and first transformation).
4. The number of terms in row multi-equations (last transformation).

This guarantees the termination. Stability of separation rules is proved above. Stability is obvious for all other rules. ■

For instance, with the signature of record terms Σ' of section 3.4, all terms are always E-canonical. This is no longer the case with the signature Σ'', as shown by examples at the beginning of the section. However, the type inference algorithm will only generate E-canonical unificands in disjoint form.

Of course, E-canonical unificands can be approximated by smaller unificands, but this has less interest in this case, since the completely decomposed forms are automatically E-canonical if the input unificands are always separable. The type systems for languages with records presented in [Rém93] are based on signatures Σ' and Σ'', and the unificands generated by the type system are always separable, which prevents us from computing E-canonical forms at the end of typechecking.

5 Generic record terms

In record terms, row variables can always be substituted by rows with independent fields; the expansion "duplicates" row variables. For instance, in record types of the form $\Pi(a : \tau; pre(\alpha))$, it is always possible to substitute α by $b : \alpha_b \,;\, c : \alpha_c \,;\, \alpha'$. In some cases, it is useful to express the fact that some part of a row must be shared on all fields. Those parts need to be of the sort type, and consequently we need a new symbol ∂ to inject types into rows. For instance, let σ be $\alpha \rightarrow \Pi(a : \tau \;;\; \partial(\alpha))$. The variable α cannot be replaced by a row defined on b, but σ is equal to $\alpha \rightarrow \Pi(a : \tau \;;\; b : \alpha \;;\; \partial(\alpha))$, which is realized by non-linear idempotent axioms asserting that $\partial(\tau)$ is equal to $b : \alpha \;;\; \partial(\alpha)$.

This formalizes the algebra of generic record terms. We can show that it is a decidable syntactic theory, and derive a unitary unifying unification algorithm. Generic record terms provide a type system for a very raw view of records [Rém92b]. They also illustrate the treatment of generic variables in ML, which they generalizes to a degenerate form of intersection types [Pie91].

5.1 Presentation of generic Record Terms

As for record terms, we first define the unsorted generic record terms, and restrict them later by a compatible signature. We assume given a collection of symbols with their arities $(\mathcal{C}_n)_{n \in N}$. Let $\mathcal{L}$ be a denumerable set of labels. Let $\mathcal{K}$ be composed of

- a sort *Type* and
- a finite collection of sorts $(Row(L)_{L \in \mathcal{P}_f(\mathcal{L})}$.

Let Σ be the signature composed of the following symbols, given with their sorts:

$$
\begin{array}{ll}
\Sigma \vdash \Pi :: Row(\emptyset) \Rightarrow Type & \\
\Sigma \vdash f^\iota :: \iota^{\varrho(f)} \Rightarrow \iota & f \in \mathcal{C}, \iota \in \mathcal{K}, \varrho(f) \neq 0 \\
\Sigma \vdash \partial^L :: Type \Rightarrow Row(L) & L \in \mathcal{P}_f(\mathcal{L}) \\
\Sigma \vdash (\ell^L : \_ \;;\; \_) :: Type \otimes Row(\ell.L) \Rightarrow Row(L) & \ell \in \mathcal{L}, L \in \mathcal{P}_f(\mathcal{L} \setminus \{\ell\})
\end{array}
$$

We write $\mathcal{D}$ for the new set of symbols. Let E be the following set of axioms:

- Left commutativity. For any labels a and b and any finite subset of labels L that contains neither a nor b,

$$a^L : \alpha \;;\; b^{a.L} : \beta \;;\; \gamma = b^L : \beta \;;\; a^{b.L} : \alpha \;;\; \gamma \qquad (a \triangleright b, L)$$

- Distributivity. For any symbol f, any label a and any finite subset of labels L that do not contain a,

$$f^{Row(L)}(a^L : \alpha_1 \;;\; \beta_1, \ldots a^L : \alpha_p \;;\; \beta_p) = a^L : f^{Type}(\alpha_1, \ldots \alpha_p) \;;\; f^{Row(a.L)}(\beta_1, \ldots \beta_p) \qquad (f \triangleright a, L)$$

- Idempotence.

