

Subtyping Recursive Types

ROBERTO M. AMADIO

CNRS-CRIN, Nancy

and

LUCA CARDELLI

DEC, Systems Research Center

We investigate the interactions of subtyping and recursive types, in a simply typed λ -calculus. The two fundamental questions here are whether two (recursive) types are in the subtype relation and whether a term has a type. To address the first question, we relate various definitions of type equivalence and subtyping that are induced by a model, an ordering on infinite trees, an algorithm, and a set of type rules. We show soundness and completeness among the rules, the algorithm, and the tree semantics. We also prove soundness and a restricted form of completeness for the model. To address the second question, we show that to every pair of types in the subtype relation we can associate a term whose denotation is the uniquely determined coercion map between the two types. Moreover, we derive an algorithm that, when given a term with implicit coercions, can infer its least type whenever possible.

Categories and Subject Descriptors: D.3.1 [**Programming Languages**]: Formal Definitions and Theory; D.3.3 [**Programming Languages**]: Language Constructs and Features; F.3.2 [**Logics and Meanings of Programs**]: Semantics of Programming Languages; F.3.3 [**Logics and Meanings of Programs**]: Studies of Program Constructs; F.4.3 [**Mathematical Logic and Formal Languages**]: Formal Languages

General Terms: Languages, Theory

Additional Key Words and Phrases: Coercions, lambda-calculus, partial-equivalence relations, recursive types, regular trees, subtyping, tree orderings, type equivalence, typechecking algorithm

1. INTRODUCTION

Subtyping is an inclusion relation between types that is present to some degree in many programming languages. Subtyping is especially important in object-oriented languages, where it is crucial for understanding the much more complex notions of inheritance and subclassing.

A preliminary version of this paper appears in the Conference Record of the Eighteenth Annual ACM Symposium on Principles of Programming Languages, ACM Press, 1991. R. M. Amadio's work has been supported in part by Digital Equipment Corporation and in part by the Stanford-CNR Collaboration Project.

Authors' addresses: R. M. Amadio, CRIN-CNRS, BP239, F-54506, Vandoeuvre-les-Nancy, France; L. Cardelli, DEC Systems Research Center, 130 Lytton Avenue, Palo Alto, CA 94301.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1993 ACM 0164-0925/93/0900-0575 \$01.50

ACM Transactions on Programming Languages and Systems, Vol. 15, No. 4, September 1993, Pages 575-631

Recursive types are also present in most languages. These types are supposed to *unfold* recursively to match other types. Moreover, unfolding must preserve typing soundness and not cause the compiler to diverge.

In this paper we investigate the interaction of *unrestricted* recursive types with subtyping. This interaction is present in some modern languages based on *structural type matching* (where type equality or subtyping is determined by some abstract type structure, and not by how types are syntactically presented). In the past, recursive types have often been restricted by other language features, for example, by explicit unfolding in ML, and by *name matching* in Modula-2. Algol68 was the first language to rely on a structural type equality algorithm for recursive types. Thereafter name matching became popular, largely because it is easier to implement but also because it prevents accidental matches based on type structure.

Name matching determines type equality by relying, at least partially, on the names assigned to types in a given program, instead of on their structure. With name matching, recursive analysis can stop at occurrences of type names. Unfortunately there is no general definition of name matching; each language, and sometimes each compiler, implements it slightly differently. Types with the same meaning (in the eye of the programmer) may or may not be equated in different runs of the compiler, depending on irrelevant textual perturbations that affect the name-matching rules.

The inconsistency of name-matching rules becomes a problem in distributed environments, where type definitions and data may migrate outside the compiler or program run in which they are created. Types and data should have a meaning independent of particular runs; hence, languages such as Modula-3 [22] and other experimental languages such as Amber [10] and Quest [9, 11], concerned with data persistence and data migration, have again adopted structural matching. Since these languages also rely on subtyping, structural subtyping becomes an issue. Because of various language design issues, Modula-3 restricts itself to structural equivalence plus a limited form of structural subtyping; in this paper we deal with the unrestricted combination of recursion and subtyping, which forms the basis of Amber and Quest.

With this motivation, we investigate type systems with recursive types and subtyping and the related problems of structural matching and structural subtyping. Structural-matching techniques are well known and have strong connections with well-understood theoretical concepts. Structural subtyping is a much newer subject. We provide the first complete theory of recursive subtypes that leads naturally to an effective type theory and to type-checking algorithms. In practice it is easy to adapt algorithms for structural typing to structural subtyping (although to our knowledge, this was first done in Amber), but formalizing the type rules and the proofs of correctness of the algorithms is more challenging. We show that both our algorithm and our type rules are complete with respect to a natural notion of subtyping.

In the rest of the Introduction we provide the basic intuitions about recursive subtypes, and we illustrate the main problems along with several

nonsolutions. Section 2 formalizes the syntax of a basic calculus with recursive types, and Section 3 introduces a subtyping relation based on a tree ordering. Section 4 describes a subtyping algorithm, and Section 5 describes the corresponding type rules. A partial equivalence relation model is given in Section 6. Finally, Section 7 relates subtyping to type coercions.

1.1 Types

A *type*, as normally intended in programming languages, is a collection of values sharing a common structure or shape. Examples of *basic* types are: *Unit*, the trivial type containing a single element, and *Int*, the collection of integer numbers. Examples of *structured* types are: $\text{Int} \rightarrow \text{Int}$, the functions from integers to integers; $\text{Int} \times \text{Int}$, the pairs of two integers; and $\text{Unit} + \text{Int}$, the *disjoint union* of *Unit* and *Int* consisting of either a unit value marked “left” or an integer marked “right” (given two arbitrary but distinct marks).

A *recursive type* is a type that satisfies a *recursive type equation*. Common examples are:

$$\text{Tree} = \text{Int} + (\text{Tree} \times \text{Tree})$$

the collection of binary trees with integer leaves, and:

$$\text{List} = \text{Unit} + (\text{Int} \times \text{List})$$

the collection of lists of integers. Note that these are not definitions of *Tree* and *List*; they are equational properties that any definition of *Tree* and *List* must satisfy.

There are also useful examples of recursion involving function spaces, typical of the object-oriented style of programming:

$$\text{Cell} = (\text{Unit} \rightarrow \text{Int}) \times (\text{Int} \rightarrow \text{Cell}) \times (\text{Cell} \rightarrow \text{Cell}).$$

A *Cell* is interpreted as the collection of integer-containing memory cells, implemented as triples of functions *read*: $\text{Unit} \rightarrow \text{Int}$, *write*: $\text{Int} \rightarrow \text{Cell}$, and *add*: $\text{Cell} \rightarrow \text{Cell}$. In each of these functions the current cell is implicit, so, for example, *add* needs only to receive another cell in order to perform a binary addition.

Recursive types can hence be described by equations, and we see that in fact they can be unambiguously *defined* by equations. To see this, we need some formal way of reasoning about the solutions of type equations. These formal tools become particularly useful if we start examining problematic equations such as $t = t$, $s = s \times s$, $r = r \rightarrow r$, etc., for which it is not clear whether there are solutions or whether the solutions are unique.

It is appealing to set up sufficient conditions so that type equations have *canonical* solutions. Then, if we have an equation such as $t = \text{Unit} + (\text{Int} \times t)$, we can talk about *the* solution of the equation. Such a canonical solution can then be indicated by a term such as $\mu t. \text{Unit} + (\text{Int} \times t)$; *the* type t that is equal to $\text{Unit} + (\text{Int} \times t)$. Here $\mu t. \alpha$ is a new type construction just introduced for denoting canonical solutions.

To say that $L \triangleq \mu t. \text{Unit} + (\text{Int} \times t)$ (where \triangleq means *equal by definition*) is the solution of the *List* equation implies that L must satisfy the equation, that is, $L = \text{Unit} + (\text{Int} \times L)$ must be provable. This requirement suggests

the most important rule for the $\mu t.\alpha$ construction, which amounts to a one-step unfolding of the recursion:

$$\mu t.\alpha = [\mu t.\alpha/t]\alpha$$

meaning that $\mu t.\alpha$ is equal to α where we replace t by $\mu t.\alpha$ itself. In our example we have:

$$L = \mu t.\text{Unit} + (\text{Int} \times t) = [L/t](\text{Unit} + (\text{Int} \times t)) = \text{Unit} + (\text{Int} \times L)$$

which is the equation we expected to hold.

Having discussed recursive types, we now need to determine when a value belongs to a recursive type. The rule above for $\mu t.\alpha$ allows us to expand recursive types arbitrarily far, for a finite number of expansions. Hence, we can postulate that a finite value belongs to a recursive type if it belongs to one of its finite expansions according to the ordinary typing rules. That is, we push the troublesome μ 's far enough until we no longer need to consider them.

However, if the values are not finite (for example, if they are defined recursively), we may not be able to push the μ 's out of the way. In that case, we need to provide adequate notions of finite *approximations* of values and types and postulate that a value belongs to a type when every approximation of the value belongs to some approximation of the type. An approximation α^n of a type expression α is an appropriate truncation of α at depth n ; hence it is different from an unfolding. This will be made precise in later sections.

1.2 Subtypes

If types are collections of values, *subtypes* should be subcollections. For example, we can introduce two new basic types \perp (*bottom*), the collection containing only the divergent computation, and \top (*top*), the collection of all values. Then \perp should be a subtype of every type, and every type should be a subtype of \top . We write these relations as $\perp \leq \alpha$ and $\alpha \leq \top$.

Function spaces $\alpha \rightarrow \beta$ have a subtyping rule that is *antimonotonic* in the first argument, i.e.,

$$\alpha \rightarrow \beta \leq \alpha' \rightarrow \beta' \quad \text{if} \quad \alpha' \leq \alpha \quad \text{and} \quad \beta \leq \beta'.$$

For example, if $\text{Nat} \leq \text{Int}$, and $f: \text{Int} \rightarrow \text{Cell}$ stores an integer into a cell, then f is also willing to store a natural number into a cell, i.e., $f: \text{Nat} \rightarrow \text{Cell}$. Hence, it is sound to have $\text{Int} \rightarrow \text{Cell} \leq \text{Nat} \rightarrow \text{Cell}$, but not the opposite. This antimonotonic rule is familiar in object-oriented programming, where it is one of the main considerations for the correct type checking of methods.

Adequate subtyping rules can be found for all the other type constructions we may have. For example, for products we have $\alpha \times \beta \leq \alpha' \times \beta'$ if $\alpha \leq \alpha'$ and $\beta \leq \beta'$. Similarly, for disjoint unions we have $\alpha + \beta \leq \alpha' + \beta'$ if $\alpha \leq \alpha'$ and $\beta \leq \beta'$.

What is, then, subtyping for recursive types? The intuition we adopt is that two recursive types α and β are in the subtype relation if their

infinite unfoldings are also in this relation, in some appropriate sense. We might at first just consider finite unfoldings α^+ of a type α and require that “ $\alpha \leq \beta$ if for every α^+ of α there is a β^+ of β with $\alpha^+ \leq \beta^+$.” However, we shall see shortly that this condition is not strong enough. Hence, we insist on inclusion of infinite unfoldings. This is made precise by the notion, mentioned above, of finite approximations α^n of a type α and by defining “ $\alpha \leq \beta$ if, for every n , $\alpha^n \leq \beta^n$.”

Unfortunately, the formal subtyping rules for recursive types and the related algorithms cannot rely on approximations, since “ $\alpha^n \leq \beta^n$ for every n ” involves testing an infinite number of conditions. The subtyping rules should rely instead on “finitary” rules, and it is therefore not so obvious how to invent a collection of rules that achieves the desired effect. For example, a first idea might be simply to say that:

$$\text{if } \alpha \leq \beta \text{ then } \mu t. \alpha \leq \mu t. \beta \quad (1)$$

where t may occur freely in α and β . By this we can show that, for example, $\mu t. \top \rightarrow t \leq \mu t. \perp \rightarrow t$, just from the assumption that $t \leq t$. Unfortunately we also have:

$$(1) \text{ implies } \alpha \triangleq \mu t. t \rightarrow \perp \leq \mu t. t \rightarrow \top \triangleq \beta,$$

and this is quite wrong. By unfolding both α and β twice we get:

$$(1) \text{ implies } (\alpha \rightarrow \perp) \rightarrow \perp \leq (\beta \rightarrow \top) \rightarrow \top,$$

and these are not subtypes: the first \perp on the left and first \top on the right are in the wrong inclusion relation ($\top \leq \perp$), being in antimonotonic position.

The problem with Rule (1) comes from the *negative* occurrences (on the left of an odd number of \rightarrow 's) of the recursion variable. In fact Rule (1) is sound for types that are monotonic in the recursion variable.

A correct (and finitary) rule for inclusion of recursive types is instead the following:

$$(s \leq t \Rightarrow \alpha \leq \beta) \Rightarrow \mu s. \alpha \leq \mu t. \beta \quad (2)$$

where s occurs only in α , and t occurs only in β . That is, if by assuming the inclusion of the recursive variables we can verify the inclusion of the bodies, then we can deduce the inclusion of the recursive types. (It is interesting to check how subtyping fails now on the example above.)

Going back to the List example, if we have $\text{Nat} \leq \text{Int}$ and

$$\text{NatList} \triangleq \mu s. \text{Unit} + (\text{Nat} \times s)$$

$$\text{IntList} \triangleq \mu t. \text{Unit} + (\text{Int} \times t)$$

then we can safely deduce $\text{NatList} \leq \text{IntList}$ from Rule (2) since $s \leq t$ implies $\text{Unit} + (\text{Nat} \times s) \leq \text{Unit} + (\text{Int} \times t)$.

On the other hand, the Cell example does not work as smoothly.

$$\text{NatCell} \triangleq \mu s. (\text{Unit} \rightarrow \text{Nat}) \times (\text{Nat} \rightarrow s) \times (s \rightarrow s)$$

$$\text{IntCell} \triangleq \mu t. (\text{Unit} \rightarrow \text{Int}) \times (\text{Int} \rightarrow t) \times (t \rightarrow t).$$

Here we cannot conclude $\text{NatCell} \leq \text{IntCell}$ from Rule (2), because of antimonotonicity: both the inclusion of the second component (*write*) and the

inclusion of the third (*add*) fail. This is however not a deficiency of Rule (2); such a conclusion would be unsound. For example, a *NatCell* might have a *write* function of type $\text{Nat} \rightarrow \text{NatCell}$ that fails on negative numbers. If such a cell were considered as an *IntCell*, it would be possible to pass a negative integer to this *write* and cause it to fail. These issues are related to the type checking of object types in object-oriented languages, and are discussed at length in [15] and [8].

1.3 Equality of Recursive Types

We need now to consider strong notions of equality of recursive types. This is necessary because Rule (2) is weak in some areas; for example, we cannot deduce directly from it that

$$\mu t. t \rightarrow t \leq \mu s. s \rightarrow s$$

because this would require assuming both $s \leq t$ and $t \leq s$. The combination of Rule (2) and equality rules will finally give us all the power we need.

To check whether two recursive types $\mu s. \alpha'$ and $\mu t. \beta'$ are equivalent, we could assume $s = t$ and attempt to prove $\alpha' = \beta'$ under this assumption. This would work for $\mu t. t \rightarrow t$ and $\mu s. s \rightarrow s$. But now consider the types:

$$\alpha \triangleq \mu s. \text{Int} \rightarrow s \quad \beta \triangleq \mu t. \text{Int} \rightarrow \text{Int} \rightarrow t.$$

They both expand infinitely into $\text{Int} \rightarrow \text{Int} \rightarrow \text{Int} \rightarrow \text{Int} \rightarrow \dots$, and they also have the same set of values (for example, recursive terms like $\mu f. \lambda x. \text{Int}. f$). However, the assumption $s = t$ does not show $\text{Int} \rightarrow s = \text{Int} \rightarrow \text{Int} \rightarrow t$; we get stuck on the question whether $s = \text{Int} \rightarrow t$.

Another attempt might involve expanding the μ 's, but unfortunately we cannot expand them out of existence. By unfolding alone we can get only

$$\begin{aligned} \alpha &= \mu s. \text{Int} \rightarrow s = \text{Int} \rightarrow (\mu s. \text{Int} \rightarrow s) = \text{Int} \rightarrow \text{Int} \rightarrow (\mu s. \text{Int} \rightarrow s) \\ &= \text{Int} \rightarrow \text{Int} \rightarrow \alpha \\ \beta &= \mu t. \text{Int} \rightarrow \text{Int} \rightarrow t = \text{Int} \rightarrow \text{Int} \rightarrow (\mu t. \text{Int} \rightarrow \text{Int} \rightarrow t) = \text{Int} \rightarrow \text{Int} \rightarrow \beta \end{aligned}$$

which after a few unfoldings leaves us with the original problem of determining whether $\alpha = \beta$. This is what we meant earlier by the insufficiency of “ $\alpha \leq \beta$ if for every expansion α^+ of α there is a β^+ of β with $\alpha^+ \leq \beta^+$.”

In fact, we seem to have made some progress here; we have come back to the original question $\alpha = \beta$ only after analyzing the entire structure of α and β . It seems that we should then be able to conclude that $\alpha = \beta$, because a complete analysis of α and β has found no contradiction. This kind of reasoning is possible, but it has to be carefully justified; and in general we need to determine the conditions under which this stronger notion of equality does not lead to a circular argument.

Note that in the process above we have found a single context $C[X] \triangleq \text{Int} \rightarrow \text{Int} \rightarrow X$ such that $\alpha = C[\alpha]$ and $\beta = C[\beta]$; that is, both α and β are fixpoints of C . We shall be able to show that all the nontrivial (formally, *contractive*) type contexts $C[X]$ have unique fixpoints over infinite trees, and therefore if they have two fixpoints these must be equal. Hence, the necessary

rule for determining type equality can be formulated as follows:

$$\alpha = C[\alpha] \wedge \beta = C[\beta] \wedge C \text{ contractive} \Rightarrow \alpha = \beta. \quad (3)$$

It remains to be shown how to generate contractive contexts that allow us to equate any two types that have equal infinite expansions. This can be done via an algorithm, and in fact a natural one. We will show that this algorithm is sound (it will not equate types with different infinite expansions) and complete (it will equate all types that have equal infinite expansions). Such proofs of correctness of algorithms are among our major goals here, but first we need to carefully develop a formal framework.

1.4 Subtyping of Recursive Types

The problem of equating recursive types such as α and β can be related to well-known solvable problems, such as the equivalence of finite-state automata. However, the similar problem for subtyping has no well-known parallel. Take, for example:

$$\gamma \triangleq \mu s. \text{Int} \rightarrow s \quad \delta \triangleq \mu t. \text{Nat} \rightarrow \text{Nat} \rightarrow t.$$

Again, looking at the infinite expansions we obtain $\gamma = \text{Int} \rightarrow \text{Int} \rightarrow \dots$ and $\delta = \text{Nat} \rightarrow \text{Nat} \rightarrow \dots$, from which we would like to deduce $\gamma \leq \beta$ by anti-monotonicity. But what are the exact rules? Attempts to unfold γ and δ fall into the same difficulties as before.