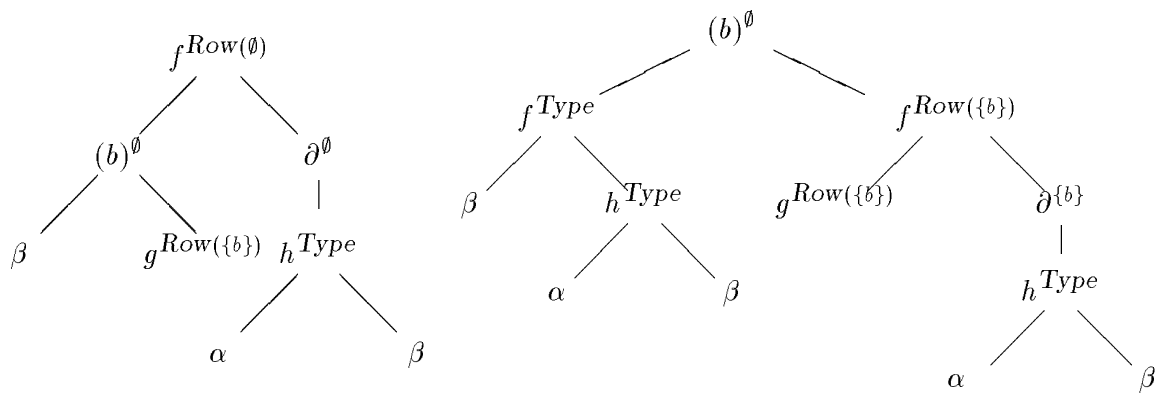
$$\partial^L \alpha = a^L : \alpha \;;\; \partial^{a.L} \alpha \qquad (\partial \triangleright a, L)$$

- Distributivity.

$$\partial^L \left(f^{Row(L)}(\alpha_1, \ldots \alpha_p) \right) = f^{Row(L)}(\partial^L \alpha_1, \ldots \partial^L \alpha_p) \qquad (\partial \triangleright f, L)$$

All axioms are collapse-free and regular. The last axiom is not linear, but it is right-linear.

For instance, the following are two E-equal record terms:



Let θ be a decreasing function on positive integers. Let Θ be the size function defined on terms by:

$$\begin{aligned}
\Theta(a^L : \tau \; ; \sigma) &= \sup\left(\Theta(\tau) + \theta(\mathit{Card}\,(L)), \Theta(\sigma) + \theta(\mathit{Card}\,(L)) - \theta(\mathit{Card}\,(a.L))\right) \\
\Theta(f^K(\tau_1, \ldots \tau_p)) &= \sup_{i \in [1,p]}(\tau_i) + 1 \\
\Theta(\partial^L(\tau)) &= \Theta(\tau) + \theta(\mathit{Card}\,(L))
\end{aligned}$$

The size Θ is constant on E-equality classes and defines a compatible ordering on terms. However, the ordering is not well-founded on $\mathcal{T}$. Let $\mathcal{T}^n$ be the subset of terms that uses only the sorts *Type* and $Row(L)$, where $\mathit{Card}\,(L)$ is at most n. The previous ordering is compatible and well-founded on all subsets $\mathcal{T}^n$.

5.2 Sufficient condition for syntacticness with non-linear axioms

The sufficient condition for syntacticness given by of theorem 1 does not apply, since the axioms $(\partial \triangleright a, L)$ are not linear and therefore the condition (h_1) is not always true. For instance, the adjoining proof, matching $\xrightarrow[\epsilon]{}\xrightarrow[\bullet\bullet]{*}$, cannot be rewritten into a proof matching $\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}$. However, it is still possible to show directly that the theory of generic record terms is syntactic, following the structure of the demonstration of theorem 1.

$$\begin{array}{c}
\overline{\partial\,(\Pi\,(f))} \\
\big\downarrow \epsilon \\
\overline{a : \underline{\Pi\,(f)} \; ; \partial\,(\Pi\,(f))} \\
\big\downarrow 2 \\
a : \overline{\Pi\,(b \; ; f : f)} \; ; \partial f
\end{array}$$

Lemma 35 *The theory of generic record terms is syntactic.*

<u>Proof</u>: The proof follows the proof of theorem 1 until the use of conditions (h_1) and (h_2). Then it resembles the proof of theorem 5, and using a tedious case analysis as we did for proving condition (h_2) of theorem 1.

The sequence of sets $(\mathcal{T}^k)_{k \in N}$ is increasing and its union is $\mathcal{T}$. Thus it is enough to show $\mathit{Synt}\,(\mathcal{T}^k)$ on all $\mathcal{T}^k$. Let H be one of them.