The strategy here is to reduce the subtyping problem to an equality problem, which we solve by Rule (3), plus Rule (2). That is, we first show that $\delta' \triangleq \mu t. \text{Nat} \rightarrow t = \mu t. \text{Nat} \rightarrow \text{Nat} \rightarrow t \equiv \delta$. After that, we can use Rule (2) to show $\gamma \leq \delta'$, and hence $\gamma \leq \delta$.

Initially, this strategy suggests a two-step algorithm that first synchronizes the recursions in some appropriate way and then uses Rule (2) without additional folding/unfolding. Instead, we present an algorithm that tests subtyping of recursive types directly; the correspondence between the algorithm and the rules is then less obvious.

The example above involves two distinct recursive types for which Rule (2) alone is not sufficient to determine subtyping. This example may seem artificial; however, this situation can easily happen in practice. As a slightly more plausible example, suppose we define the type of lists of alternating integers and naturals:

$$\text{IntNatList} \triangleq \mu t. \text{Unit} + \text{Int} \times (\text{Unit} + \text{Nat} \times t).$$

This definition could arise more naturally from a mutual-recursion construct in some programming language, for example:

$$\begin{aligned} \text{Let Rec IntNatList} &= \text{Unit} + \text{Int} \times \text{NatIntList} \\ \text{and NatIntList} &= \text{Unit} + \text{Nat} \times \text{IntNatList}. \end{aligned}$$

One would certainly expect $\text{NatList} \leq \text{IntNatList}$ to hold. But, we have

$$\text{NatList} \triangleq \mu s. \text{Unit} + \text{Nat} \times s.$$

Hence, we have to show first that $\text{NatList} = \mu s. \text{Unit} + \text{Nat} \times (\text{Unit} + \text{Nat} \times s)$, and only then can we apply Rule (2) successfully.

1.5 Algorithm Outline

We describe the algorithm informally, and we show some sample runs. This is only an approximation of the algorithm analyzed in the formal part, but it should explain the main ideas. A more detailed description is given in Section 4.4.

A recursive type of the form $\mu t. \dots t \dots$ can be represented in memory as a cyclic linked structure such that every occurrence of t in the recursive body is represented by the address of the corresponding μt structure, i.e., by a *backpointer*. Otherwise, all subexpressions of a type expression, including μ subexpressions, are uniquely determined by their address in memory. Every time the algorithm reaches a μ structure, possibly through a backpointer, it has the option of analyzing the interior of the structure (“unfolding” the recursive type) or of comparing its address with other addresses as a termination condition. The algorithm for $\alpha \leq \beta$ operates on a pair of linked structures and a *trail*. A trail is a set of address pairs that records the pairs of addresses that have been jointly encountered when following a pair of paths in the two linked structures. To avoid diverging on cyclic structures, the algorithm registers a local successful termination when it reaches a pair of addresses that have already been *seen*, that is, a pair of addresses that are contained in the trail.

The algorithm to determine whether $\alpha \leq \beta$ starts with an empty trail and proceeds through the following steps in sequence. We only consider basic types, function types, and recursive types.

- (1) Succeed if the pair of addresses of α and β (in this order) is contained in the trail. (In this situation, we have completely explored a pair of cyclic paths and found no subtyping failures; hence we declare success for these paths.)
- (2) Succeed if α and β are type constants that are equal or in the subtype relation. (This is the base case for the given collection of basic types and basic inclusions.)
- (3) When α is $\alpha' \rightarrow \alpha''$ and β is $\beta' \rightarrow \beta''$, recur on $\beta' \leq \alpha'$ and on $\alpha'' \leq \beta''$. Succeed if both recursions succeed. (This is the case for function types; note the swapping of inclusion on the domains because of antimonotonicity of \rightarrow ; no such swapping would occur for data type constructors such as \times and $+$).
- (4.1) When α is $\mu t. \alpha'$ and β is $\mu s. \beta'$, add the pair of addresses of α and β (in this order) to the trail, and recur on $\alpha' \leq \beta'$. Succeed if the recursion succeeds. (The presence of μ 's signals potential cyclic paths; hence, we store the current pair of addresses in the trail so that Case (1) can prevent looping. We use an ordered pair of addresses because inclusion is, obviously, not symmetric; this detail differs from the standard trail algorithms for type equivalence. The next two cases are similar.)
- (4.2) When α is $\mu t. \alpha'$, add the pair of addresses of α and β to the trail and recur on $\alpha' \leq \beta$. Succeed if the recursion succeeds.
- (4.3) When β is $\mu s. \beta'$, add the pair of addresses of α and β to the trail and recur on $\alpha \leq \beta'$. Succeed if the recursion succeeds.

- (5) Otherwise, fail. (This means we have found a pair of incomparable type expressions, such as a function type and a base type.)

A faithful description of a run of this algorithm would involve assigning arbitrary addresses to subexpressions of type expressions; this would only obscure the exposition. Instead, we display the type expressions, and we leave their addresses implicit; the reader is urged to keep this in mind.

The diagrams below represent execution trees. The starting goal is at the bottom; the branching represents recursive calls; and the leaves represent termination conditions. The trail is shown in curly brackets; its elements are written as $t \leq s$ and represent pairs of addresses of type expressions. We indicate in parentheses the step of the algorithm used in each line to obtain the line above it.

The first sample run involves two types with matching μ structures; their inclusion is nontrivial because of antimonotonicity.

$$\begin{array}{l}
 \{t \leq s\} t \leq s \text{ (1)} \quad \{t \leq s\} \perp \leq t \text{ (2)} \\
 \{t \leq s\} s \rightarrow \perp \leq t \rightarrow t \text{ (3)} \quad \{t \leq s\} \perp \leq \top \text{ (2)} \\
 \{t \leq s\} (t \rightarrow t) \rightarrow \perp \leq (s \rightarrow \perp) \rightarrow \top \text{ (3)} \\
 \{ \} \mu t.((t \rightarrow t) \rightarrow \perp) \leq \mu s.((s \rightarrow \perp) \rightarrow \top) \text{ (4.1)}
 \end{array}$$

The second sample run involves two types with mismatching μ structures. This mismatch introduces the need to examine a cyclic path more than once. For this, we use a *loopback* step, which corresponds to following a cyclic structure back to its original entry point (an artificial loopback step is needed only because, as we said, we keep the address information implicit). In the algorithm above, a loopback situation corresponds to a failure of Step (1) followed by some dereferencing of backpointers that leads to Step (4).

$$\begin{array}{l}
 \{t \leq s, t \leq \perp \rightarrow s\} \perp \leq \top \text{ (2)} \quad \{t \leq s, t \leq \perp \rightarrow s\} t \leq s \text{ (1)} \\
 \{t \leq s, t \leq \perp \rightarrow s\} \top \rightarrow t \leq \perp \rightarrow s \text{ (3)} \\
 \{t \leq s\} \mu t.(\top \rightarrow t) \leq \perp \rightarrow s \text{ (4.2)} \\
 \{t \leq s\} \perp \leq \top \text{ (2)} \quad \{t \leq s\} t \leq \perp \rightarrow s \text{ (loopback)} \\
 \{t \leq s\} \top \rightarrow t \leq \perp \rightarrow (\perp \rightarrow s) \text{ (3)} \\
 \{ \} \mu t.(\top \rightarrow t) \leq \mu s.(\perp \rightarrow (\perp \rightarrow s)) \text{ (4.1)}
 \end{array}$$

Hence, in this run we go around the μt loop twice in order to go around the μs loop once.

For other interesting examples, check how $\mu t.(t \rightarrow t) \leq \mu s.(s \rightarrow s)$ succeeds and how $\mu t.(t \rightarrow \perp) \leq \mu s.(s \rightarrow \top)$ fails.

One of the main aims of this paper is to show that the algorithm above is consistent with, and in fact equivalent to, Rules (2) and (3) of Sections 1.2 and 1.3. For this we need to place both the rules and the algorithm in a more formal framework.

1.6 Formal Development

Having explained most of the problems and the unsatisfactory solutions arising from subtyping recursive types, we can now proceed to the formal treatment.

So far we have discussed rules for the subtyping of recursive types which are motivated by some operational intuition. In the following we will broaden

our perspective and consider various notions of type equivalence, $\alpha = \beta$, and subtyping, $\alpha \leq \beta$. These are induced by:

- (a) An *ordering on infinite trees*: $\alpha =_T \beta$, $\alpha \leq_T \beta$ (Section 3).
- (b) An *algorithm*: $\alpha =_A \beta$, $\alpha \leq_A \beta$ (Section 4).
- (c) A collection of *typing rules*: $\alpha =_R \beta$, $\alpha \leq_R \beta$ (Section 5).
- (d) A collection of *per models*: $\alpha =_M \beta$, $\alpha \leq_M \beta$ (Section 6).

The mathematical content of the paper consists mainly in analyzing the relationships between these notions. For a simply typed lambda calculus with recursive types (described in Section 2) we show, among other properties:

$$\begin{aligned} \alpha =_T \beta &\Leftrightarrow \alpha =_A \beta \Leftrightarrow \alpha =_R \beta \Rightarrow \alpha =_M \beta \\ \alpha \leq_T \beta &\Leftrightarrow \alpha \leq_A \beta \Leftrightarrow \alpha \leq_R \beta \Rightarrow \alpha \leq_M \beta. \end{aligned}$$

Moreover, we prove a restricted form of completeness with respect to the model (Section 6.3); we show the definability in the calculus of certain maps that interpret coercions (Section 7.1); and we give an algorithm for computing the minimal type of a term with respect to \leq_T (Section 7.2). All these results support the relevance of the theory for the subtyping of recursive types sketched in this introduction.

2. A SIMPLY TYPED λ -CALCULUS WITH RECURSIVE TYPES

We consider a simply typed λ -calculus with recursive types and two ground types \perp (bottom) and \top (top); the latter play the roles of least and greatest elements in the subtype relation. Although this calculus is very simple, it already embodies the most interesting problems for which we can provide solutions sufficiently general to extend to other domains. In the Conclusion we comment on which techniques can be applied to more complex calculi.¹

2.1 Types

In an informal BNF notation, types are defined as follows:

$$\begin{aligned} t, s, \dots &\quad \text{type variables and type constants, indifferently} \\ \alpha ::= t \mid \perp \mid \top \mid \alpha \rightarrow \beta \mid \mu t. \alpha. \end{aligned}$$

Types are identified up to renaming of bound variables. We use parentheses to determine precedence; in their absence, \rightarrow associates to the right, and the scoping of μ extends to the right as far as possible. For simplicity we omit the other type constructors considered in the Introduction.

2.2 Terms

Terms are denoted with M, N, \dots ; the following rules establish when a term M has type α (written $M : \alpha$).

¹*Conventions:* \triangleq stands for equality by definition; \equiv for abbreviation or syntactic identification; \vdash precedes a judgment provable in a certain formal system; \supset is the linguistic implication; \Rightarrow is the metalinguistic implication; $[U/x]V$ denotes the substitution of U for x in V .

$$\begin{aligned}
(\text{assmp}) \quad & x^\alpha : \alpha, \\
(\rightarrow I) \quad & M : \beta \Rightarrow (\lambda x^\alpha. M) : \alpha \rightarrow \beta, \\
(\rightarrow E) \quad & M : \alpha \rightarrow \beta, N : \alpha \Rightarrow (MN) : \beta, \\
(\text{fold}) \quad & M : [\mu t. \alpha / t] \alpha \Rightarrow (\text{fold}_{\mu t. \alpha} M) : \mu t. \alpha, \\
(\text{unfold}) \quad & M : \mu t. \alpha \Rightarrow (\text{unfold}_{\mu t. \alpha} M) : [\mu t. \alpha / t] \alpha.
\end{aligned}$$

Hence, terms are either typed variables, typed λ -abstractions, applications, or fold and unfold coercions. The latter should be subscripted with the intended recursive type to facilitate type inference, but these subscripts are sometimes omitted. The fold/unfold coercions are technical devices to explicitly contract or expand the recursive type of a term; that is, such contractions and expansions do not happen automatically.

2.3 Equations

Here are some fundamental equations for the calculus. In particular, notice that the constants “fold” and “unfold” establish an isomorphism between a recursive type and its unfolding.

$$\begin{aligned}
(\beta) \quad & (\lambda x^\alpha. M)N = [N/x^\alpha]M \\
(\mu) \quad & \text{fold}(\text{unfold } x) = x \quad \text{unfold}(\text{fold } x) = x.
\end{aligned}$$

In Section 6 we consider a model in which many more types and terms are equated; for example the following are valid equations:

$$\begin{aligned}
(\text{fold-unfold}) \quad & [\mu t. \alpha / t] \alpha = \mu t. \alpha \\
(\eta) \quad & \lambda x^\alpha. Mx^\alpha = M \quad \text{if } x^\alpha \notin FV(M) \\
(\perp) \quad & x^\perp = y^\perp \\
(\top) \quad & x^\top = y^\top.
\end{aligned}$$

3. TREE ORDERING

There is a well-established theory of subtyping for the nonrecursive types. Basic motivations can be found, for example, in [9]. The notion of nonrecursive type is merely syntactic; it means that the type does not contain μ 's. The purpose of this section is to extend this theory to the recursive types, by defining a notion of approximation on infinite trees.

3.1 Subtyping Nonrecursive Types

We have the following simple rules. There is a least type \perp and a greatest type \top ; the operator \rightarrow is antimonotonic in the first argument and monotonic in the second. The relation \leq is reflexive by virtue of (var) and (\rightarrow) below.

$$\begin{array}{ll}
(\perp) & \perp \leq \alpha \\
(\top) & \alpha \leq \top \\
(\text{var}) & t \leq t \\
(\rightarrow) & \alpha' \leq \alpha, \beta \leq \beta' \Rightarrow \alpha \rightarrow \beta \leq \alpha' \rightarrow \beta'.
\end{array}$$

It is fairly easy to prove that the relation \leq , defined as $\alpha \leq \beta$ iff $\alpha \leq \beta$ is derivable in the system above, is a partial order on the collection of nonrecursive types. In particular, one has to show that the transitivity rule

$$(\text{trans}) \quad \alpha \leq \beta, \beta \leq \gamma \Rightarrow \alpha \leq \gamma$$

is derived. This can be proven by defining a collection of rewriting rules on proofs that have the property that, when applied to a proof using transitivity, produce a (trans)-free proof of the same judgment. More abstractly one can look at the rules as the clauses of an inductive definition of a binary relation \leq and show that such a relation is transitive (see Section 3.4.4).

3.2 Folding and Unfolding

Should the types $[\mu t. \alpha / t] \alpha$ and $\mu t. \alpha$ be considered as equivalent? In general they are provably isomorphic in the calculus via fold and unfold. However, in most languages fold and unfold are implicit, and most implementations do not generate run-time code for them. So it seems reasonable to require that $[\mu t. \alpha / t] \alpha \leq \mu t. \alpha$ and $\mu t. \alpha \leq [\mu t. \alpha / t] \alpha$, thereby making unfolding transparent.

In fact, we will exhibit a model of the calculus in which $\mu t. \alpha$ and $[\mu t. \alpha / t] \alpha$ are equated because recursive domain equations are solved up to equality. However, a theory of type equivalence based only on the congruence closure of

$$(\text{fold-unfold}) \quad [\mu t. \alpha / t] \alpha \leq \mu t. \alpha \quad \mu t. \alpha \leq [\mu t. \alpha / t] \alpha$$

turns out to be too weak; for example, the types $\mu t. s \rightarrow s \rightarrow t$ and $\mu t. s \rightarrow t$ are not equivalent.

Once we assume the transparency of unfolding, it seems natural to consider types with the same infinite expansions as equivalent. Infinite expansion can be rephrased as an *approximation property* such that the semantics of a type is completely determined by the semantics of its finite syntactic approximations. In fact, this is a very desirable property in the semantics of programming languages (see, for example, the approximation theorem in [29]).

3.3 Tree Expansions

As we have seen, simple unfolding does not induce a sufficiently strong notion of type equivalence. A stronger condition of approximation seems required to deal with infinite expansions. Let us explain first how to associate a finitely branching, labeled, regular tree with any recursive type.

Paths in a tree are represented by finite sequences of natural numbers $\pi, \sigma \in \omega^*$, with $\pi\sigma$ for concatenation and nil as the empty sequence.

Nodes in a tree are labeled by a ranked alphabet $L = \{\perp^0, \top^0, \rightarrow^2\} \cup \{t^0 | t \text{ is a type variable}\}$, where the superscripts indicate arity.

A tree $A \in \omega^* \rightarrow L$ is a partial function from (paths) ω^* into (node labels) L , whose domain is nonempty and prefix closed, and such that each node has a number of children equal to the rank of the associated label.

Formally, let $A(\pi) \downarrow$ indicate that π is in the domain of A (and $A(\pi) \uparrow$ indicate the opposite). Then the collection $\text{Tree}(L)$ of finitely branching labeled trees over L is given by the partial maps:

$$\begin{aligned} A: \omega^* \rightarrow L \quad \text{such that:} \\ A(\text{nil}) \downarrow \\ A(\pi\sigma) \downarrow &\Rightarrow A(\pi) \downarrow \\ A(\pi) = p^i &\Rightarrow \forall 0 \leq j < i. A(\pi j) \downarrow. \end{aligned}$$

We can now define a function $T: \text{Type} \rightarrow \text{Tree}(L)$ from recursive types (as defined in Section 2.1) to $\text{Tree}(L)$. Let $h\mu$ be the function that counts the number of μ 's in the head position of a type. We define $T(\alpha)(\pi)$ by induction in $(|\pi|, h\mu(\alpha))$:

$$\begin{aligned} T(\perp)(\text{nil}) &\triangleq \perp \\ T(\top)(\text{nil}) &\triangleq \top \\ T(t)(\text{nil}) &\triangleq t \\ T(\alpha \rightarrow \beta)(\text{nil}) &\triangleq \rightarrow \\ T(\alpha \rightarrow \beta)(0\pi) &\triangleq T(\alpha)(\pi) \\ T(\alpha \rightarrow \beta)(1\pi) &\triangleq T(\beta)(\pi) \\ T(\mu t. \alpha)(\text{nil}) &\triangleq \perp \quad \text{if } \alpha \text{ has the shape } \mu t_1 \dots \mu t_n. t \ (t_i \neq t, i \in 1..n, n \geq 0) \\ T(\mu t. \alpha)(\pi) &\triangleq T([\mu t. \alpha / t](\alpha))(\pi) \quad \text{if } \alpha \text{ does not have the shape above} \\ T(\alpha)(\pi) &\triangleq \uparrow \quad \text{in all other cases.} \end{aligned}$$

Note that the $\alpha \rightarrow \beta$ case reduces $|\pi|$, and the second $\mu t. \alpha$ case preserves $|\pi|$ while reducing $h\mu(\alpha)$; this entails that the definition is well founded.

Here are some simple examples; the tree on the right repeats itself after the "...":

$$T(s \rightarrow \mu t. t) = \begin{array}{c} \rightarrow \\ \swarrow \quad \searrow \\ s \quad \perp \end{array} \qquad T(\mu t. \perp \rightarrow (\top \rightarrow t)) = \begin{array}{c} \rightarrow \\ \swarrow \quad \searrow \\ \perp \quad \rightarrow \\ \quad \swarrow \quad \searrow \\ \quad \top \quad \dots \end{array}$$

Finally, define the collection of finite trees, $\text{Tree}_{\text{fin}}(L)$, as follows:

$$\text{Tree}_{\text{fin}}(L) \triangleq \{A \in \text{Tree}(L) | \exists k. \forall \pi \in \omega^*. |\pi| > k \Rightarrow A(\pi) \uparrow\}.$$

3.3.1. T induces a bijection between $\text{Tree}_{\text{fin}}(L)$ and nonrecursive types. We denote its inverse with T^{-1} .

3.3.2 $\text{Tree}(L)$ is a complete metric space with respect to the usual metric on trees [4]. In fact it is the completion of the space of finite trees, $\text{Tree}_{\text{fin}}(L)$.

We recall:

- A metric space is *complete* iff every Cauchy sequence converges.
- A map $f: M \rightarrow M$ over a metric space M with distance d is *contractive* iff there is a real number $q < 1$ such that $\forall a, b \in M; d(f(a), f(b)) \leq q \cdot d(a, b)$.
- Banach's fixpoint theorem asserts that a contractive map over a complete metric space has a unique fixpoint.
- The distance $d(A, B)$ on $\text{Tree}(L)$ is defined as either 0 if $A = B$; or else $2^{-c(A, B)}$, where $c(A, B)$ is either ∞ if $A = B$, or else it is the length of a shortest path that distinguishes A from B .

3.3.3 For every α , $T\alpha$ is a regular tree, that is, a tree with a finite number of different subtrees. Every tree is completely specified by the language of its occurrences, where if $p \in L$ and $A \in \text{Tree}(L)$ then the occurrences are $\text{Occ}(p, A) \triangleq \{\pi \in \omega^* \mid A(\pi) = p\}$. In particular, every regular tree A has an associated set $\{\pi p \mid \pi \in \text{Occ}(p, A), p \in L\}$ which is a regular language [16]. From this it follows that given types α, β , the problem of deciding if $T\alpha = T\beta$ is reducible to the problem of the equivalence of deterministic finite-state automata.

3.3.4 Going back to the example in 3.2, observe that $T(\mu t.s \rightarrow s \rightarrow t) = T(\mu t.s \rightarrow t)$.

3.4 Finite Approximations

Finite trees are in one-one correspondence with the nonrecursive types; therefore they have a partial order as defined in Section 3.1. The problem we are going to consider now is how to extend this partial order on finite trees to $\text{Tree}(L)$.

Hence, we introduce the notion of finite approximation of a tree. It is crucial to keep in mind the antimonotonic behavior of the \rightarrow in its first argument.

We define a family of functions:

$$\{\downarrow_k : \text{Tree}(L) \rightarrow \text{Tree}_{\text{fin}}(L)\}_{k \in \omega}.$$

Given $A \in \text{Tree}(L)$ its *cut at the k th level* is defined as follows:

$$A_{\downarrow k}(\pi) \triangleq \begin{cases} \uparrow & \text{if } |\pi| > k \\ A(\pi) & \text{if } |\pi| < k, \text{ or } |\pi| = k \text{ and } A(\pi) \uparrow \\ \perp & \text{if } |\pi| = k, A(\pi) \downarrow, \text{ and } \pi \text{ is positive in } A \\ \top & \text{if } |\pi| = k, A(\pi) \downarrow, \text{ and } \pi \text{ is negative in } A \end{cases}$$

where we say that π is positive (negative) in A if along the path π from the root we select the left sibling of a node labeled \rightarrow an even (odd) number of times.

We can extend this definition to types:

$$\alpha_{\downarrow k} \triangleq T^{-1}((T\alpha)_{\downarrow k}) \quad (\text{a nonrecursive type}).$$

Convention. The bijection T, T^{-1} between $\text{Tree}_{\text{fin}}(L)$ and nonrecursive types is from now on often omitted. That is, given any finite tree $A \in \text{Tree}_{\text{fin}}(L)$, we ambiguously identify it with the corresponding nonrecursive type. Similarly, for $A \in \text{Tree}(L)$, we denote with $A_{|k}$ both its cut and the corresponding nonrecursive type.

3.4.1 Definition (tree ordering).

For $A, B \in \text{Tree}_{\text{fin}}(L)$: $A \leq_{\text{fin}} B \Leftrightarrow T^{-1}A \leq T^{-1}B$ (as finite types; see Section 3.1)

For $A, B \in \text{Tree}(L)$: $A \leq_{\infty} B \Leftrightarrow \forall k. (A_{|k} \leq_{\text{fin}} B_{|k})$

For $\alpha, \beta \in \text{Type}$: $\alpha \leq_T \beta \Leftrightarrow T\alpha \leq_{\infty} T\beta$

3.4.2 \leq_{∞} is a partial order on $\text{Tree}(L)$.

3.4.3 $\alpha \leq_T \beta$ is a preorder on recursive types, and is such for all k $\alpha_{|k} \leq_T \alpha$. We can now show, for example, $\alpha \triangleq \mu t. \top \rightarrow t \leq_T \mu t. \perp \rightarrow (\perp \rightarrow t) \triangleq \beta$; consider the tree expansions:



Observe that \top and \perp always occur in negative position; so from $\perp \leq \top$ we can conclude $\forall k. \alpha_{|k} \leq \beta_{|k}$, and this gives us the statement.

3.4.4 One can think of other tree orderings; for example, consider the following inductive definition that gives an ordering \leq_{Ind} on $\text{Tree}(L)$.

\leq_{Ind} is the least *reflexive* relation such that, $\forall A, B, A', B' \in \text{Tree}(F)$:

$$\perp \leq_{\text{Ind}} A; A \leq_{\text{Ind}} \top; A' \leq_{\text{Ind}} A, B \leq_{\text{Ind}} B' \Rightarrow \begin{array}{c} \rightarrow \\ \swarrow \quad \searrow \\ A \quad B \end{array} \leq_{\text{Ind}} \begin{array}{c} \rightarrow \\ \swarrow \quad \searrow \\ A' \quad B' \end{array}$$

Equivalently, $\leq_{\text{Ind}} = \bigcup_{n < \omega} \leq^n$ where:

$$\begin{aligned} \leq^0 &= \{(\perp, A), (A, \top) \mid A \in \text{Tree}(L)\} \cup \text{Id}_{\text{Tree}(L)} \\ \leq^{n+1} &= \leq^n \cup \left\{ \left(\begin{array}{c} \rightarrow \\ \swarrow \quad \searrow \\ A \quad B \end{array}, \begin{array}{c} \rightarrow \\ \swarrow \quad \searrow \\ A' \quad B' \end{array} \right) \mid A' \leq^n A, B \leq^n B' \right\}. \end{aligned}$$

It is not difficult to prove by induction on n that \leq_{Ind} is a partial order on $\text{Tree}(L)$; it conservatively extends the ordering on $\text{Tree}_{\text{fin}}(L)$, and it is contained in \leq_{∞} . Moreover, such containment is strict, as shown by the example in Section 3.4.3. In fact, \leq_{Ind} lacks the crucial approximation property possessed by \leq_{∞} .

4. AN ALGORITHM

In this section we show that the tree ordering we have defined on types (Section 3.4.1) can be decided by a rather natural modification of the algo-

Table I

Type	Equations (v_0 is the root)	A memory representation			
		Address	Node	Child1	Child2
$\mu t. \perp \rightarrow t$	$v_0 = v_1 \rightarrow v_0$ $v_1 = \perp$	0: 1;	\rightarrow \perp	1 —	0 —

rithm that tests directly (that is, without reduction to a minimal form) the tree equivalence of two types.

4.1 Canonical Forms

The first step towards formalizing the algorithm is to introduce canonical forms for types and systems of equations. Canonical forms of types allow us to ignore the trivial type equivalences due to redundant uses of μ binders. For example, the recursive type $(\mu t. \mu s. t \rightarrow s) \rightarrow ((\mu t. t) \rightarrow (\mu t. \top))$ can be simplified to the canonical form $(\mu v. v \rightarrow v) \rightarrow (\perp \rightarrow \top)$ without changing the denoted tree. In a canonical form, the body of each μ is an \rightarrow type, and each μ variable is used in its μ body. Note, however, that different canonical forms may generate the same tree, for example, $\mu t. s \rightarrow t$, $s \rightarrow \mu t. s \rightarrow t$ and $\mu t. s \rightarrow s \rightarrow t$.

Implementations of the subtyping algorithm manipulate cyclic linked data structures in computer memory. We represent these data structures abstractly as special sets of equations. Informally, each equation relates a memory address, represented by a variable, to a node of the data structure, represented by a type constant or a type constructor applied to variables. For example, here is a simple type with a corresponding equational representation and a possible memory representation (Table I).

Sets of equations in this stylized form are called canonical. In this section we show that a canonical set of equations, along with a root variable, determines a unique tree which is called the *solution* of the equations. Moreover, we give effective ways of going from a type to a canonical set of equations, and vice versa, while preserving the represented tree.

Proviso. In order to have a simple correspondence between recursive types and systems of regular equations, we assume that all variables, both bound and free, in the types $\alpha_1, \dots, \alpha_n$ under consideration are distinct. When a type is unfolded, the necessary renaming of bound variables must be performed. For example, $(\mu t. t \rightarrow s) \rightarrow (\mu s. t \rightarrow s)$ should be rewritten as $(\mu v. v \rightarrow s) \rightarrow (\mu r. t \rightarrow r)$.

4.1.1 Recursive Types in Canonical Form. Henceforth, Tp denotes the collection of nonrecursive types, and μTp denotes the collection of *recursive types in canonical form*, defined as follows:

$$\alpha ::= \perp \mid \top \mid t \mid \alpha \rightarrow \beta \mid \mu t. \alpha \rightarrow \beta$$

where in the case $\mu t. \alpha \rightarrow \beta$, t must occur free in $\alpha \rightarrow \beta$. Hence, the body of a μ in canonical form must immediately start with an \rightarrow ; in particular, it

cannot be another μ . The introduction of μTp simplifies the case analysis in the following proofs.

4.1.2 PROPOSITION (Existence of canonical forms). *For every type α there is a type β in canonical form such that $T\alpha = T\beta$.*

PROOF. The crucial observation is that $T\mu t.\mu s.\gamma[t, s] = T\mu v.\gamma[v, v]$. See also Section 5.1.3 for a proof of this fact that uses the rules for type equivalence. \square

4.1.3 Regular System of Equations in Canonical Form. Systems of regular equations are a well-known tool for representing regular trees (see for example [16] and [17]).

For our purposes a *regular system of equations in canonical form* is an element of Tenv , that is, a finite association of distinct type variables (members of Tvar) with types in a specific form:

$\text{Tenv} \triangleq$

$\{\varepsilon \in \text{Tvar} \rightarrow Tp \mid \text{Dom}(\varepsilon) \text{ is finite and } \forall t \in \text{Dom}(\varepsilon) \text{ we have that } \varepsilon(t) \text{ is one of } \perp, \top, t_1, t_2 \rightarrow t_3, \text{ where } t_1 \notin \text{Dom}(\varepsilon) \text{ and } t_2, t_3 \in \text{Dom}(\varepsilon)\}.$

A pair $(\alpha, \varepsilon) \in Tp \times \text{Tenv}$ represents the following *system of regular equations* (not necessarily in canonical form because α may be complex):

$$\begin{aligned} t_\alpha &= \alpha && (t_\alpha \text{ a fresh variable}) \\ t &= \varepsilon(t) && \text{for each } t \in \text{Dom}(\varepsilon). \end{aligned}$$

It is important to observe that, by the definition of Tenv , this system defines a *contractive functional* (G_0, \dots, G_n) over $\text{Tree}(L)^{n+1}$ (see Section 3.3.2) where $n = |\text{Dom}(\varepsilon)|$, $\text{Dom}(\varepsilon) = \{t_1, \dots, t_n\}$ and:

$$\begin{aligned} G_0(A_0, \dots, A_n) &\triangleq [A_0/t_\alpha, A_1/t_1, \dots, A_n/t_n]T\alpha \\ G_i(A_0, \dots, A_n) &\triangleq [A_0/t_\alpha, A_1/t_1, \dots, A_n/s_n]T\varepsilon(t_i) \quad (1 \leq i \leq n). \end{aligned}$$

The predicate $\text{Reach}(\alpha, \varepsilon)$ denotes the variables reachable from the free variables in α by applying the equations in ε . Formally:

$$\begin{aligned} \text{Reach}(t, \varepsilon)^0 &\triangleq \{t\} \\ \text{Reach}(t, \varepsilon)^{n+1} &\triangleq \begin{aligned} &\text{if } t \notin \text{Dom}(\varepsilon) \text{ then } \{t\} \\ &\text{if } t \in \text{Dom}(\varepsilon) \text{ then} \\ &\quad \text{if } \varepsilon(t) = \perp \text{ or } \varepsilon(t) = \top \text{ then } \emptyset \\ &\quad \text{if } \varepsilon(t) = s \text{ then } \{s\} \\ &\quad \text{if } \varepsilon(t) = t_1 \rightarrow t_2 \text{ then } \text{Reach}(t_1, \varepsilon)^n \cup \text{Reach}(t_2, \varepsilon)^n \end{aligned} \\ \text{Reach}(t, \varepsilon) &\triangleq \bigcup_{n \in \omega} \text{Reach}(t, \varepsilon)^n \\ \text{Reach}(\alpha, \varepsilon) &\triangleq \bigcup_{t \in FV(\alpha)} \text{Reach}(t, \varepsilon) \end{aligned}$$

4.1.4 Definition (solution of a system). We denote with $\text{Sol}(\alpha, \varepsilon)$ the first component B_0 of the solution (B_0, \dots, B_n) in $\text{Tree}(L)^{n+1}$ of the system associated with (α, ε) . The solution is given by Banach's unique fixpoint theorem (see Section 3.3.2).

Remark. Given a system of regular equations in canonical form, it is possible to minimize the number of variables by a procedure that is analo-

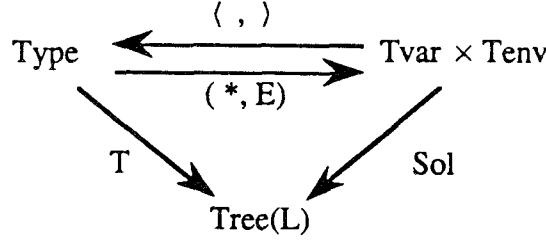


Figure 1

gous to the one for minimizing the number of states in a deterministic finite-state automaton. This immediately provides an algorithm for deciding the equality of the trees represented by two regular systems of equations in canonical form.

In the rest of this section we describe maps between types and regular systems of equations in canonical form, as summarized by the diagram in Figure 1, where all the paths leading from a node to $\text{Tree}(L)$ commute.

4.1.5 PROPOSITION (From recursive types to regular systems). *There is a pair of maps $*$ \in $\text{Type} \rightarrow \text{Tvar}$, $E \in \text{Type} \rightarrow \text{Tenv}$ such that:*

$$\forall \alpha \in \text{Type}. \quad T\alpha = \text{Sol}(\alpha^*, E\alpha).$$

PROOF. It is enough to prove the result for every term in μTp . Then the lemma follows by 4.1.2. We now define $(*, E)$ by induction on the structure of $\gamma \in \mu Tp$.

Cases $\gamma \equiv t$, $\gamma \equiv \perp$, and $\gamma \equiv \top$. Take $\gamma^* \triangleq s$ and $E\gamma \triangleq \{s = \gamma\}$, for any s not appearing in the original type α .

Case $\gamma \equiv \alpha \rightarrow \beta$. We denote with $p[t_1 \dots t_k]$ for $p \in L \setminus \{t_1 \dots t_k\}$ a type in Tp of the form $p(u_1 \dots u_{\#p})$, where $\#p$ is the arity of p , and $\{u_1 \dots u_{\#p}\} \subseteq \{t_1 \dots t_k\}$.

Assume, by induction hypothesis, that $E\alpha = \{t_{i+1} = p_i[t_2 \dots t_{n+1}] \mid i \in 1..n\}$, $\alpha^* = t_2$, $E\beta = \{t_{n+1+j} = q_j[t_{n+2} \dots t_{n+m+1}] \mid j \in 1..m\}$, and $\beta^* = t_{n+2}$. (We require here that $t_1, t_2 \dots t_{n+1}$ and $t_{n+2} \dots t_{n+m+1}$ are disjoint variables not appearing in the original α ; otherwise a consistent renaming must be performed.) Then $E\gamma$ is the following system and $\gamma^* \triangleq t_1$:

$$\begin{aligned} t_1 &= r_1[t_1 \dots t_{n+m+1}] \equiv t_2 \rightarrow t_{n+2}, \\ t_2 &= r_2[t_1 \dots t_{n+m+1}] \equiv p_1[t_2 \dots t_{n+1}] \quad \dots \\ t_{n+1} &= r_{n+1}[t_1 \dots t_{n+m+1}] \equiv p_n[t_2 \dots t_{n+1}] \\ t_{n+2} &= r_{n+2}[t_1 \dots t_{n+m+1}] \equiv q_1[t_{n+2} \dots t_{n+m+1}] \quad \dots \\ t_{n+m+1} &= r_{n+m+1}[t_1 \dots t_{n+m+1}] \equiv q_m[t_{n+2} \dots t_{n+m+1}] \end{aligned}$$

The property $\text{Sol}(\gamma^*, E\gamma) = \text{Sol}(t_1, E\gamma) = \text{Sol}(t_2 \rightarrow t_{n+2}, E\gamma) = \text{Sol}(t_2, E\alpha) \rightarrow \text{Sol}(t_{n+2}, E\beta) = T\gamma$ follows easily from the induction hypothesis.