For any integer n, let H_n be the subset of H composed of all the terms that do not start any decreasing sequence of length n. All these sets are closed under $\xrightarrow[\upharpoonright H]{*}$. In particular, they are closed subsets of $\mathcal{T}$. Any term smaller than a term in H_{n+1} is in H_n. The sequence is increasing, and its limit is H. We show $\mathit{Synt}\,(H_n)$ by induction on n. In fact, it is enough to show that

$$\xrightarrow[\epsilon]{}\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{} \subset_{H_n} \xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

holds.

The set H_0 is composed of variables and constant symbols. Thus, the only instance of the premise is $\underset{\epsilon}{\rightarrow}\underset{\epsilon}{\rightarrow}$, for which the inclusion is satisfied, since the two axioms must be inverses and annihilate each other.

Let us assume the property $Synt(H_n)$ and show $Synt(H_{n+1})$. Let $\mathcal{D}$ be the set of all directions. A relation $\xrightarrow[\bullet\infty]{*}$ in H_{n+1} can be written as the composition

$$\left(\xrightarrow[k\infty]{*}\right)^{k\in\mathcal{D}}$$

since disjoint occurrences commute. For each direction k, the subproof at k is in H_n. Since $Synt(H_n)$, it can be rewritten in H_n so that it matches

$$\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

Re-assembling all subproofs, we get a proof matching

$$\left(\xrightarrow[k\bullet\infty]{*}\xrightarrow[k]{\delta}\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}$$

which is a proof in H_{n+1} since H_{n+1} is closed. It can be reordered as

$$\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}$$

We have shown

$$\xrightarrow[\bullet\infty]{*}\subset_{H_{n+1}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}$$

Composing the step at the root on both sides, we get

$$\underset{\epsilon}{\rightarrow}\xrightarrow[\bullet\infty]{*}\underset{\epsilon}{\rightarrow}\subset_{H_{n+1}}\underset{\epsilon}{\rightarrow}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\underset{\epsilon}{\rightarrow}$$

and we are left with proving

$$\xrightarrow[r,\epsilon]{}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k]{\delta}\right)^{k\in\mathcal{D}}\left(\xrightarrow[k\bullet\infty]{*}\right)^{k\in\mathcal{D}}\xrightarrow[s,\epsilon]{}\subset_{H_{n+1}}\xrightarrow[\bullet\infty]{*}\xrightarrow[\epsilon]{\delta}\xrightarrow[\bullet\infty]{*}$$

We show this inclusion condition by cases on the outermost axioms r and s. We omit the superscripts of symbols.

Case r is $(\partial \triangleright a)$: The axiom s must be of the form $(a \triangleright X)$, since $a : \_ \,;\, \_$ is the top symbol of the term before the last step.

Subcase X is ∂: The proof is of the form:

$$\partial(\tau)\underset{\epsilon}{\rightarrow} a : \tau \;;\; \partial(\tau)\xrightarrow[1\infty]{*} a : \sigma \;;\; \partial(\tau)\xrightarrow[2\infty]{*} a : \sigma \;;\; \partial(\sigma)\underset{\epsilon}{\rightarrow}\partial(\sigma)$$

A shorter proof is the subproof at occurrence 1 applied to $\partial(\tau)$.

Subcase X is b: The label b must be distinct from a. The proof is of the form:

$$\partial(\tau) \xrightarrow[\epsilon]{} a:\tau\ ;\ \partial(\tau) \xrightarrow[R,1\infty]{*} a:\tau'\ ;\ \partial(\tau) \xrightarrow[S,2\infty]{*} a:\tau'\ ;\ b:\tau''\ ;\ \sigma \xrightarrow[\epsilon]{} b:\tau''\ ;\ a:\tau'\ ;\ \sigma$$

The subproof

$$\partial(\tau) \xrightarrow[S,\infty]{*} b:\tau''\ ;\ \sigma$$

at the occurrence 2 is in H_{n-1}. Thus, it can be rewritten into a proof of the form

$$\partial(\tau) \xrightarrow[T,1\infty]{*} \partial(\rho) \xrightarrow[\epsilon]{} b:\rho\ ;\ \partial(\rho) \xrightarrow[P,1*]{*} \xrightarrow[Q,2*]{*} b:\tau''\ ;\ \sigma$$

Putting the pieces together, there is a proof matching

$$\partial(\tau) \xrightarrow[1\infty]{*} \partial(\rho) \xrightarrow[\epsilon]{} b:\rho\ ;\ \partial(\rho) \xrightarrow[P^{-1},1\infty]{*} b:\tau''\ ;\ \partial(\rho) \xrightarrow[22]{} b:\tau''\ ;\ a:\rho\ ;\ \partial(\rho) \xrightarrow[T^{-1},21]{*}$$
$$b:\tau''\ ;\ a:\tau\ ;\ \partial(\rho) \xrightarrow[R,21]{*} b:\tau''\ ;\ a:\tau'\ ;\ \partial(\rho) \xrightarrow[Q,22]{*} b:\tau''\ ;\ a:\tau'\ ;\ \sigma$$

Subcase X is f: Then first term must be of the form $f(\tau_1,\dots\tau_p)$. The proof is of the form

$$\partial\left(f(\tau_1,\dots\tau_p)\right) \xrightarrow[\epsilon]{} a:f(\tau_1,\dots\tau_p)\ ;\ \partial\left(f(\tau_1,\dots\tau_p)\right) \xrightarrow[R,1\infty]{*}$$
$$a:f(\sigma_1,\dots\sigma_p)\ ;\ \partial\left(f(\tau_1,\dots\tau_p)\right) \xrightarrow[S,12\infty]{*} a:f(\sigma_1,\dots\sigma_p)\ ;\ f(\rho_1,\dots\rho_p) \xrightarrow[\epsilon]{}$$
$$f(a:\sigma_1\ ;\ \rho_1,\dots a:\sigma_p\ ;\ \rho_p)$$

The subproof using axioms S at occurrence 1 is H_n, and can be rewritten into a proof matching

$$\partial\left(f(\tau_1,\dots\tau_p)\right) \xrightarrow[T,1\infty]{*} \partial\left(f(\tau_1',\dots\tau_p')\right) \xrightarrow[\epsilon]{} f(\partial(\tau_1'),\dots\partial(\tau_p')) \left(\xrightarrow[P,\bullet\infty]{*}\right) f(\rho_1,\dots\rho_p)$$

Again, the subproof using axioms R at occurrence 1 and the subproof using axioms T at the occurrence 1 are in H_n, and cannot have exactly one axiom at the empty occurrence. Therefore they are composed of a succession of proofs in all directions. Then we can prove:

$$\partial\left(f(\tau_1,\dots\tau_p)\right) \xrightarrow[\epsilon]{} f(\partial(\tau_1),\dots\partial(\tau_p)) \xrightarrow[2]{*} f(a:\tau_1\ ;\ \partial(\tau_1),\dots a:\tau_p\ ;\ \partial(\tau_p)) \xrightarrow[T,\bullet 21\infty]{*}$$
$$f(a:\tau_1\ ;\ \partial(\tau_1'),\dots a:\tau_p\ ;\ \partial(\tau_p')) \xrightarrow[P,\bullet 21\infty]{*} f(a:\tau_1\ ;\ \partial(\rho_1),\dots a:\tau_p\ ;\ \partial(\rho_p)) \xrightarrow[R,\bullet 1\infty]{*}$$
$$f(a:\sigma_1\ ;\ \partial(\rho_1),\dots a:\sigma_p\ ;\ \partial(\rho_p))$$

Case r is $(\partial \triangleright f)$: Then s must be an axiom $(f \triangleright X)$.

Subcase X is ∂: The proof is of the form:

$$\partial(f(\tau)) \xrightarrow[\epsilon]{} f(\partial(\tau_1),\dots\partial(\tau_p)) \xrightarrow[\bullet 1\infty]{*} f(\partial(\sigma_1),\dots\partial(\sigma_p)) \xrightarrow[\epsilon]{} \partial(f(\sigma))$$

A shorter proof is the composition of subproofs at occurrences $i1$ applied at occurrences $1i$ to $\partial(f(\tau))$.