Case $\gamma \equiv \mu t. \alpha \rightarrow \beta$. Let $\gamma' = [\gamma/t]\alpha \rightarrow [\gamma/t]\beta$ (of course $T\gamma' = T\gamma$). As in the previous case, assume $E\alpha = \{t_{i+1} = p_i[t_2 \dots t_{n+1}] \mid i \in 1..n\}$, $\alpha^* = t_2$, $E\beta = \{t_{n+1+j} = q_j[t_{n+2} \dots t_{n+m+1}] \mid j \in 1..m\}$ and $\beta^* = t_{n+2}$. Then $E\gamma$ is the following system and $\gamma^* \triangleq t_1$:

$$\begin{aligned} t_1 &= r_1[t_1 \dots t_{n+m+1}] \equiv t_2 \rightarrow t_{n+2} \\ t_2 &= r_2[t_1 \dots t_{n+m+1}] \equiv t_2 \rightarrow t_{n+2} \mid p_1[t_2 \dots t_{n+1}] \quad \dots \\ t_{n+1} &= r_{n+1}[t_1 \dots t_{n+m+1}] \equiv t_2 \rightarrow t_{n+2} \mid p_n[t_2 \dots t_{n+1}] \\ t_{n+2} &= r_{n+2}[t_1 \dots t_{n+m+1}] \equiv t_2 \rightarrow t_{n+2} \mid q_1[t_{n+2} \dots t_{n+m+1}] \quad \dots \\ t_{n+m+1} &= r_{n+m+1}[t_1 \dots t_{n+m+1}] \equiv t_2 \rightarrow t_{n+2} \mid q_m[t_{n+2} \dots t_{n+m+1}] \end{aligned}$$

By $t_2 \rightarrow t_{n+2} \mid p_i[t_2 \dots t_{n+1}]$ we denote $t_2 \rightarrow t_{n+2}$ if $p_i \equiv t$, and $p_i[t_2 \dots t_{n+1}]$ otherwise. Analogously, $t_2 \rightarrow t_{n+2} \mid q_j[t_{n+2} \dots t_{n+m+1}]$ denotes $t_2 \rightarrow t_{n+2}$ if $q_j \equiv t$, and $q_j[t_{n+2} \dots t_{n+m+1}]$ otherwise. Next proceed by induction on $(|\pi|, \gamma)$ to prove $T\gamma(\pi) = \text{Sol}(\gamma^*, E\gamma)(\pi)$. The only difficulty arises for $\gamma \equiv \mu t. \alpha \rightarrow \beta$. In order to apply the induction hypothesis one needs a lemma. Following the notation above we show, for instance, $\text{Sol}(t_2, E\gamma) = \text{Sol}([\gamma/t]\alpha^*, E[\gamma/t]\alpha)$. See Section 4.1.7 for a proof of a related fact that gives the main insights while being slightly simpler. \square

Here is an example of the procedure described in the proof above. Consider:

$$\gamma \triangleq \mu t. t \rightarrow \perp.$$

For the base cases t and \perp we have (cunningly choosing the names t_2 and t_3):

$$\begin{aligned} t^* &= t_2; & Et &= \{t_2 = t\} \\ \perp^* &= t_3; & E\perp &= \{t_3 = \perp\}. \end{aligned}$$

From the μ case of the proof we obtain:

$$\gamma^* = t_1; \quad E\gamma = \{t_1 = t_2 \rightarrow t_3, t_2 = t_2 \rightarrow t_3, t_3 = \perp\}.$$

Note the first two equations of the system $E\gamma$; the redundancy facilitates the uniform treatment of the μ case.

4.1.6 Definition (From regular systems to recursive types). We define a function $\langle -, \varepsilon \rangle: Tp \times \text{Tenv} \rightarrow \text{Type}$ by induction on $(|\text{Dom}(\varepsilon)|, \alpha)$:

$$\begin{aligned} \langle \perp, \varepsilon \rangle &\triangleq \perp \\ \langle \top, \varepsilon \rangle &\triangleq \top \\ \langle \alpha \rightarrow \beta, \varepsilon \rangle &\triangleq \langle \alpha, \varepsilon \rangle \rightarrow \langle \beta, \varepsilon \rangle \\ \langle t, \varepsilon \rangle &\triangleq t && \text{if } t \notin \text{Dom}(\varepsilon) \\ \langle t, \varepsilon \rangle &\triangleq \mu t. \langle \varepsilon(t), \varepsilon \setminus t \rangle && \text{if } t \in \text{Dom}(\varepsilon) \end{aligned}$$

where $\varepsilon \setminus t$ is like ε except that it is undefined on t .

Continuing the example above, we have:

$$\begin{aligned} \langle \gamma^*, E\gamma \rangle &= \langle t_1, \{t_1 = t_2 \rightarrow t_3, t_2 = t_2 \rightarrow t_3, t_3 = \perp\} \rangle \\ &= \mu t_1. \langle t_2 \rightarrow t_3, \{t_2 = t_2 \rightarrow t_3, t_3 = \perp\} \rangle \\ &= \mu t_1. \langle t_2, \{t_2 = t_2 \rightarrow t_3, t_3 = \perp\} \rangle \rightarrow \langle t_3, \{t_2 = t_2 \rightarrow t_3, t_3 = \perp\} \rangle \end{aligned}$$

$$\begin{aligned}
&= \mu t_1.(\mu t_2.\langle t_2 \rightarrow t_3, \{t_3 = \perp\} \rangle) \rightarrow (\mu t_3.\langle \perp, \{t_2 = t_2 \rightarrow t_3\} \rangle) \\
&= \mu t_1.(\mu t_2.\langle t_2, \{t_3 = \perp\} \rangle \rightarrow \langle t_3, \{t_3 = \perp\} \rangle) \rightarrow (\mu t_3.\perp) \\
&= \mu t_1.(\mu t_2 t_2 \rightarrow (\mu t_3.\langle \perp, \{ \} \rangle)) \rightarrow (\mu t_3.\perp) \\
&= \mu t_1.(\mu t_2.t_2 \rightarrow (\mu t_3.\perp)) \rightarrow (\mu t_3.\perp).
\end{aligned}$$

The last line is equivalent to the original type $\gamma = \mu t.t \perp$, as established in general by the following proposition.

4.1.7 PROPOSITION (More on commuting translations).

- (1) *For any system of equations, the first component of the solution coincides with the tree expansion of the associated recursive type:*

$$\begin{aligned}
\forall(\alpha, \varepsilon) \in T_p \times Tenv. \text{Sol}(\alpha, \varepsilon) &= T\langle \alpha, \varepsilon \rangle \\
&\quad (\text{we abbreviate } T\langle \alpha, \varepsilon \rangle \text{ as } T\langle \alpha, \varepsilon \rangle).
\end{aligned}$$

- (2) *The map $\langle \cdot, \cdot \rangle$ satisfies the conditions:*

1. $\langle \perp, \varepsilon \rangle = \perp$
2. $\langle \top, \varepsilon \rangle = \top$
3. $\langle t, \varepsilon \rangle = t$ if $t \notin \text{Dom}(\varepsilon)$
4. $T\langle t, \varepsilon \rangle = T\langle \varepsilon(t), \varepsilon \rangle$ if $t \in \text{Dom}(\varepsilon)$
5. $T\langle \alpha \rightarrow \beta, \varepsilon \rangle = T(\langle \alpha, \varepsilon \rangle \rightarrow \langle \beta, \varepsilon \rangle)$
6. $T\langle \alpha^*, E\alpha \rangle = T\alpha$.

PROOF OF PART (1). Show by induction on $(|\pi|, \alpha)$ that $T\langle \alpha, \varepsilon \rangle(\pi) = \text{Sol}(\alpha, \varepsilon)(\pi)$. The interesting case arises when $\alpha \equiv t$, $t \in \text{Dom}(\varepsilon)$, $\varepsilon(t) = t_1 \rightarrow t_2$. Then, $\text{Sol}(t, \varepsilon) = \text{Sol}(\varepsilon(t), \varepsilon) = \text{Sol}(t_1 \rightarrow t_2, \varepsilon)$; and, $T\langle t, \varepsilon \rangle = T\mu t.\langle \varepsilon(t), \varepsilon \setminus t \rangle = T[\langle t, \varepsilon \rangle / t] \langle t_1 \rightarrow t_2, \varepsilon \setminus t \rangle$. In order to apply the induction hypothesis and complete this case we need to prove $T\langle t_i, \varepsilon \rangle = T([\langle t, \varepsilon \rangle / t] \langle t_i, \varepsilon \setminus t \rangle)$ ($i = 1, 2$). To obtain the latter, we show the following lemma:

*For any canonical system ε , and type variables, t, t' ,
we have $T\langle t', \varepsilon \rangle = T([\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \setminus t \rangle)$.*

We proceed by induction on the depth of the path π , and by case analysis, to show:

$$T\langle t', \varepsilon \rangle(\pi) = T([\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \setminus t \rangle)(\pi).$$

Case $t \equiv t'$. $[\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \setminus t \rangle = [\langle t, \varepsilon \rangle / t] t' = \langle t, \varepsilon \rangle = \langle t', \varepsilon \rangle$.

Case $t \neq t'$.

Subcase $t \in \text{Dom}(\varepsilon)$, $t' \in \text{Dom}(\varepsilon)$.

Say: $\varepsilon(t) = t_1 \rightarrow t_2$, $\varepsilon(t') = t'_1 \rightarrow t'_2$.

Then: $\alpha \equiv \langle t', \varepsilon \rangle = \mu t'. \langle t'_1, \varepsilon \setminus t' \rangle \rightarrow \langle t'_2, \varepsilon \setminus t' \rangle$.

and $T\alpha = \rightarrow$

$$\begin{aligned}
&\quad / \quad \backslash \\
&T([\langle t', \varepsilon \rangle / t'] \langle t'_1, \varepsilon \setminus t' \rangle) \quad T([\langle t', \varepsilon \rangle / t'] \langle t'_2, \varepsilon \setminus t' \rangle)
\end{aligned}$$

$$\begin{aligned} \text{Also: } \beta &\equiv [\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \setminus t \rangle = [\langle t, \varepsilon \rangle / t] \mu t'. \langle t'_1, \varepsilon \setminus t \setminus t' \rangle \rightarrow \\ &\quad \langle t'_2, \varepsilon \setminus t \setminus t' \rangle \\ &= [\langle t, \varepsilon \rangle / t] ([\langle t', \varepsilon \setminus t \rangle / t'] (\langle t'_1, \varepsilon \setminus t \setminus t' \rangle \rightarrow \langle t'_2, \varepsilon \setminus t \setminus t' \rangle)). \\ \text{So: } T\beta &= \rightarrow \\ &\quad T[\langle t, \varepsilon \rangle / t] ([\langle t', \varepsilon \setminus t \rangle / t'] (\langle t'_1, \varepsilon \setminus t \setminus t' \rangle)) \quad / \quad \backslash \\ &\quad \quad \quad T[\langle t, \varepsilon \rangle / t] \\ &\quad \quad \quad ([\langle t', \varepsilon \setminus t \rangle / t'] \\ &\quad \quad \quad (\langle t'_2, \varepsilon \setminus t \setminus t' \rangle)). \end{aligned}$$

If $n\pi$ is the current path, we can apply the inductive hypothesis on the shorter path π w.r.t.:

(a) the variables t', t'_i , ($i = 1, 2$) and the system ε to show:

$$T\langle t'_i, \varepsilon \rangle(\pi) = T([\langle t', \varepsilon \rangle / t'] \langle t'_i, \varepsilon \setminus t' \rangle(\pi)$$

(b) the variables t', t'_i , ($i = 1, 2$) and the system $\varepsilon \setminus t$ to show:

$$T\langle t'_i, \varepsilon \setminus t \rangle(\pi) = T[\langle t', \varepsilon \setminus t \rangle / t'](\langle t'_i, \varepsilon \setminus t \setminus t' \rangle)(\pi)$$

(c) the variables t, t'_i , ($i = 1, 2$) and the system ε to show:

$$T[\langle t, \varepsilon \rangle / t] \langle t'_i, \varepsilon \setminus t \rangle (\pi) = T \langle t'_i, \varepsilon \rangle (\pi).$$

Finally, we use the substitutivity of the T operation, $T[\gamma/t]\delta = [T\gamma/t]T\delta$, to conclude $T\alpha = T\beta$.

Subcase $t \in \text{Dom}(\varepsilon)$, $t' \notin \text{Dom}(\varepsilon)$.

Say: $\varepsilon(t) = t_1 \rightarrow t_2$.

Then: $T\langle t', \varepsilon \rangle = t'$

$$T([\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \rangle) = T([\langle t, \varepsilon \rangle / t] \ t' = t').$$

Subcase $t \notin \text{Dom}(\varepsilon)$, $t' \in \text{Dom}(\varepsilon)$.

Say: $\varepsilon(t') = t'_1 \rightarrow t'_2$.

Then: $T([\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \setminus t \rangle) = T([t / t] \langle t', \varepsilon \setminus t \rangle) = T \langle t', \varepsilon \setminus t \rangle = T \langle t', \varepsilon \rangle$.

Subcase $t \notin \text{Dom}(\varepsilon)$, $t' \notin \text{Dom}(\varepsilon)$:

Then: $T([\langle t, \varepsilon \rangle / t] \langle t', \varepsilon \setminus t \rangle) = T[t/t] t' = Tt' = T\langle t', \varepsilon \rangle$.

PROOF OF PART (2). Conditions 1, 2, 3, and 5 follow by definition. Condition 4 follows from $\text{Sol}(t, \varepsilon) = \text{Sol}(\varepsilon(t), \varepsilon)$ and part (1). Condition 6 follows from Prop. 4.1.5 and part (1). $T\alpha = \text{Sol}(\alpha^*, E\alpha) = T\langle \alpha^*, E\alpha \rangle$. \square

4.2 Computational Rules

The subtyping algorithm described in this section is based on the canonical sets of equations described in the previous section (again, these equations can be interpreted as linked data structures in memory). The algorithm involves a single set of equations ε , with two distinct roots α and β representing the types to be compared. It also involves a *trail* Σ of the form $\{t_1 \leq s_1, \dots, t_n \leq$

$s_n\}$, which records inclusions of variables discovered as the algorithm progresses. An invocation of the algorithm with parameters Σ , ε , α , and β is written as the *judgment* $\Sigma, \varepsilon \supset \alpha \leq \beta$.

The algorithm is not expressed as an ordinary procedure, but as a collection of rules that resembles a Prolog program. The typical rule is written as a logical implication of judgments:

$$\Sigma_1, \varepsilon_1 \supset \alpha_1 \leq \beta_1, \quad \Sigma_2, \varepsilon_2 \supset \alpha_2 \leq \beta_2 \Rightarrow \Sigma, \varepsilon \supset \alpha \leq \beta.$$

Operationally, this means that in order to determine whether $\Sigma, \varepsilon \supset \alpha \leq \beta$ holds we must invoke the “subroutines” $\Sigma_1, \varepsilon_1 \supset \alpha_1 \leq \beta_1$ and $\Sigma_2, \varepsilon_2 \supset \alpha_2 \leq \beta_2$ and check whether they hold. In general, given a logical deduction in this system of rules, the algorithm execution can be recovered by reading the rules backwards from the conclusion to the assumptions.

In the following, t, s, r, u denote arbitrary variables; a, b denote variables not in the domain of ε ; Σ is a finite set of subtyping assumptions on pairs of type variables; and $\alpha, \beta \in Tp$.

The algorithm can then be written as follows:

$$\begin{aligned} (\text{assmp}_A) \quad & \Sigma \varepsilon \supset t \leq s \quad \text{if } t \leq s \in \Sigma \\ (\perp_A) \quad & \Sigma, \varepsilon \supset \perp \leq \beta \\ (\top_A) \quad & \Sigma, \varepsilon \supset \alpha \leq \top \\ (\text{var}_A) \quad & \Sigma, \varepsilon \supset a \leq a \\ (\rightarrow_A) \quad & \Sigma, \varepsilon \supset \alpha' \leq \alpha, \Sigma, \varepsilon \supset \beta \leq \beta' \Rightarrow \Sigma, \varepsilon \supset \alpha \rightarrow \beta \leq \alpha' \rightarrow \beta' \\ (\mu_A) \quad & \Sigma \cup \{t \leq s\}, \varepsilon \supset \varepsilon(t) \leq \varepsilon(s) \Rightarrow \Sigma, \varepsilon \supset t \leq s \text{ if } t, s \in \text{Dom}(\varepsilon). \end{aligned}$$

The initial judgment $\Sigma, \varepsilon \supset \alpha \leq \beta$ that starts an execution of the algorithm must obey a special condition expressing some reasonable assumptions. This condition says (1) that the initial type structures α, β are simple root variables denoting disjoint structures and (2) that Σ has not yet come into play. For $\Sigma = \{t_1 \leq s_1, \dots, t_n \leq s_n\}$, define:

$$\begin{aligned} \text{Vars}(\Sigma) &\triangleq \{t_1, s_1, \dots, t_n, s_n\} \\ \Sigma \nVdash \varepsilon &\Leftrightarrow \text{Vars}(\Sigma) \cap \text{Dom}(\varepsilon) = \emptyset. \end{aligned}$$

Then, a judgment $\Sigma, \varepsilon \supset \alpha \leq \beta$ satisfies the *initiality condition* (or equivalently, is an *initial goal*) iff $\alpha \equiv t$, $\beta \equiv s$, ε can be decomposed in $\varepsilon_1 \cup \varepsilon_2$ so that $t \in \text{Dom}(\varepsilon_1)$ and $s \in \text{Dom}(\varepsilon_2)$, $\text{Dom}(\varepsilon_1) \cap \text{Dom}(\varepsilon_2) = \emptyset$, and $\Sigma \nVdash \varepsilon$.

By the way canonical systems are constructed and by the fact of starting with an initial goal, the expansions of variables according to ε , as in (μ_A) , is always synchronized. That is, in a call to $\Sigma, \varepsilon \supset \alpha \leq \beta$ during the execution of the algorithm we never have a situation where α is a variable in $\text{Dom}(\varepsilon)$ and β is not, or vice versa; hence (μ_A) covers all the cases that may arise. If one desires to treat more general systems of equations, then it may be necessary to introduce other μ -rules that take into account situations in which just an ε -expansion on the left (or the right) is needed. In these cases we would have

rules like:

$$\begin{aligned}
 (\text{assmp}'_A) \quad & \Sigma, \varepsilon \supset \alpha \leq \beta \quad \text{if } \alpha \leq \beta \in \Sigma \\
 (\mu_{lA}) \quad & \Sigma \cup \{t \leq \alpha' \rightarrow \beta'\}, \varepsilon \supset \varepsilon(t) \leq \alpha' \rightarrow \beta' \Rightarrow \Sigma, \varepsilon \supset t \leq \alpha' \rightarrow \beta' \\
 & \quad \text{if } t \in \text{Dom}(\varepsilon) \\
 (\mu_{rA}) \quad & \Sigma \cup \{\alpha' \rightarrow \beta' \leq s\}, \varepsilon \supset \alpha' \rightarrow \beta' \leq \varepsilon(s) \Rightarrow \Sigma, \varepsilon \supset \alpha' \rightarrow \beta' \leq s \\
 & \quad \text{if } s \in \text{Dom}(\varepsilon).
 \end{aligned}$$

Note also that there are two conceptually distinct uses of the rule (assmp'_A) in the algorithm: one for the initial assumptions contained in Σ , which represent known inclusions on type constants, and one for the assumptions inserted during the computation, which come from the unfolding of μ 's.

4.2.1 Generating the Execution Tree. Given a goal $\Sigma, \varepsilon \supset t \leq s$, the algorithm consists in applying the inference rules backwards, generating subgoals in the cases (\rightarrow_A) and (μ_A) . This process is completely determined once we establish that (1) (assmp'_A) has priority over the other rules and (2) (\perp_A) has priority over (\top_A) .