Subcase X is a: The proof is of the form

$$\partial(f(\tau_1, \ldots \tau_p)) \xrightarrow[\epsilon]{} f(\partial(\tau_1), \ldots \partial(\tau_p)) \xrightarrow[R, \bullet 1\infty]{*} f(\partial(\tau'_1), \ldots \partial(\tau'_p)) \xrightarrow[\partial \rhd a, \bullet]{*}$$
$$f(a : \tau'_1 \; ; \; \partial(\tau'_1), \ldots a : \tau'_p \; ; \; \partial(\tau'_p)) \xrightarrow[S, \bullet 1\infty]{*} f(a : \sigma_1 \; ; \; \partial(\tau'_1), \ldots a : \sigma_p \; ; \; \partial(\tau'_p)) \xrightarrow[T, \bullet 21\infty]{*}$$
$$f(a : \sigma_1 \; ; \; \partial(\rho_1), \ldots a : \sigma_p \; ; \; \partial(\rho_p)) \xrightarrow[\epsilon]{} a : f(\sigma_1, \ldots \sigma_p) \; ; \; f(\partial(\rho)_1, \ldots \partial(\rho)_p)$$

Then, we can prove

$$\partial(f(\tau_1, \ldots \tau_p)) \xrightarrow[\epsilon]{} a : f(\tau_1, \ldots \tau_p) \; ; \; f(\partial(\tau), \ldots \partial(\tau)) \xrightarrow[R \cup S, 1\bullet\infty]{*}$$
$$a : f(\sigma_1, \ldots \sigma_p) \; ; \; f(\partial(\tau), \ldots \partial(\tau)) \xrightarrow[R \cup T, 2\bullet\infty]{*} a : f(\tau_1, \ldots \tau_p) \; ; \; f(\partial(\rho), \ldots \partial(\rho))$$

Subcase X is g: is not possible.

Case s is $(a \rhd \partial)$: This case is symmetric to the case r where is $(\partial \rhd a)$.

Case neither r is $(\partial \rhd X)$, nor is s $(\partial \rhd X)$: Then the axioms r and s are linear on the right and left, respectively and commutes with their right and left neighbors, so we get a proof matching:

$$\left(\xrightarrow[k\bullet\infty]{*}\right)^{k \in \mathcal{D}} \xrightarrow[r,\epsilon]{} \left(\xrightarrow[k]{\delta}\right)^{k \in \mathcal{D}} \xrightarrow[\epsilon]{} \left(\xrightarrow[k\bullet\infty]{*}\right)^{k \in \mathcal{D}}$$

We show that the inclusion

$$\xrightarrow[r,\epsilon]{} \left(\xrightarrow[k]{\delta}\right)^{k \in \mathcal{D}} \xrightarrow[s,\epsilon]{} \subset_{H_n} \xrightarrow[\bullet\infty]{*} \xrightarrow[s,\epsilon]{\delta} \xrightarrow[\bullet\infty]{*}$$

by cases on r and s.

Case r is $(X \rhd \partial)$ or s is $\partial \rhd X$: The middle equality must be empty. The two axioms r and s must be inverse and annihilate each other.

Other cases: In the remaining cases, neither r nor s are ∂ axioms. The middle axiom cannot be a ∂ axiom. The remaining cases are then exactly those of the record term algebra.

We derive the mutation in the record term algebra (figure 2): For all other pairs of terms (τ, σ), if they have identical top symbols, they are decomposable; otherwise they produce a collision.

Theorem 11 *Unification in the generic record algebra is decidable and unitary unifying.*

Proof: The theory is strict (we exhibited a compatible ordering). Therefore, the rules mutation, decomposition, collision, fusion and generalization applied any order form a complete semi-algorithm for unification. It is unitary since mutation does not introduce any disjunction.

All transformations are stable in the sets $\mathcal{T}^n$ and decrease in the following lexicographic ordering:

1. the number of symbols $f^{Row(L)}$ in the lexicographic order of decreasing $Card(L)$ (that is, bigger sets count less: MUTATE $(a \rhd f)$),

2. the number of symbols $a^L : \_ \; ; \; \_$ counted in the lexicographic order of decreasing $Card(L)$ (MUTATE $(a \rhd b)$),

$$\frac{f^{Row(L)}(\sigma_i)_{i\in[1,p]} \doteq a^L : \alpha \mathrel{;} \gamma \doteq e}{\exists\, \alpha_i, \gamma_{i\, i\in[1,p]} \cdot \bigwedge \begin{cases} a^L : \alpha \mathrel{;} \gamma \doteq e \\ \alpha \doteq f^{Type}(\alpha_1, \ldots \alpha_p) \\ \gamma \doteq f^{Row(a.L)}(\gamma_1, \ldots \gamma_p) \\ \sigma_i \doteq a^L : \alpha_i \mathrel{;} \gamma_i \qquad i \in [1,p] \end{cases}} \leadsto \qquad \textsc{Mutate}(a \triangleright f)$$