A tree of goals built this way is called an *execution tree*. If no rules are applicable to a certain subgoal, that branch of the execution tree is abandoned, and execution is resumed at the next subgoal, until all subgoals are exhausted.

4.2.2 Termination. The execution tree is always *finite*. Observe that if $t \leq s$ is the assumption that we add to Σ , then t and s are type variables in $\text{Dom}(\varepsilon)$. Also observe that the (\rightarrow) rule shrinks the size of the current goal by replacing it with subexpressions of the goal and that each application of a μ -rule enlarges Σ .

The bound on the depth of the execution tree for $\alpha \leq_A \beta$ is of the order of the product of the sizes of the two systems $E\alpha, E\beta$.

4.2.3 Algorithm Ordering. An execution tree *succeeds* if all the leaves correspond to an application of one of the rules (assmp'_A) , (\perp_A) , (\top_A) , and (var_A) . Dually, it *fails* if at least one leaf is an unfulfilled goal (no rule can be applied).

We have $\vdash_A \Sigma, \varepsilon \supset t \leq s$ iff (1) $\Sigma, \varepsilon \supset t \leq s$ is an initial goal (Section 4.2) and (2) the corresponding execution tree succeeds.

Given recursive types α, β we have:

$$\alpha \leq_A \beta \Leftrightarrow \vdash_A \emptyset, \quad E\alpha \cup E\beta \supset \alpha^* \leq \beta^*.$$

For testing type equality, we can define:

$$\alpha =_A \beta \Leftrightarrow \alpha \leq_A \beta \wedge \beta \leq_A \alpha.$$

Alternatively, we could directly define a (more efficient) type equality algorithm, along the same lines as the subtyping algorithm.

4.3 Soundness and Completeness of the Algorithm

We now show that the subtyping algorithm described in the previous section is sound and complete with respect to the infinite-tree interpretation of types.

That is, the algorithm precisely embodies our intuition of recursive types as infinite trees.

First we prove soundness and completeness for nonrecursive types. Soundness is then derived by observing that a successful execution of the algorithm on some input must also be successful on all the finite approximations of the input. Completeness is proven by examining a failing execution tree and concluding that the trees corresponding to the input must have been different to start with.

4.3.1 LEMMA (Derived structural computational rules). *Given the definition of $Tenv$ in Section 4.1.3, the algorithm in 4.2, and the ordering in 4.2.3, we have:*

Σ -weaken

If $\vdash_A \Sigma, \varepsilon \supset t \leq s$ and $\Sigma \cup \Sigma' \Vdash \varepsilon$ then $\vdash_A \Sigma \cup \Sigma', \varepsilon \supset t \leq s$.

Σ -strengthen

If $\vdash_A \Sigma \cup \Sigma', \varepsilon \supset t \leq s$ and $Reach(t \rightarrow s, \varepsilon) \cap Vars(\Sigma') = \emptyset$ then $\vdash_A \Sigma, \varepsilon \supset t \leq s$.

ε -weaken

If $\vdash_A \Sigma, \varepsilon \supset t \leq s$, $Reach(t \rightarrow s, \varepsilon) \cap Dom(\varepsilon') = \emptyset$, $\Sigma \Vdash \varepsilon \cup \varepsilon'$, and $Dom(\varepsilon) \cap Dom(\varepsilon') = \emptyset$ then $\vdash_A \Sigma, \varepsilon \cup \varepsilon' \supset t \leq s$.

ε -strengthen

If $\vdash_A \Sigma, \varepsilon \cup \varepsilon' \supset t \leq s$ and $Reach(t \rightarrow s, \varepsilon) \cap Dom(\varepsilon') = \emptyset$ then $\vdash_A \Sigma, \varepsilon \supset t \leq s$.

4.3.2 PROPOSITION (Completeness of \leq_A for nonrecursive types). *Given $\alpha, \beta \in Tp$ nonrecursive types, then $\alpha \leq_T \beta \Rightarrow \alpha \leq_A \beta$.*

PROOF. Let $\varepsilon \triangleq E\alpha \cup E\beta$. We show $\alpha \leq_T \beta \Rightarrow \forall \Sigma. \Sigma \Vdash \varepsilon \Rightarrow \vdash_A \Sigma, \varepsilon \supset \alpha^* \leq \beta^*$ by induction on the structure of α and β .

Case $\alpha \equiv \perp$. Then $\varepsilon = \{\alpha^* = \perp\} \cup E\beta$. Take any Σ s.t. $\Sigma \Vdash \varepsilon$:

$$\begin{aligned} &\Rightarrow \Sigma \cup \{\alpha^* \leq \beta^*\}, \varepsilon \supset \varepsilon(\alpha^*) \leq \varepsilon(\beta^*) && \text{by } (\perp_A) \text{ since } \varepsilon(\alpha^*) = \perp \\ &\Rightarrow \Sigma, \varepsilon \supset \alpha^* \leq \beta^* && \text{by } (\mu_A) \text{ since } \\ & && \alpha^*, \beta^* \notin Vars(\Sigma). \end{aligned}$$

Cases $\alpha \equiv \top$, $\alpha \equiv a$. Similar.

Case $\alpha \equiv \alpha' \rightarrow \alpha''$. Since $\alpha \leq_T \beta$, we have either:

Case $\beta \equiv \top$, similar to the case $\alpha \equiv \perp$.

Case $\beta \equiv \beta' \rightarrow \beta''$, with $\beta' \leq_T \alpha'$ and $\alpha'' \leq_T \beta''$.

Then $\varepsilon = \{\alpha^* = \alpha'^* \rightarrow \alpha''^*\} \cup \{\beta^* = \beta'^* \rightarrow \beta''^*\} \cup \varepsilon' \cup \varepsilon''$ where $\varepsilon' \triangleq E\alpha' \cup E\beta'$ and $\varepsilon'' \triangleq E\alpha'' \cup E\beta''$.

By induction hypothesis $\forall \Sigma'. \Sigma' \Vdash \varepsilon' \Rightarrow \vdash_A \Sigma', \varepsilon' \supset \beta'^* \leq \alpha'^*$ and $\forall \Sigma''. \Sigma'' \Vdash \varepsilon'' \Rightarrow \vdash_A \Sigma'', \varepsilon'' \supset \alpha''^* \leq \beta''^*$.

Take any Σ such that $\Sigma \Vdash \varepsilon$ then:

$$\vdash_A \Sigma, \varepsilon' \supset \beta'^* \leq \alpha'^* \text{ and } \vdash_A \Sigma, \varepsilon'' \supset \alpha''^* \leq \beta''^*.$$

By ε -weaken (note $\Sigma \Vdash \varepsilon \Rightarrow \Sigma \Vdash \varepsilon' \cup \varepsilon''$):

$$\vdash_A \Sigma, \varepsilon' \cup \varepsilon'' \supset \beta'^* \leq \alpha'^* \text{ and } \vdash_A \Sigma, \varepsilon' \cup \varepsilon'' \supset \alpha''^* \leq \beta''^*.$$

By Σ -weaken:

$$\vdash_A \Sigma \cup \{\alpha^* \leq \beta^*\}, \varepsilon' \cup \varepsilon'' \supset \beta'^* \leq \alpha'^* \text{ and } \vdash_A \Sigma \cup \{\alpha^* \leq \beta^*\}, \\ \varepsilon' \cup \varepsilon'' \supset \alpha''^* \leq \beta''^*.$$

Hence, by applying (\rightarrow_A) and (μ_A) we can conclude $\vdash_A \Sigma, \varepsilon \supset \alpha^* \leq \beta^*$.

□

4.3.3 PROPOSITION (Soundness of \leq_A for nonrecursive types). *Given $\alpha, \beta \in Tp$ nonrecursive types, then $\alpha \leq_A \beta \Rightarrow \alpha \leq_T \beta$.*

PROOF. We show $\vdash_A \emptyset, \varepsilon \supset \alpha^* \leq \beta^* \Rightarrow \alpha \leq_T \beta$, where $\varepsilon \triangleq E\alpha \cup E\beta$, by induction on the structure of α and β .

Case $\alpha \equiv \perp$. Then $\perp \leq_A \beta$ by (\perp_A) ((assmp_A) does not apply), and also $\perp \leq_T \beta$.

Cases $\alpha \equiv \top, \alpha \equiv a$. Similar.

Case $\alpha \equiv \alpha' \rightarrow \alpha''$. Assume $\vdash_A \emptyset, \varepsilon \supset \alpha^* \leq \beta^*$; then the first step is either:

Case (\top_A) . Similar to the case $\alpha \equiv \perp$.

Case (μ_A) . Then the second step is (\rightarrow_A) , that is also $\beta \equiv \beta' \rightarrow \beta''$

with $\{\alpha^* \leq \beta^*\}, \varepsilon \supset \beta'^* \leq \alpha'^*$ and $\{\alpha^* \leq \beta^*\}, \varepsilon \supset \alpha''^* \leq \beta''^*$,
where $\varepsilon = \{\alpha^* = \alpha'^* \rightarrow \alpha''^*\} \cup \{\beta^* = \beta'^* \rightarrow \beta''^*\} \cup \varepsilon' \cup \varepsilon''$ and
 $\varepsilon' \triangleq E\alpha' \cup E\beta'$ and $\varepsilon'' \triangleq E\alpha'' \cup E\beta''$.

Since α, β contain no μ , $\text{Reach}(\beta'^* \rightarrow \alpha'^*, \varepsilon) \cap (\text{Dom}(\varepsilon'') \cup \{\alpha^*, \beta^*\}) = \emptyset$.

By a simple analysis we have: $\vdash_A \{\alpha^* \leq \beta^*\}, \varepsilon' \cup \varepsilon'' \supset \beta'^* \leq \alpha'^*$.

By ε -strengthen $\vdash_A \{\alpha^* \leq \beta^*\}, \varepsilon' \supset \beta'^* \leq \alpha'^*$. Similarly, $\vdash_A \{\alpha^* \leq \beta^*\}, \varepsilon'' \supset \alpha''^* \leq \beta''^*$.

Now, by Σ -strengthen $\vdash_A \emptyset, \varepsilon' \supset \beta'^* \leq \alpha'^*$ and $\vdash_A \emptyset, \varepsilon'' \supset \alpha''^* \leq \beta''^*$.

By induction hypothesis $\beta' \leq_T \alpha'$ and $\alpha'' \leq_T \beta''$; hence, $\alpha' \rightarrow \alpha'' \leq_T \beta' \rightarrow \beta''$. □

4.3.4 LEMMA (Uniformity of \leq_A). *Let $\alpha, \beta \in Type$. If $\alpha \leq_A \beta$ then $\forall k. \alpha|_k \leq_A \beta|_k$.*

PROOF (SKETCH). Given any k , from the execution tree of $\alpha \leq_A \beta$ it is possible to extract a successful execution tree for $\alpha|_k \leq_A \beta|_k$. The point is that the use of the (assmp_A) rule can be arbitrarily delayed by repeating a certain pattern of computation.

For example, consider $\mu t. \top \rightarrow t \leq_A \mu s. \perp \rightarrow s$, which gives rise to:

$$\varepsilon \triangleq \varepsilon_1 \cup \varepsilon_2, \varepsilon_1 \triangleq \{t_1 = t_2 \rightarrow t_1, t_2 = \top\}, \varepsilon_2 \triangleq \{s_1 = s_2 \rightarrow s_1, s_2 = \perp\}.$$

The execution tree of the initial goal $\emptyset, \varepsilon \supset t_1 \leq s_1$ is:

$$\begin{aligned}
 & (\perp_A) \\
 \Rightarrow & \{t_1 \leq s_1, s_2 \leq t_2\}, \varepsilon \supset \perp \leq \top & (\text{assmp}_A) \\
 \Rightarrow & \{t_1 \leq s_1\}, \varepsilon \supset s_2 \leq t_2 & \Rightarrow \{t_1 \leq s_1\}, \varepsilon \supset t_1 \leq s_1 \\
 & \Rightarrow \{t_1 \leq s_1\}, \varepsilon \supset t_2 \rightarrow t_1 \leq s_2 \rightarrow s_1 \\
 & \Rightarrow \emptyset, \varepsilon \supset t_1 \leq s_1
 \end{aligned}$$

The goal under (assmp_A) can be replaced by a copy of the entire tree, appropriately renamed. At the same time, ε must be appropriately expanded:

$$\begin{aligned}
 \underline{\varepsilon} & \triangleq \underline{\varepsilon}_1 \cup \underline{\varepsilon}_2, \\
 \underline{\varepsilon}_1 & \triangleq \{t_1 = t_2 \rightarrow u_1, t_2 = \top, u_1 = u_2 \rightarrow u_1, u_2 = \top\}, \\
 \underline{\varepsilon}_2 & \triangleq \{s_1 = s_2 \rightarrow v_1, s_2 = \perp, v_1 = v_2 \rightarrow v_1, v_2 = \perp\}. \\
 & (\perp_A) \\
 \Rightarrow & \{t_1 \leq s_1, u_1 \leq v_1, v_2 \leq u_2\}, \underline{\varepsilon} \supset \perp \leq \top & (\text{assmp}_A) \\
 \Rightarrow & \{t_1 \leq s_1, u_1 \leq v_1\}, \underline{\varepsilon} \supset v_2 \leq u_2 & \Rightarrow \{t_1 \leq s_1, u_1 \leq v_1\}, \underline{\varepsilon} \supset u_1 \leq v_1 \\
 & \Rightarrow \{t_1 \leq s_1, u_1 \leq v_1\}, \underline{\varepsilon} \supset u_2 \rightarrow u_1 \leq v_2 \rightarrow v_1 \\
 & \Rightarrow \{t_1 \leq s_1\}, \underline{\varepsilon} \supset u_1 \leq v_1 \\
 & (\perp_A) \quad \vdots \\
 \Rightarrow & \{t_1 \leq s_1, s_2 \leq t_2\}, \underline{\varepsilon} \supset \perp \leq \top \\
 \Rightarrow & \{t_1 \leq s_1\}, \underline{\varepsilon} \supset s_2 \leq t_2 \\
 & \Rightarrow \{t_1 \leq s_1\}, \underline{\varepsilon} \supset t_2 \rightarrow u_1 \leq s_2 \rightarrow v_1 \\
 & \Rightarrow \emptyset, \underline{\varepsilon} \supset t_1 \leq s_1
 \end{aligned}$$

This is now the execution tree of a different initial goal, which might have originated from the problem $\top \rightarrow (\mu t. \top \rightarrow t) \leq_A \perp \rightarrow \mu s. (\perp \rightarrow s)$, which is equivalent to the original problem.

In a similar way, this execution tree can be further transformed into one for $\top \rightarrow (\top \rightarrow \perp) \leq_A \perp \rightarrow (\perp \rightarrow \perp)$ by replacing the (assmp_A) leaf with a (\perp_A) leaf. By repeating this process we can obtain an execution tree for $\alpha_{|k} \leq_A \beta_{|k}$, for an arbitrarily large k . \square

4.3.5 PROPOSITION (Soundness of \leq_A). *Let $\alpha, \beta \in \text{Type}$; if $\alpha \leq_A \beta$ then $\alpha \leq_T \beta$.*

PROOF. From 4.3.4 we have: $\alpha \leq_A \beta \Rightarrow \forall k. \alpha_{|k} \leq_A \beta_{|k}$. From 4.3.3 and the definition of \leq_T we have: $\forall k. \alpha_{|k} \leq_{\text{fin}} \beta_{|k}$ and $\alpha \leq_T \beta$. \square

4.3.6 LEMMA (Faithfulness of \leq_A w.r.t. paths). *Let $\text{lead}(\alpha, \varepsilon) \triangleq \text{Sol}(\alpha, \varepsilon)(\text{nil})$ be the first label of α in ε (that is, skipping initial variables in α, ε).*

Let $\Sigma, \varepsilon \supset \alpha \leq \beta$ be the root of an execution tree, terminating with success or failure leaves, obtained from the rules in 4.2. Every node $\Sigma', \varepsilon \supset \alpha' \leq \beta'$ in

Table II

$\alpha' \setminus \beta'$	\perp	\top	s	b	$\beta' \rightarrow \beta''$
\perp	\perp	\perp	n.a.	\perp	\perp
\top	fail	\top	n.a.	fail	fail
t	n.a.	n.a.	assmp- μ	n.a.	n.a.
a	fail	\top	n.a.	var-fail	fail
$\alpha' \rightarrow \alpha''$	fail	\top	n.a.	fail	\rightarrow

the execution tree determines a path π from the root to itself, given by considering the occurrences of (\rightarrow_A) and ignoring the other rules. Then:

- (1) Either α' and β' are both (bound) type variables, or neither is.
- (2) $T\alpha(\pi) = \text{lead}(\alpha', \varepsilon)$ and $T\beta(\pi) = \text{lead}(\beta', \varepsilon)$.

PROOF. By induction on the depth of the execution tree. \square

4.3.7 PROPOSITION (Completeness of \leq_A). *Let $\alpha, \beta \in \text{Type}$; if $\alpha \leq_T \beta$ then $\alpha \leq_A \beta$.*

PROOF. We show $\neg \alpha \leq_A \beta \Rightarrow \neg T\alpha \leq_\infty T\beta$. By assumption, we have an execution tree for $\alpha \leq_A \beta$ which contains a failure node $\Sigma, \varepsilon \supset \alpha' \leq \beta'$, determining a path π as in Lemma 4.3.6. By 4.3.6.(2), $T\alpha(\pi) = \text{lead}(\alpha', \varepsilon)$ and $T\beta(\pi) = \text{lead}(\beta', \varepsilon)$. Hence we have a *common* path in $T\alpha$ and $T\beta$ corresponding to the failure node. Table II summarizes the possible cases for α', β' where the entry indicates either failure or the rule being applied by the algorithm; the n.a. (not applicable) cases come from 4.3.6.(1).

Every “fail” in the algorithm corresponds to a situation where the two trees cannot be in the inclusion relation. \square

4.4 An Implementation

In order to facilitate the proofs, the representation of data structures and algorithms given in 4.1 and 4.2 was rather abstract. In this section we show the beginning of a similar treatment for more concrete and traditional representations.

The computational rules in 4.2 can be converted into a straightforward and practical algorithm, based on the method of *trails* [28]. To reflect more closely actual implementations, we adopt the additional rules (assmp'_A), (μ_{lA}), and (μ_{rA}) described in 4.2. This results in the algorithm discussed informally in 1.5, but differs slightly from the one treated formally in 4.2 and 4.3 where the additional rules are not needed because the systems of equations are taken in canonical form.

A member α of μTp is represented as a directed cyclic graph l, S where the nodes in S are uniquely labeled (for example, by memory addresses) and where l is the starting label. Each μ in α corresponds to a cycle in S .