$$\frac{a^L : \tau \mathrel{;} \sigma \doteq b^L : \alpha \mathrel{;} \beta \doteq e}{\exists\, \gamma \cdot \bigwedge \begin{cases} b^L : \alpha \mathrel{;} \beta \doteq e \\ \sigma \doteq b^{a.L} : \alpha \mathrel{;} \gamma \\ \beta \doteq a^{b.L} : \tau \mathrel{;} \gamma \end{cases}} \leadsto \qquad \textsc{Mutate}(a \triangleright b)$$

$$\frac{a^L : \tau \mathrel{;} \rho \doteq \partial^L(\alpha) \doteq e}{\partial^L(\alpha) \doteq e \wedge \tau \doteq \alpha \wedge \rho \doteq \partial^{a.L}(\alpha)} \leadsto \qquad \textsc{Mutate}(\partial \triangleright b)$$

$$\frac{f^{Row(L)}(\sigma_1, \ldots \sigma_p) \doteq \partial^L(\alpha) \doteq e}{\exists\, (\alpha)_{i\in[1,p]} \cdot \bigwedge \begin{cases} \partial^L(\alpha) \doteq e \\ \alpha \doteq f^{Type}(\alpha_1, \ldots \alpha_p) \\ \sigma_i \doteq \partial^L(\alpha_i) \qquad i \in [1,p] \end{cases}} \leadsto \qquad \textsc{Mutate}(f \triangleright \partial)$$

Figure 2: Mutation in the generic algebra of record terms

3. the number of other symbols (other mutations and decomposition),
4. the sum of heights of terms (generalization),
5. the number of multi-equations (fusion).

Thus, applying the rules in any order always terminates. ■

6 Comparison with other work

In our approach, records are terms of a sorted algebra modulo regular equations. We have infinitely many equations (indexed by labels), but all equations act locally on terms, since the axioms are of depth at most 2.

In record calculi, it is possible to define a record s by adding a field $(a = x)$ to the record r whether r already defines field a or not. It is tempting to reflect the structure of record objects into record types and assign the type $(a : \tau \mathrel{;} \sigma)$ to s provided s has type σ. In the case, where r already defines field a, it has a type of the form $(a : \tau' \mathrel{;} \rho)$; then s has type $(a : \tau \mathrel{;} (a : \tau' \mathrel{;} \rho))$. The type component of r on field a is meaningless in the type of s, since the field of s is x. This can be realized by adding a non-regular absorption axiom

$$a : \alpha \mathrel{;} a : \alpha' \mathrel{;} \gamma = a : \alpha \mathrel{;} \gamma$$

Such an axiom much be treated carefully. This approach has been taken in two recent proposals for record terms, one by A. Hense and G. Smolka [HS92], the other by B. Berthomieu [Ber93].

Both approaches also restrict terms to the equivalent of our canonical terms. The notion of expansion is not based on distributivity equations; instead, it is incorporated in the substitutions themselves, using more complex sorts to control substitution.

For instance, the record term $a : \tau \;;\; abs$ is coded as $a : \tau \;;\; \alpha^{abs}$ in Berthomieu's system; the sort abs assigned to α restricts any substitution to be at least of the form $\alpha \mapsto \ell : abs \;;\; \beta^{abs}$. The following E-equality in record terms

$$a : \tau \;;\; abs = a : \tau \;;\; (b : abs \;;\; abs)$$

is not an equality in Berthomieu's system [Ber93]; the two equivalent terms are only in the instance relation

$$a : \tau \;;\; \alpha^{abs} < a : \tau \;;\; (b : abs \;;\; \alpha'^{abs})$$

The right hand side is equal to the substitution of variable α of sort abs by $b : abs ; \alpha'^{abs}$ in the left hand side. This is the smallest possible substitution for α as a result of its sort constraint. Berthomieu's approach is more complex in the case of simple record algebras, but it seems to simplify the treatment of generic record algebras: the generic record term

$$a : \tau \;;\; \beta \rightarrow \partial(\alpha)$$

would be represented in Berthomiue's system as

$$a : \tau \;;\; \gamma^{\forall \beta \cdot \beta \rightarrow \alpha}$$

using the explicit quantifiers in sorts. Then, the variable γ can be replaced by $\ell : \rho \;;\; \gamma'^{\forall \beta \cdot \beta \rightarrow \alpha}$ provided ρ is an instance of the type scheme $\forall \beta \cdot \beta \rightarrow \alpha$, that is, of the form $\sigma \rightarrow \alpha$. The sorts are terms and can also be instantiated during unification. This is an additional difficulty but also a gain of expressiveness.