More concretely, using an informal programming notation, S is a Store, where $\text{Store} \triangleq \text{Label} \rightarrow \text{Node}$ are the partial functions from labels to nodes (from memory addresses to memory locations). Then $\text{Graph} \triangleq \text{Label} \times \text{Store}$,

where $\text{Label} \triangleq \text{Nat}$, and $\text{Node} \triangleq \text{Bot} + \text{Top} + \text{Var}(\text{Tvar}) + \text{Arrow}(\text{Label} \times \text{Label}) + \text{Rec}(\text{Label})$.

An *allocator* transforms a type into a graph structure:

$$\text{Alloc}: \mu Tp \times \text{Store} \times (\text{Tvar} \rightarrow \text{Label}) \rightarrow \text{Graph}.$$

Let $\text{new}(S)$ be a label l (for example, the least one) such that $l \notin \text{dom}(S)$. We denote by $S[l = \text{Bot}]$ a store that is just like S except that $S(l) = \text{Bot}$.

$$\begin{aligned} \text{Alloc}(\perp, S, e) &\triangleq \\ &\quad \text{let } l = \text{new}(S) \text{ in } l, S[l = \text{Bot}] \\ \text{Alloc}(\top, S, e) &\triangleq \\ &\quad \text{let } l = \text{new}(S) \text{ in } l, S[l = \text{Top}] \\ \text{Alloc}(t, S, e) &\triangleq \\ &\quad \text{if } t \in \text{dom}(e) \text{ then } e(t), S \\ &\quad \text{else let } l = \text{new}(S) \text{ in } l, S[l = \text{Var}(t)] \\ \text{Alloc}(\alpha \rightarrow \beta, S, e) &\triangleq \\ &\quad \text{let } l', S' = \text{Alloc}(\alpha, S, e) \\ &\quad \text{and } l'', S'' = \text{Alloc}(\beta, S', e) \\ &\quad \text{let } l = \text{new}(S'') \text{ in } l, S''[l = \text{Arrow}(l', l'')] \\ \text{Alloc}(\mu t. \alpha, S, e) &\triangleq \\ &\quad \text{let } l = \text{new}(S) \\ &\quad \text{let } l', S' = \text{Alloc}(\alpha, S[l = \text{Bot}], e[t = l]) \\ &\quad \text{in } l, S'[l = \text{Rec}(l')] \end{aligned}$$

The allocation of $\mu t. \alpha$ is done by reserving a new memory location l , then allocating the body α by binding every occurrence of t to l , and finally storing a Rec node containing the allocation of α back into l . The store $S[l = \text{Bot}]$ is used in the recursion to prevent l from being returned again by new .

Given a path in ω^* , we can define a (partial) access function that returns the node corresponding to that path in a graph, but skipping over Rec nodes:

$$\begin{aligned} GT: \text{Graph} &\rightarrow \text{Tree}(L) \quad (\text{Where } \text{Tree}(L) = \omega^* \rightarrow L, \text{Section 3.3}) \\ GT(l, S)(\text{nil}) &\triangleq \\ &\quad \text{if } S(l) = \text{Rec}(l') \text{ then } GT(l', S)(\text{nil}) \\ &\quad \text{if } S(l) = \text{Bot} \text{ then } \perp \\ &\quad \text{if } S(l) = \text{Top} \text{ then } \top \\ &\quad \text{if } S(l) = \text{Var}(t) \text{ then } t \\ &\quad \text{if } S(l) = \text{Arrow}(l', l'') \text{ then } \rightarrow \\ GT(l, S)(0.s) &\triangleq \\ &\quad \text{if } S(l) = \text{Rec}(l') \text{ then } GT(l', S)(0.s) \\ &\quad \text{if } S(l) = \text{Arrow}(l', l'') \text{ then } GT(l', S)(s) \\ &\quad \text{else } \uparrow \\ GT(l, S)(1.s) &\triangleq \\ &\quad \text{if } S(l) = \text{Rec}(l') \text{ then } GT(l', S)(1.s) \\ &\quad \text{if } S(l) = \text{Arrow}(l', l'') \text{ then } GT(l'', S)(s) \\ &\quad \text{else } \uparrow \\ GT(l, S)(n + 2.s) &\triangleq \uparrow \end{aligned}$$

We now show that Alloc is correct and that the initial state S is irrelevant.

4.4.1 PROPOSITION. $\forall \alpha \in Tp. \forall S, l', S'.$

$$Alloc(\alpha, S, []) = l', S' \Rightarrow GT(l', S') = T\alpha.$$

PROOF (SKETCH). If $l \notin \text{dom}(S)$ then $S[l = v]$ is a *single extension* of S . S' is an *extension* of S if it is S or if it is the single extension of an extension of S .

We indicate by S^+ an arbitrary (finite) extension of S . Note that:

If $l \in \text{dom}(S)$ then $S(l) = S^+(l)$.

If $Alloc(\alpha, S, e) = l', S'$ then $l' \in \text{dom}(S')$ and S' is an extension of S .

$\forall S^+. l \in \text{dom}(S) \Rightarrow GT(l, S) = GT(l, S^+).$

To obtain the proposition, we need to prove a stronger statement:

$$\begin{aligned} &\forall \alpha \in \mu Tp. \forall n \geq 0. \forall S, m_1..m_n, \alpha_1..\alpha_n, l', S', \pi. \\ &Alloc(\alpha, S, [t_i = m_i]) = l', S' \wedge \\ &(\forall \pi' \text{ s.t. } |\pi'| \leq |\pi|. \forall S^+. GT(m_i, S^+)(\pi') = T\alpha_i(\pi') \text{ for all } i \in 1..n) \Rightarrow \\ &\forall S^+. GT(l', S^+)(\pi) = T([\alpha_i/t_i]\alpha)(\pi). \end{aligned}$$

The proof is then by induction on $|\pi|$; the hard case is $\pi = \text{i.s.}, i \in \{0, 1\}$, and $\alpha \equiv \mu t. \alpha' \rightarrow \alpha''$. \square

In the implementation of the algorithm, the assumption set Σ is represented as a *trail*, that is, a set of label pairs. This has the task of remembering the pairs of labels in the cyclic graphs that have been jointly visited.

From two types α and β we produce two graphs $l^\alpha, S^\alpha, l^\beta, S^{\alpha\beta}$ such that $S^{\alpha\beta}$ extends S^α . Then $\text{Alg}(\emptyset, S^{\alpha\beta}, l^\alpha, l^\beta)$ proceeds as follows, mimicking the rules in 4.2:

```

Alg(Tr, S, l, l')  $\triangleq$ 
  if  $\langle l, l' \rangle \in Tr$  then ok
  else if  $S(l) = \text{Bot}$  ok
  else if  $S(l') = \text{Top}$  then ok
  else if  $S(l) = \text{Var}(t)$  and  $S(l') = \text{Var}(t)$  then ok
  else if both  $S(l) = \text{Arrow}(l_1, l_2)$  and  $S(l') = \text{Arrow}(l'_1, l'_2)$  then
    Alg(Tr, S, l'_1, l'_2); Alg(Tr, S, l_2, l'_2)
  else if  $S(l) = \text{Rec}(l_1)$  and  $S(l') \neq \text{Rec}(l'_1)$  then
    Alg(Tr  $\cup \langle l, l' \rangle$ , S, l_1, l')
  else if  $S(l) \neq \text{Rec}(l_1)$  and  $S(l') = \text{Rec}(l'_1)$  then
    Alg(Tr  $\cup \langle l, l' \rangle$ , S, l, l'_1)
  else if  $S(l) = \text{Rec}(l_1)$  and  $S(l') = \text{Rec}(l'_1)$  then
    Alg(Tr  $\cup \langle l, l' \rangle$ , S, l_1, l'_1)
  else fail

```

An alternative approach is to avoid Rec nodes completely and have the allocator construct direct loops in the graph. This leads to an algorithm where trails must be kept of every pair of nodes, instead of every pair of nodes of which one is a Rec node. This algorithm is closer to the formulation of the rules in 4.2, while the present algorithm, which in practice produces much shorter trails, uses the equivalent of the (μ_{l_A}) and (μ_{r_A}) rules described there.

4.4.2 *Definition.* $\alpha \leq_C \beta \Leftrightarrow \forall S, l', S', l'', S''.$

$$\text{Alloc}(\alpha, S, [\]) = l', S' \wedge \text{Alloc}(\beta, S', [\]) = l'', S'' \Rightarrow$$

$$\text{Alg}(\emptyset, S'', l', l'') = \text{ok}.$$

From this point on it seems possible to mimic sections 4.3.4–4.3.7, modulo the use of the (μ_{l_A}) and (μ_{r_A}) rules, and show $\alpha \leq_C \beta$ iff $\alpha \leq_T \beta$, but we have not checked the cumbersome details.

5. TYPING RULES

In this section we introduce a certain number of axioms and rules for type equality and subtyping. These are intended as natural rules for a language based on subtyping and as a specification of a subtyping algorithm for such a language. In Section 4 we have studied such a subtyping algorithm; here we see that the algorithm and the rules match each other perfectly, by relating them both to trees.

5.1 Type Equivalence Rules

We say that a type α is *contractive* in the type variable t if either t does not occur free in α , or α can be rewritten via unfolding as a type of the shape $\alpha_1 \rightarrow \alpha_2$. We write this fact as $\alpha \downarrow t$.

It is now easy to observe that the contractiveness of α in t is a sufficient (and necessary) condition to enforce the contractiveness of the following functional on the space $\text{Tree}(L)$ (3.3):

$$G_{\alpha, t}(A) \triangleq [A/t]T\alpha \quad A \in \text{Tree}(L).$$

($[A/t]T\alpha$ denotes the substitution of the tree A for the occurrences of t in $T\alpha$.)

This remark suggests the following rule that is generalized to a larger calculus in [12]:

$$(\text{contract}) \quad [\beta/t]\alpha = \beta, \quad [\beta'/t]\alpha = \beta', \quad \alpha \downarrow t \Rightarrow \beta = \beta'.$$

In words, if two types β and β' are fixpoints of the same functional $\alpha[t]$, then they are equal since contractive functionals have unique fixpoints. This rule was also inspired by a standard proof technique for bisimulation [23]. Moreover, it is convenient to identify $\mu t.t = \perp$.

In this section we consider the equivalence

$$\vdash \alpha = \beta \quad (\text{or } \alpha =_R \beta).$$

meaning that $\alpha = \beta$ can be derived in the congruence induced by the (contract) rule and the (fold-unfold) and $(\mu\text{-}\perp)$ axioms below. Here is the complete

axiomatization:

(refl)	$\alpha = \alpha$
(symm)	$\alpha = \beta \Rightarrow \beta = \alpha$
(trans)	$\alpha = \beta, \beta = \gamma \Rightarrow \alpha = \gamma$
(\rightarrow -congr)	$\alpha = \alpha', \beta = \beta' \Rightarrow \alpha \rightarrow \beta = \alpha' \rightarrow \beta'$
(μ -congr)	$\alpha = \beta \Rightarrow \mu t. \alpha = \mu t. \beta$
($\mu \cdot \perp$)	$\mu t. t = \perp$
(fold-unfold)	$[\mu t. \alpha / t] \alpha = \mu t. \alpha$
(contract)	$[\beta / t] \alpha = \beta, [\beta' / t] \alpha = \beta', \alpha \downarrow t \Rightarrow \beta = \beta'.$

5.1.1 PROPOSITION (Soundness of the equivalence rules w.r.t. the trees).

$$\alpha =_R \beta \Rightarrow T\alpha = T\beta.$$

PROOF. Immediate by the previous considerations. \square

5.1.2 *Derived Rules.* By means of (contract) and (fold-unfold) it is possible to prove new interesting equivalences, for example:

- (1) $\mu t. s \rightarrow t = \mu t. s \rightarrow (s \rightarrow t).$
- (2) $\mu t. \mu s. \alpha = \mu v. [v/t, v/s] \alpha$ (μ -contraction).

We make explicit a free variable by writing, for example, $\alpha[t]$. Then we have:

- (1) Consider $\gamma[r] = s \rightarrow (s \rightarrow r).$
 $\mu t. s \rightarrow t = s \rightarrow (\mu t. s \rightarrow t) = s \rightarrow (s \rightarrow (\mu t. s \rightarrow t)) = \gamma[\mu t. s \rightarrow t].$
 $\mu t. s \rightarrow (s \rightarrow t) = s \rightarrow (s \rightarrow (\mu t. s \rightarrow (s \rightarrow t))) = \gamma[\mu t. s \rightarrow (s \rightarrow t)].$
- (2) Let: $\alpha \equiv \mu t. \mu s. \gamma[t, s], \alpha' \equiv \mu s. \gamma[\alpha, s] = \alpha, \beta = \mu v. \gamma[v, v].$
 Consider $\gamma[w, w]$. Then: $\gamma[\alpha, \alpha] = \gamma[\alpha, \alpha'] = \alpha' = \alpha$ and $\gamma[\beta, \beta] = \beta.$

5.1.3 *Reduction to Canonical Form.* It easy to show that any recursive type is provably equivalent ($=_R$) to a type in canonical form. The strategy can be described as follows:

- (a) Use unfold to get rid of all μ 's that do not bind any variable.
- (b) Use μ -contraction to reduce sequences of μ 's to one μ .
- (c) Use $\mu \cdot \perp$ to reduce to \perp all subtypes of the shape $\mu t. t.$

5.2 Completeness of Equivalence Rules

By the strong connection between regular trees and recursive types we show that any time two recursive types α, β have the same tree expansion $T\alpha = T\beta$, then we can conclude $\vdash \alpha = \beta$.

First we show how to solve systems of type equations. Then we introduce the notion of *equational characterization* of a type: that is, how to characterize a type by a system of type equations. Finally we use equational characterizations to prove the completeness theorem.

In this section we use the following notation. If γ has free variables $\{u_1 \dots u_p\} \subseteq \{t_1 \dots t_n\}$, then we write $\gamma[\alpha_1 \dots \alpha_n]$ for the substitution $[\alpha_1/t_1 \dots \alpha_n/t_n]\gamma$. In particular, $\gamma[t_1 \dots t_n]$ emphasizes a superset of the free variables of γ .

5.2.1 LEMMA (A system of equations has a solution, by iterated elimination). *Every system of n equations in n variables:*

$$t_i = \gamma_i[t_1 \dots t_n] \quad (i \in 1..n)$$

has a solution in the congruence induced by the axiom (fold-unfold). That is, there are $\alpha_1 \dots \alpha_n$ such that $\vdash \alpha_i = \gamma_i[\alpha_1 \dots \alpha_n]$ ($i \in 1..n$).

PROOF. By induction on n .

Case $n = 1$. Given the equation $t = \gamma[t]$ just take $\mu t. \gamma[t]$.

Case $n \geq 2$. Given the equations $t_i = \gamma_i[t_1 \dots t_n]$ ($i \in 1..n$) take $\alpha'_n[t_1 \dots t_{n-1}] \equiv \mu t_n. \gamma_n[t_1 \dots t_n]$. Consider the system of $n - 1$ equations: $t_i = \gamma_i[t_1 \dots t_{n-1} \alpha'_n[t_1 \dots t_{n-1}]]$ ($i \in 1..n - 1$) which by inductive hypothesis has solution $\alpha_1 \dots \alpha_{n-1}$, that is:

$$\alpha_i = \gamma_i[\alpha_1 \dots \alpha_{n-1} \mu t_n. \gamma_n[\alpha_1 \dots \alpha_{n-1} t_n]] \quad (i \in 1..n - 1).$$

Now take $\alpha_n \equiv \mu t_n. \gamma_n[\alpha_1 \dots \alpha_{n-1} t_n]$ and check that $\alpha_1 \dots \alpha_n$ is a solution for the original system. \square

5.2.2 LEMMA (A system of contractive equations has a unique solution). *Assume that, for $i \in 1..n$, we have two sets of types α_i, β_i , related by two systems of equations:*

$$\vdash \alpha_i = \gamma_i[\alpha_1 \dots \alpha_n] \quad \vdash \beta_i = \gamma_i[\beta_1, \dots, \beta_n]$$

such that $\gamma_i[t_1 \dots t_n] \downarrow t_j$ for $i, j \in 1..n$. Then, for all i : $\vdash \alpha_i = \beta_i$.

PROOF. By induction on n .

Case $n = 1$. We have $\vdash \alpha = \gamma[\alpha]$ and $\vdash \beta = \gamma[\beta]$. Consider the context $\gamma[t]$, by (contract) we have $\vdash \alpha = \beta$.

Case $n \geq 2$. Consider the n th equation. We have, by (fold-unfold),

$$\begin{aligned} \vdash \mu t_n. \gamma_n[\alpha_1 \dots \alpha_{n-1} t_n] &= \gamma_n[\alpha_1 \dots \alpha_{n-1} \mu t_n. \gamma_n[\alpha_1 \dots \alpha_{n-1} t_n]] \\ \vdash \mu t_n. \gamma_n[\beta_1 \dots \beta_{n-1} t_n] &= \gamma_n[\beta_1 \dots \beta_{n-1} \mu t_n. \gamma_n[\beta_1 \dots \beta_{n-1} t_n]]. \end{aligned}$$

Hence, by (contract) on t_n , $\vdash \alpha_n = \mu t_n. \gamma_n[\alpha_1 \dots \alpha_{n-1} t_n]$, $\vdash \beta_n = \mu t_n. \gamma_n[\beta_1 \dots \beta_{n-1} t_n]$.

Take $\gamma'[t_1 \dots t_{n-1}] \equiv \mu t_n. \gamma_n[t_1 \dots t_n]$. We can now construct a system of size $n - 1$:

$$t_i = \gamma_i[t_1 \dots t_{n-1} \gamma'[t_1 \dots t_{n-1}]] \quad (i \in 1..n - 1)$$

and check that both $\alpha_1 \dots \alpha_{n-1}$ and $\beta_1 \dots \beta_{n-1}$ are solutions. Hence, by inductive hypothesis $\vdash \alpha_i = \beta_i$ for $i \in 1..n - 1$. Moreover, by congruence we obtain $\vdash \alpha_n = \beta_n$. \square

5.2.3 *Definition.* A node context $p[t_1 \dots t_n]$ for $p \in L \setminus \{t_1 \dots t_n\}$ (see Section 3.3 and Proof 4.1.5) is a type of the form $p(u_1..u_{\#p})$, where $\#p$ is the arity of p , and $\{u_1..u_{\#p}\} \subseteq \{t_1 \dots t_n\}$.

Node contexts provide a convenient metanotation for nodes whose children are all type variables. For example, the type $\rightarrow(r, r)$ can be denoted by the

node context $\rightarrow [r]$ or (redundantly) by $\rightarrow [r, s, t]$, and the type $\rightarrow (r, s)$ by $\rightarrow [s, r]$ among others.

Note that a node context $p[t_1 \dots t_n]$ is contractive in each t_i , because either t_i is prefixed by p or does not occur in the type.