We find the approach of Berthomieu quite interesting for the extension to generic record algebras, but we prefer our approach for simple record terms, since it fits nicely in a known framework.

Conclusion

We introduced a framework in which syntacticness of an equational presentation can be studied more easily. We defined record algebras over an initial set of symbols as the quotient of a free sorted algebras by left commutativity and distributivity axioms. We showed that it is syntactic and decidable and we deduced an efficient, unitary unifying algorithm for unification. Many variants of record algebras can be obtained by restricting the terms by a signature that is compatible with the equations. Different instances have already been used to provide type systems for languages with records.

The extension of record algebras to recursive types has not been addressed here. In practice, the algorithm that we described also works with non-strict systems of multi-equations, which represent recursive terms. However, the notion of regular trees modulo equations has to be defined before any correspondence between these and non-strict systems of multi-equations can be studied. It seems that the algebra of record terms is sufficiently constrained by the sorts that there would be a close correspondence between the two, which cannot be expected in general.

The extension of record algebras to generic record algebras is a difficult step. Even if it can be extended to higher order genericity, it seems too difficult to be the right notion. Making a closer connection with the record terms of Berthomieu is a promising approach. The generality of record algebras suggests that there should be other useful applications.

References

[Ber93] Bernard Berthomieu. Tagged types, a theory of order sorted types for tagged expressions. Research Report 93083, LAAS, 7, Avenue du Colonnel Roche, 31077 Toulouse, France, March 1993.

[HS92] Andreas V. Hense and Gert Smolka. A verification of extensible record types. In Zhongzhi Shi, editor, *Proceedings of the IFIP TC12/WG12.3 International Workshop on Advances in Automated Reasoning*, pages 13–16, Beijing, P.R. China, July 1992. International Federation for Information Processing, Elsevier, North-Holland, Excerpta Medica. to appear.

[Kir85] Claude Kirchner. *Méthodes et outils de conception systématique d'algorithmes d'unification dans les théories équationnelles.* Thèse de doctorat d'état en informatique, Université de Nancy 1, 1985.

[KJ90] Claude Kirchner and Jean-Pierre Jouannaud. Solving equations in abstract algebras: a rule-based survey of unification. Research Report 561, Université de Paris Sud, Orsay, France, April 1990.

[KK89] Claude Kirchner and François Klay. Syntactic theories and unification. CRIN & INRIA-Loraine, Nancy (France), 1989.

[MM82] Alberto Martelli and Ugo Montanari. An efficient unification algorithm. *ACM Transactions on Programming Languages and Systems*, 4(2):258–282, 1982.

[Nip89] Tobias Nipkow. Proof transformations for equational theories. Technical Report 181, University of Cambridge, Computer laboratory, 1989.

[Pie91] Benjamin C. Pierce. Programming with intersection types, union types, and polymorphism. Technical Report CMU-CS-91-106, Carnegie Mellon University, Pittsburg, Pensylvania, February 1991.

[Rém90] Didier Rémy. *Algèbres Touffues. Application au Typage Polymorphe des Objects Enregistrements dans les Langages Fonctionnels.* Thèse de doctorat, Université de Paris 7, 1990.

[Rém92a] Didier Rémy. Extending ML type system with a sorted equational theory. Technical Report 1766, INRIA-Rocquencourt, BP 105, F-78 153 Le Chesnay Cedex, 1992.

[Rém92b] Didier Rémy. Projective ML. In *1992 ACM Conference on Lisp and Functional Programming*, pages 66–75, New-York, 1992. ACM press.

[Rém93] Didier Rémy. Type inference for records in a natural extension of ML. In Carl A. Gunter and John C. Mitchell, editors, *Theoretical Aspects Of Object-Oriented Programming. Types, Semantics and Language Design.* MIT Press, 1993. To appear. Also research report 1431, Mai 1991, INRIA-Rocquencourt, France.