5.2.4 Definition. A type $\alpha \in \text{Type}$ is *equationally characterized* (*eq. char.*) if there are types $\alpha_1 \dots \alpha_n$ with $\alpha \equiv \alpha_1$, and there are node contexts $p_i[t_1 \dots t_n]$, $i \in 1..n$, for which $\vdash \alpha_i = p_i[\alpha_1 \dots \alpha_n]$.

An equation, $t_j = p_j[t_1 \dots t_n]$, in a system is reachable from a variable t_k if $k = j$ or if it is reachable from the variables in $p_k[t_1 \dots t_n]$ (see Section 4.1.3). An equation is reachable from another if it is reachable from any of the variables in the other.

5.2.5 LEMMA (Building an equational characterization). *Every term $\alpha \in \text{Type}$ has an equational characterization such that all equations are reachable from the first one.*

PROOF. The construction is basically the same as the one in 4.1.5. It is enough to prove by induction on the structure of γ that every term in $\mu T p$ is equationally characterized. Then the lemma follows by 5.1.3 and by the invariance of equational characterization modulo provable equivalence. \square

5.2.6 LEMMA. *Assume $T\alpha = T\beta$ and $\vdash \alpha = p(\alpha_1 \dots \alpha_{\#p})$, $\vdash \beta = q(\beta_1 \dots \beta_{\#q})$, where $p, q \in L$. Then $p = q$ and $T\alpha_i = T\beta_i$ for all $i \in 1..\#p$.*

PROOF. By soundness, $T\alpha = Tp(\alpha_1 \dots \alpha_{\#p})$ and $T\beta = Tq(\beta_1 \dots \beta_{\#q})$. Hence, $p = q$ and $T\alpha_i = T\beta_i$ by definition of T . \square

5.2.7 THEOREM (Completeness of type equivalence rules). *If $T\alpha = T\beta$ then $\alpha =_R \beta$.*

PROOF. The idea of the proof is as follows: given α and β such that $T\alpha = T\beta$ we produce their corresponding equational characterizations, say $ec(\alpha)$ and $ec(\beta)$. By a collapse of “equivalent” equations we derive a new equational characterization $ec(\gamma)$. The solutions of the (smaller) system associated with $ec(\gamma)$ can be replicated to produce solutions for the systems associated with $ec(\alpha)$ and $ec(\beta)$. Hence, we can apply Lemma 5.2.2 twice (uniqueness of solutions) and then transitivity to conclude $\alpha =_R \beta$.

Let $T\alpha = T\beta$; by Lemma 5.2.5, α, β are equationally characterized by α_i , $t_i = p_i[t_1 \dots t_n]$ and β_j , $t_j = q_j[t_1 \dots t_m]$ so that all equations are reachable from the first ones.

From these α_i, β_j we generate a sequence of pairs (A^h, B^h) where A^h, B^h are equivalence classes of α_i and β_j respectively. Moreover, for each h , $\alpha_{i_1}, \alpha_{i_2} \in A^h$, and $\beta_{j_1}, \beta_{j_2} \in B^h$, we shall have the invariant $T\alpha_{i_1} = T\alpha_{i_2} = T\beta_{j_1} = T\beta_{j_2}$. We start with the pair $(A^1, B^1) \equiv (\{\alpha\}, \{\beta\})$. At each step we consider all the pairs α_i, β_j such that $\alpha_i \in A^h$ and $\beta_j \in B^h$ for some h . We indicate by $\alpha_{(i', i'')}$ some α_i depending on both i' and i'' , similarly for $\beta_{(j', j'')}$. If $\alpha_i = p_i(\alpha_{(i, 1)} \dots \alpha_{(i, \#p_i)})$ and $\beta_j = q_j(\beta_{(j, 1)} \dots \beta_{(j, \#q_j)})$, we have, by Lemma 5.2.6, $p_i = q_j$ and $T\alpha_{(i, 1)} = T\beta_{(j, 1)} \dots T\alpha_{(i, \#p_i)} = T\beta_{(j, \#q_j)}$. We add all the pairs $(\alpha', \beta') \in \{(\alpha_{(i, 1)}, \beta_{(j, 1)}), \dots, (\alpha_{(i, \#p_i)}, \beta_{(j, \#q_j)})\}$ in the following way, respecting

the invariant above:

- if $\alpha' \in A^h$ and $\beta' \in B^h$ for some h , then nothing is done;
- else, if $\alpha' \in A^{h_1}$, and $\beta' \in B^{h_2}$, with $h_1 \neq h_2$, then we replace the pairs (A^{h_1}, B^{h_1}) and (A^{h_2}, B^{h_2}) by $(A^{h_1} \cup A^{h_2}, B^{h_1} \cup B^{h_2})$;
- else, if $\alpha' \in A^h$ we replace the pair (A^h, B^h) by $(A^h, B^h \cup \{\beta'\})$;
- else, if $\beta' \in B^h$ we replace the pair (A^h, B^h) by $(A^h \cup \{\alpha'\}, B^h)$;
- else we add a new pair $(\{\alpha'\}, \{\beta'\})$.

We stop when the list of pairs no longer changes. This process terminates because there are at most $n \cdot m$ pairs to consider.

The process above produces two partitions of α_i and β_j of size $k \leq n$, $k \leq m$, for some k . These are total partitions since all equations are reachable from the first ones. These partitions determine two functions $\sigma: 1..n \rightarrow 1..k$ and $\pi: 1..m \rightarrow 1..k$ such that:

- $\sigma(i) = \pi(j) \Leftrightarrow \alpha_i \in A^h, \beta_j \in B^h$ for some h .
- $\sigma(i_1) = \sigma(i_2) \Leftrightarrow \alpha_{i_1}, \alpha_{i_2} \in A^h$ for some h .
- $\pi(i_1) = \pi(i_2) \Leftrightarrow \beta_{i_1}, \beta_{i_2} \in B^h$ for some h .

Given these partitions, we now define a system of k equations $t_h = r_h[t_1..t_k]$, which will turn out to be equivalent both to the p_i and q_j systems. For $h \in 1..k$ we have:

$$\begin{aligned} t_h &= r_h[t_1..t_k] \quad \text{where} \\ r_{\sigma(i)}[t_1..t_k] &\equiv p_i[t_{\sigma(1)} \dots t_{\sigma(n)}] \\ r_{\pi(j)}[t_1..t_k] &\equiv q_j[t_{\pi(1)} \dots t_{\pi(m)}]. \end{aligned}$$

We need to argue that this is a proper definition, since we can have, for example, $\sigma(i) = \pi(j)$ for some i, j . We show that when this happens, we also have by construction that $r_{\sigma(i)}[t_1..t_k] \equiv r_{\pi(j)}[t_1..t_k]$. Similarly for the other possible conflicts: $\sigma(i_1) = \sigma(i_2)$ for some i_1, i_2 , and $\pi(i_1) = \pi(i_2)$ for some j_1, j_2 . To show these facts, we further investigate the properties of σ and π .

- $\sigma(1) = \pi(1)$ since α, β start in the same pair (A^1, B^1) .
- If $\sigma(i) = \pi(j)$ then $p_i \equiv q_j$. Moreover, let $\alpha_i = p_i[\alpha_1 \dots \alpha_n] \equiv p_i(\alpha_{(i,1)} \dots \alpha_{(i,\#p_i)})$ and $\beta_j = q_j[\beta_1 \dots \beta_m] \equiv q_j(\beta_{(j,1)} \dots \beta_{(j,\#q_j)})$ be the i th and j th equations in the respective systems. Then $\alpha_i \in A^h, \beta_j \in B^h$ for some h (property above); the pair α_i, β_j was considered in the process above; that is, the pairs $\alpha_{(i,1)}, \beta_{(j,1)} \dots \alpha_{(i,\#p_i)}, \beta_{(j,\#q_j)}$ were also added to the list. Therefore $\sigma(i, 1) = \pi(j, 1) \dots \sigma(i, \#p_i) = \pi(j, \#q_j)$, and $p_i(t_{\sigma(i,1)} \dots t_{\sigma(i,\#p_i)}) \equiv q_j(t_{\pi(j,1)} \dots t_{\pi(j,\#q_j)})$. This is the same as saying $p_i[t_{\sigma(1)} \dots t_{\sigma(n)}] \equiv q_j[t_{\pi(1)} \dots t_{\pi(m)}]$.
- If $\sigma(i_1) = \sigma(i_2)$ then $p_{i_1} \equiv p_{i_2}$. Moreover, let $\alpha_{i_1} = p_{i_1}(\alpha_{(i_1,1)} \dots \alpha_{(i_1,\#p_{i_1})})$ and $\alpha_{i_2} = p_{i_2}(\alpha_{(i_2,1)} \dots \alpha_{(i_2,\#p_{i_2})})$ be the i_1 th and i_2 th equations in the α system. Then $\alpha_{i_1}, \alpha_{i_2} \in A^h$ for some h (property above). Consider any $\beta_j \in B^h$; the pairs $\alpha_{i_1}, \beta_j, \alpha_{i_2}, \beta_j$ were considered in the process above; that is, the pairs $\alpha_{(i_1,1)}, \beta_{(j,1)}, \alpha_{(i_2,1)}, \beta_{(j,1)}$ were also added to the list. Therefore $\sigma(i_1, 1) = \pi(j, 1) = \sigma(i_2, 1)$, and similarly up to $\sigma(i_1, \#p_{i_1}) =$

$\sigma(i2, \#p_{i2})$. Hence, $p_{i1}(t_{\sigma(i1,1)} \dots t_{\sigma(i1, \#p_{i1})}) \equiv p_{i2}(t_{\sigma(i2,1)} \dots t_{\sigma(i2, \#p_{i2})})$. This is the same as saying $p_{i1}[t_{\sigma(1)} \dots t_{\sigma(n)}] \equiv p_{i2}[t_{\sigma(1)} \dots t_{\sigma(n)}]$.
 —Similarly for $\pi(i1) = \pi(i2)$.

Hence we conclude:

—if $\sigma(i) = \pi(j)$
 then $r_{\sigma(i)}[t_1..t_k] \equiv p_i[t_{\sigma(1)} \dots t_{\sigma(n)}] \equiv q_j[t_{\pi(1)} \dots t_{\pi(m)}] \equiv r_{\pi(j)}[t_1..t_k]$
 —if $\sigma(i1) = \sigma(i2)$
 then $r_{\sigma(i1)}[t_1..t_k] \equiv p_{i1}[t_{\sigma(1)} \dots t_{\sigma(n)}] \equiv p_{i2}[t_{\sigma(1)} \dots t_{\sigma(n)}] \equiv r_{\sigma(i2)}[t_1..t_k]$
 —similarly for $\pi(j1) = \pi(j2)$.

Now by Lemma 5.2.1 we can construct a solution of the system $t_h = r_h[t_1..t_k]$; that is, we can obtain $\gamma_1.. \gamma_k$ such that $\vdash \gamma_h = r_h[\gamma_1.. \gamma_k]$.

Then $\vdash \gamma_{\sigma(i)} = r_{\sigma(i)}[\gamma_1.. \gamma_k] \equiv p_i[\gamma_{\sigma(1)} \dots \gamma_{\sigma(n)}]$ for all i . Therefore, the γ 's (when appropriately replicated) satisfy the same system as the α 's, and by Lemma 5.2.2 we have $\vdash \alpha_i = \gamma_{\sigma(i)}$. Similarly, the γ 's satisfy the β 's system, and $\vdash \beta_j = \gamma_{\pi(j)}$. Moreover, $\sigma(1) = \pi(1)$; hence, we have $\vdash \alpha \equiv \alpha_1 = \gamma_{\sigma(1)} = \gamma_{\pi(1)} = \beta_1 \equiv \beta$ by transitivity. \square

This constructive proof is based on the one in [25] (see also [21]), but differs in an important point as, in addition, we must deal with equivalence classes of types.

5.2.8 Example. In this example, arising from a discussion with Mario Coppo, we consider the types:

$$\alpha \triangleq \mu t. t \rightarrow (t \rightarrow t) \quad \beta \triangleq \mu t. (t \rightarrow t) \rightarrow t.$$

We have $T\alpha = T\beta$, but note that there is no single context that can prove them equivalent by the (contract) rule. We must find a third type γ which is independently provably equal to α and β (by contract), and then we can obtain $\vdash \alpha = \beta$ by transitivity. To find this γ , we instantiate the proof of 5.2.7.

We start with two equational characterizations for α and β :

$$\begin{array}{ll} \alpha_1 \triangleq \alpha & \beta_1 \triangleq \beta \\ \alpha_2 \triangleq \alpha \rightarrow \alpha & \beta_2 \triangleq \beta \rightarrow \beta \\ p_1[t_1, t_2] \triangleq t_1 \rightarrow t_2 & q_1[t_1, t_2] \triangleq t_2 \rightarrow t_1 \\ p_2[t_1, t_2] \triangleq t_1 \rightarrow t_1 & q_2[t_1, t_2] \triangleq t_1 \rightarrow t_1. \end{array}$$

That is, the following are provable, by (fold-unfold):

$$\begin{array}{ll} \alpha_1 = \alpha_1 \rightarrow \alpha_2 & \beta_1 = \beta_2 \rightarrow \beta_1 \\ \alpha_2 = \alpha_1 \rightarrow \alpha_1 & \beta_2 = \beta_1 \rightarrow \beta_1. \end{array}$$

Starting with the list $(\{\alpha_1\}, \{\beta_1\})$, we must match the equations for α_1 and β_1 . This involves equating the pairs α_1, β_2 (obtaining $(\{\alpha_1\}, \{\beta_1, \beta_2\})$, and α_2, β_1 (obtaining $(\{\alpha_1, \alpha_2\}, \{\beta_1, \beta_2\})$). Matching the newly inserted pairs does not further modify the situation; hence, we have reached termination with the

partitions:

$$(\{\alpha_1, \alpha_2\}, \{\beta_1, \beta_2\}) \quad \text{with } k = 1 \text{ and } \sigma = \pi = \{1 \mapsto 1, 2 \mapsto 1\}.$$

The associated system of one equation is $t_1 = r_1[t_1]$, where:

$$r_1[t_1] = \left\{ \begin{array}{l} r_{\sigma(1)}[t_1] \equiv p_1[t_{\sigma(1)}, t_{\sigma(2)}] \\ r_{\sigma(2)}[t_1] \equiv p_2[t_{\sigma(1)}, t_{\sigma(2)}] \\ r_{\pi(1)}[t_1] \equiv q_1[t_{\pi(1)}, t_{\pi(2)}] \\ r_{\pi(2)}[t_1] \equiv q_2[t_{\pi(1)}, t_{\pi(2)}] \end{array} \right\} = t_1 \rightarrow t_1.$$

We now generate a solution for this system:

$$\gamma_1 \triangleq \mu t. t \rightarrow t \quad \text{such that } \vdash \gamma_1 = \gamma_1 \rightarrow \gamma_1 = r_1[\gamma_1].$$

We can verify that (γ_1, γ_1) solves the α and β systems:

$$\begin{array}{ll} \vdash \gamma_1 = p_1[\gamma_1, \gamma_1] & \vdash \gamma_1 = q_1[\gamma_1, \gamma_1] \\ \vdash \gamma_1 = p_2[\gamma_1, \gamma_1] & \vdash \gamma_1 = q_2[\gamma_1, \gamma_1]. \end{array}$$

Hence, a proof of $\vdash \alpha_1 = \gamma_1$ can be constructed by Lemma 5.2.2 (more simply, by unfolding $\vdash \alpha_1 = \alpha_1 \rightarrow (\alpha_1 \rightarrow \alpha_1)$ and $\vdash \gamma_1 = \gamma_1 \rightarrow \gamma_1 = \gamma_1 \rightarrow (\gamma_1 \rightarrow \gamma_1)$, hence $\vdash \alpha_1 = \gamma_1$ by (contract)). Similarly, $\vdash \gamma_1 = \beta_1$. Hence, by transitivity, $\vdash \alpha_1 = \beta_1$.

5.3 Subtyping Rules

At first it is not clear how to define a rule for the subtyping of recursive types that is sufficiently powerful. In particular, observe that the computational rule (μ_A) in Section 4.2 does not have any apparent logical meaning since the premise is always valid under a classical reading of the entailment relation.

We now introduce a rule (μ_R) , whose soundness is clear. Later, in Section 5.4, we will show that, in conjunction with the type equivalence rules, (μ_R) leads to a subtyping system complete with respect to the tree ordering.

We denote with Γ a set $\{t_1 \leq s_1, \dots, t_n \leq s_n\}$ of subtyping assumptions on type variables. We write a subtype judgment as: $\Gamma \supset \alpha \leq \beta$.

Define a formal system for deriving this kind of judgments as follows; this is based on the $\alpha = \beta$ congruence in 5.1:

$$\begin{array}{ll} (\text{eq}_R) & \alpha = \beta \Rightarrow \Gamma \supset \alpha \leq \beta \\ (\text{trans}_R) & \Gamma \supset \alpha \leq \beta, \Gamma \supset \beta \leq \gamma \Rightarrow \Gamma \supset \alpha \leq \gamma \\ (\text{assmp}_R) & t \leq s \in \Gamma \Rightarrow \Gamma \supset t \leq s \\ (\perp_R) & \Gamma \supset \perp \leq \alpha \\ (\top_R) & \Gamma \supset \alpha \leq \top \\ (\rightarrow_R) & \Gamma \supset \alpha' \leq \alpha, \Gamma \supset \beta \leq \beta' \Rightarrow \Gamma \supset \alpha \rightarrow \beta \leq \alpha' \rightarrow \beta' \\ (\mu_R) & \Gamma \cup \{t \leq s\} \supset \alpha \leq \beta \Rightarrow \Gamma \supset \mu t. \alpha \leq \mu s. \beta \text{ with } t \text{ only} \\ & \text{in } \alpha; s \text{ only in } \beta; t, s \text{ not in } \Gamma \end{array}$$

We say $\alpha \leq_R \beta$ if we can derive $\emptyset \supset \alpha \leq \beta$. The last rule was proposed in [10] in the specification of the Amber programming language as a first attempt to define a theory for the subtyping of recursive types.

5.3.1 PROPOSITION (Soundness of the rule ordering w.r.t. the tree ordering).
If $\alpha \leq_R \beta$ then $\alpha \leq_T \beta$.

PROOF. We prove the more general statement:

If $\vdash_R \{t_1 \leq s_1, \dots, t_n \leq s_n\} \supset \alpha \leq \beta$
and $\alpha_1 \leq_T \beta_1, \dots, \alpha_n \leq_T \beta_n$ so that $\{t_1, s_1, \dots, t_n, s_n\} \cap FV$
 $(\alpha_1, \beta_1, \dots, \alpha_n, \beta_n) = \emptyset$
then $[\alpha_1/t_1, \beta_1/s_1, \dots, \alpha_n/t_n, \beta_n/s_n]\alpha \leq_T$
 $[\alpha_1/t_1, \beta_1/s_1, \dots, \alpha_n/t_n, \beta_n/s_n]\beta$.

The proof goes by induction on the length of the derivation \vdash_R . The only interesting cases arise for (μ_R) and (eq_R) .

For brevity we write lists such as $t_1 \leq s_1, \dots, t_n \leq s_n$ in the form $t_i \leq s_i$ for a free i .

Case (μ_R) . $\{t_i \leq s_i, t \leq s\} \supset \alpha \leq \beta \Rightarrow \{t_i \leq s_i\} \supset \mu t. \alpha \leq \mu s. \beta$ with $t \notin FV(\beta)$; $s \notin FV(\alpha)$; $t, s \neq t_i, s_i$ for any i .

By induction hypothesis:

$$\forall \alpha_i \leq_T \beta_i, \underline{\alpha} \leq_T \underline{\beta} \text{ such that } \{t_i, s_i, t, s\} \cap FV(\alpha_i, \beta_i, \underline{\alpha}, \underline{\beta}) = \emptyset.$$

$$[\alpha_i/t_i, \beta_i/s_i, \underline{\alpha}/t]\alpha \leq_T [\alpha_i/t_i, \beta_i/s_i, \underline{\beta}/s]\beta.$$

Define

$$\alpha^0 \triangleq \perp \quad \alpha^{n+1} \triangleq [\alpha_i/t_i, \beta_i/s_i, \alpha^n/t]\alpha$$

$$\beta^0 \triangleq \perp \quad \beta^{n+1} \triangleq [\alpha_i/t_i, \beta_i/s_i, \beta^n/s]\beta.$$

Applying the induction hypothesis with $\underline{\alpha} = \alpha^n$, $\underline{\beta} = \beta^n$ we obtain $\alpha^{n+1} \leq_T \beta^{n+1}$ for every n .

For every k we can then choose an n sufficiently large so that:

$$(\mu t. [\alpha_i/t_i, \beta_i/s_i]\alpha)_{|k} =_T \alpha^n|_k \leq_T \beta^n|_k =_T (\mu s. [\alpha_i/t_i, \beta_i/s_i]\beta)_{|k}$$

(such n is found by examining how t and s occur in α and β).

Hence, by definition of \leq_T for recursive types, we have shown:

$$[\alpha_i/t_i, \beta_i/s_i](\mu t. \alpha) \leq_T [\alpha_i/t_i, \beta_i/s_i](\mu s. \beta).$$

Case (eq_R) . $\alpha =_R \beta \Rightarrow \{t_i \leq s_i\} \supset \alpha \leq \beta$.

Since $=_R$ is a congruence, we have $[\alpha_i/t_i, \beta_i/s_i]\alpha =_R [\alpha_i/t_i, \beta_i/s_i]\beta$. By soundness of $=_R$ we have $[\alpha_i/t_i, \beta_i/s_i]\alpha =_T [\alpha_i/t_i, \beta_i/s_i]\beta$. Finally, $[\alpha_i/t_i, \beta_i/s_i]\alpha \leq_T [\alpha_i/t_i, \beta_i/s_i]\beta$ since \leq_T is a preorder. \square

5.3.2 It is easy to observe that if we prove something in the system without using (eq_R) and $(trans_R)$ then all the assumptions, $t \leq s$, inserted in Γ when applying the rule (μ_R) can be used only with respect to a pair of *positive* occurrences of t in α and s in β .

5.3.3 Then one may wonder whether the following rule suffices for our purposes [5]:

$$\alpha \leq \beta, \text{ Monotonic}(t, \alpha), \text{ Monotonic}(t, \beta) \Rightarrow \mu t. \alpha \leq \mu t. \beta$$

where $\text{Monotonic}(t, \alpha)$ iff t does not occur negatively in α .

Unfortunately it does not, as we cannot prove inclusions involving negative occurrences, as in $\mu t.t \rightarrow t \leq t.\perp \rightarrow t$.

Moreover, one must be careful in defining “ t does not occur negatively in α ” for recursive types, in order to ensure that α is really monotonic in t (for example, $\mu s.s \rightarrow t$ is not monotonic in t):

$$\begin{aligned} \text{PosAlso}(t, t) &\triangleq \text{True}, \text{PosAlso}(t, s) \triangleq \text{False} \ (s \neq t) \\ \text{PosAlso}(t, \perp) &\triangleq \text{False}, \text{PosAlso}(t, \top) \triangleq \text{False} \\ \text{PosAlso}(t, \alpha \rightarrow \beta) &\triangleq \text{NegAlso}(t, \alpha) \vee \text{PosAlso}(t, \beta) \\ \text{PosAlso}(t, \mu s.\alpha) &\triangleq (\text{NegAlso}(s, \alpha) \wedge t \in \text{FV}(\alpha)) \vee \text{PosAlso}(t, \alpha) \ (s \neq t) \end{aligned}$$

$$\begin{aligned} \text{NegAlso}(t, s) &\triangleq \text{False} \ (\text{even when } s = t) \\ \text{NegAlso}(t, \perp) &\triangleq \text{False}, \text{NegAlso}(t, \top) \triangleq \text{False} \\ \text{NegAlso}(t, \alpha \rightarrow \beta) &\triangleq \text{PosAlso}(t, \alpha) \vee \text{NegAlso}(t, \beta) \\ \text{NegAlso}(t, \mu s.\alpha) &\triangleq (\text{NegAlso}(s, \alpha) \wedge t \in \text{FV}(\alpha)) \vee \text{NegAlso}(t, \alpha) \ (s \neq t) \end{aligned}$$

$$\text{Monotonic}(t, \alpha) \triangleq \neg \text{NegAlso}(t, \alpha).$$

Under these conditions, it is possible to show that the rule above is provable from the system in 5.3.

5.4 Completeness of Subtyping Rules

In proving the completeness of the subtyping rules w.r.t. the tree ordering, it seems helpful to go through the algorithm. The rather obvious approach of extracting a proof from a successful execution tree is complicated by the lack of correspondence between the computational rule (μ_A) and the rule (μ_R) , as the former can be applied repeatedly on the same variable, whereas the latter can be applied at most once.

One may wonder if it is possible to rearrange the regular systems, while preserving type equivalence, so that during the execution we never have to expand twice the same variable by means of (μ_A) .

Naively this corresponds to a *controlled* unfolding of the recursive types so that the corresponding μ 's appear at the same time in the visit of the trees. For example, to prove $\mu t.t \rightarrow t \leq_R \mu s.(\mu s'.\top \rightarrow s') \rightarrow s$, we unfold the first type to $\mu t.(\mu t'.t' \rightarrow t') \rightarrow t$; note that this is not the unfolding given by the (fold-unfold) rule.

If $\vdash_A \Sigma, \varepsilon \supset t \leq s$ (see 4.2.3), we say that (the successful execution tree of) the initial goal $\Sigma, \varepsilon \supset t \leq s$ has the *one-expansion property* iff the following is true: for every $t \in \text{Dom}(\varepsilon)$ and for each path p of the execution tree, t is expanded in a (μ_A) node of p at most once.

It follows that with one-expansion, each variable can be inserted in Σ in a unique way, so that for each pair of assumptions $t_1 \leq s_1, t_2 \leq s_2 \in \Sigma$ we have that t_1, s_1, t_2, s_2 are pairwise distinct. Moreover, if we consider two (μ_A) nodes $\Sigma, \varepsilon \supset t_1 \leq s_1, \Sigma, \varepsilon \supset t_2 \leq s_2$ on the same path then t_1, s_1, t_2, s_2 are pairwise distinct, and if we consider a (μ_A) node $\Sigma, \varepsilon \supset t_1 \leq s_1$ and an (assmp_A) node $\Sigma, \varepsilon \supset t_2 \leq s_2$ on the same path then either $t_1 \equiv t_2, s_1 \equiv s_2$, or t_1, s_1, t_2, s_2 are pairwise distinct.

5.4.1 LEMMA (Putting recursions in lockstep). *If $\vdash_A \Sigma, \varepsilon \supset t \leq s$ then there are θ, r, u such that $\vdash_A \Sigma, \theta \supset r \leq u$, $\text{Sol}(r, \theta) = \text{Sol}(t, \varepsilon)$, $\text{Sol}(u, \theta) = \text{Sol}(s, \varepsilon)$, and $\Sigma, \theta \supset r \leq u$ satisfy the one-expansion property.*

PROOF. Given the initial goal $\Sigma, \varepsilon \supset t \leq s$ and the related successful execution tree, we build a new judgment $\Sigma, \theta \supset r \leq u$ such that the following properties hold:

- (a) $\Sigma, \theta \supset r \leq u$ is an initial goal.
- (b) $\text{Sol}(r, \theta) = \text{Sol}(t, \varepsilon)$ and $\text{Sol}(u, \theta) = \text{Sol}(s, \varepsilon)$.
- (c) $\vdash_A \Sigma, \theta \supset r \leq u$, and the execution tree is equal to the one for $\Sigma, \varepsilon \supset t \leq s$ modulo variable renaming.
- (d) $\Sigma, \theta \supset r \leq u$ satisfies the one-expansion property.

First we build the execution tree of $\Sigma, \varepsilon \supset t \leq s$. Then we associate with every node of the tree a couple (r, u) (or (u, r) on negative branches) of fresh variables with the following constraint; with every assumption leaf for $t \leq s$ we associate the same pair of variables as with the μ node where the assumption $t \leq s$ has been introduced into Σ (if any).

Next generate θ according to the following cases.

Case ($\mu\text{-}\perp$). Say we are in the situation: $\Sigma', \varepsilon \supset \perp \leq \beta \Rightarrow \Sigma', \varepsilon \supset t \leq s_0$ where $\varepsilon(t) = \perp$. If (r, u_0) is the pair of variables associated with the μ -node add the equations:

$$r = \perp \\ [u_0/s_0, u_1/s_1, \dots, u_n/s_n](s_i = \varepsilon(s_i)) \quad \text{for } i \in 0..n$$

where $u_1 \dots u_n$ are fresh variables and $s_1 \dots s_n$ are the variables reachable from s_0 in the system ε , that is $\{s_1 \dots s_n\} = \text{Reach}(s_0, \varepsilon) \cap \text{Dom}(\varepsilon)$.

Case ($\mu\text{-}\top$). Analogous.

Case ($\mu\text{-var}$). Say we are in the situation $\Sigma', \varepsilon \supset a \leq a \Rightarrow \Sigma', \varepsilon \supset t \leq s$. If (r, u) is the pair of variables associated with the μ -node, we add a pair of equations: $r = a, u = a$.

Case ($\mu\text{-}\rightarrow$). Say we are in the situation

$$\Sigma' \cup \{t \leq s\}, \varepsilon \supset s_1 \leq t_1, \Sigma' \cup \{t \leq s\}, \varepsilon \supset t_2 \leq s_2 \\ \Rightarrow \Sigma' \cup \{t \leq s\}, \varepsilon \supset t_1 \rightarrow t_2 \leq s_1 \rightarrow s_2 \Rightarrow \Sigma', \varepsilon \supset t \leq s$$

where we have the fresh variables r, r_1, r_2 for t, t_1, t_2 and u, u_1, u_2 for s, s_1, s_2 (the variables associated to an \rightarrow -node are inessential) then we generate the equations $r = r_1 \rightarrow r_2$ and $u = u_1 \rightarrow u_2$.

Case ($\mu\text{-assmp1}$). Say we are in the situation: $\Sigma', \varepsilon \supset a \leq b \Rightarrow \Sigma', \varepsilon \supset t \leq s$ where $a \leq b \in \Sigma$. If (r, u) is the pair of variables associated with the μ -node, we add a pair of equations: $r = a, u = b$.

Case ($\mu\text{-assmp2}$). Finally, if we visit a node in which we apply the rule (assmp_A) w.r.t. an assumption added during the computation then we do not generate any equation. In fact, the equations corresponding to those variables are defined in the corresponding μ -node in which the assumption was made.

Let us now consider the properties (a–d):

- (a) Follows from the use of fresh variables.

- (b) In the first place one establishes a relation R , say, between the variables reachable from t and those reachable from r . In general we will have a situation in which a variable t may correspond to many variables $r_1 \dots r_n$. Next, prove by induction on the lowest level of the appearance of r in the execution tree and $|\pi|$ that $(t, r) \in R$ implies $\text{Sol}(r, \theta)(\pi) = \text{Sol}(t, \varepsilon)(\pi)$.
- (c) By construction at each step we can apply the same computational rule.
- (d) This is a consequence of the constraint on the assignment of fresh variables to nodes. \square

5.4.2 Example. Consider the types: $\mu t. \top \rightarrow t \leq_{\mathcal{T}} \perp \rightarrow (\mu s. s \rightarrow s)$. These types are in minimal form, i.e., they are the smallest types that can describe the corresponding regular trees, but still the recursions are not in lockstep; we need to transform them into more redundant forms, in order to synchronize them. In the following we pedantically apply the procedure described in the proof of the previous lemma.

Let us assume that the types are described by the canonical system ε :

$$\begin{aligned}\varepsilon &\triangleq \varepsilon_1 \cup \varepsilon_2 \\ \varepsilon_1 &\triangleq \{t_1 = t_2 \rightarrow t_1, t_2 = \top\} \\ \varepsilon_2 &\triangleq \{s_1 = s_2 \rightarrow s_3, s_2 = \perp, s_3 = s_3 \rightarrow s_3\}.\end{aligned}$$

The following describes the successful execution tree associated to the initial goal $\emptyset, \varepsilon \supset t_1 \leq s_1$:

$$\begin{array}{ll}(\top_A) & (\text{assmp}_A) \\ \Rightarrow \{t_1 \leq s_1, t_1 \leq s_3\}, \varepsilon \supset s_3 \leq t_2 & \Rightarrow \{t_1 \leq s_1, t_1 \leq s_3\}, \varepsilon \supset t_1 \leq s_3 \\ & \Rightarrow \{t_1 \leq s_1, t_1 \leq s_3\}, \varepsilon \supset t_2 \rightarrow t_1 \leq s_3 \rightarrow s_3 \\ & \Rightarrow \{t_1 \leq s_1\}, \varepsilon \supset t_1 \leq s_3 \\ (\perp_A) & \vdots \\ \Rightarrow \{t_1 \leq s_1\}, \varepsilon \supset s_2 \leq t_2 & \\ & \Rightarrow \{t_1 \leq s_1\}, \varepsilon \supset t_2 \rightarrow t_1 \leq s_2 \rightarrow s_3 \\ & \Rightarrow \emptyset, \varepsilon \supset t_1 \leq s_1\end{array}$$

Observe that this execution tree *does not* have the one-expansion property, since the variable t_1 is expanded twice. Hence, we start associating fresh variables to each node according to the rules described in the proof. The following describes which rule is being applied at each node of the execution tree and which pair of fresh variables we associate to each node.

$$\begin{array}{ll}(\top_A)(u_6, r_6) & (\text{assmp}_A)(r_3, u_3) \\ & (\rightarrow_A)(r_5, u_5) \\ (\perp_A)(u_4, r_4) & (\mu_A)(r_3, u_3) \\ & (\rightarrow_A)(r_2, u_2) \\ & (\mu_A)(r_1, u_1)\end{array}$$

We now compute the new type environment $\theta = \theta_1 \cup \theta_2$, where:

$$\begin{aligned}\theta_1 &= \{r_1 = r_4 \rightarrow r_3, r_4 = \top, r_3 = r_6 \rightarrow r_3, r_6 = \top\}, \\ \theta_2 &= \{u_1 = u_4 \rightarrow u_3, u_4 = \perp, u_3 = u_6 \rightarrow u_3, u_6 = u_6 \rightarrow u_6\}.\end{aligned}$$

Observe here that the equation $u_6 = u_6 \rightarrow u_6$ is generated by calculating: $[u_6/s_3](s_3 = \varepsilon(s_3))$. No more equations are needed as s_3 is the only variable reachable from s_3 . Verify that:

$$\begin{aligned}T(\mu t. \top \rightarrow t) &= \text{Sol}(t_1, \varepsilon) = \text{Sol}(r_1, \theta), \\ T(\perp \rightarrow (\mu s. s \rightarrow s)) &= \text{Sol}(s_1, \varepsilon) = \text{Sol}(u_1, \theta).\end{aligned}$$

We finally compute the successful execution tree, *with one-expansion property*, associated to the initial goal $\emptyset, \theta \supset r_1 \leq u_1$:

$$\begin{array}{ll}(\top_A) & (\text{assmp}_A) \\ \Rightarrow \{r_1 \leq u_1, r_3 \leq u_3\}, \theta \supset u_6 \leq r_6 & \Rightarrow \{r_1 \leq u_1, r_3 \leq u_3\}, \theta \supset r_3 \leq u_3 \\ & \Rightarrow \{r_1 \leq u_1, r_3 \leq u_3\}, \theta \supset r_6 \rightarrow r_3 \leq u_6 \rightarrow u_3 \\ & \Rightarrow \{r_1 \leq u_1\}, \theta \supset r_3 \leq u_3 \\ (\perp_A) & \vdots \\ \Rightarrow \{r_1 \leq u_1\}, \theta \supset u_4 \leq r_4 & \\ & \Rightarrow \{r_1 \leq u_1\}, \theta \supset r_4 \rightarrow r_3 \leq u_4 \rightarrow u_3 \\ & \Rightarrow \emptyset, \theta \supset r_1 \leq u_1\end{array}$$

5.4.3 LEMMA (From the execution tree to the proof tree). *If $\vdash_A \Sigma, \varepsilon \supset t \leq s$ (see 4.2.3) and its execution tree has the one-expansion property, then $\vdash_R \Sigma \supset \langle t, \varepsilon \rangle \leq \langle s, \varepsilon \rangle$.*

PROOF. We proceed by induction on the depth k of the successful execution tree of an initial goal $\Sigma, \varepsilon \supset t \leq s$ (see 4.2). Depth is measured by the number of adjacent pairs of nodes (μ_A) – (\rightarrow_A) in the longest branch from the root. In the inductive case, each subgoal is converted into an initial goal of the same depth, in order to apply the induction hypothesis.

Case $k = 0$. The tree consists of a (μ_A) root (since the goal is initial) and a single leaf which is either (assmp_A) , (\perp_A) , (\top_A) , or (var_A) . Then after the application of the (μ_A) rule, with $s, t \in \text{Dom}(\varepsilon)$, we are in a terminal case $\Sigma \cup \{t \leq s\}, \varepsilon \supset \varepsilon(t) \leq \varepsilon(s)$.

Subcase (assmp_A) . $\Sigma \cup \{t \leq s\}, \varepsilon \supset a \leq b$, where $\varepsilon(t) = a$, $\varepsilon(s) = b$, and $a \leq b \in \Sigma$. Then $a, b \notin \text{Dom}(\varepsilon)$ (by definition of Tenv), and $\langle t, \varepsilon \rangle = \mu t. a = a$, $\langle s, \varepsilon \rangle = \mu s. b = b$. By (assmp_R) , $a \leq b \in \Sigma \Rightarrow \vdash_R \Sigma \supset a \leq b$. Conclude by (eq_R) : $\vdash_R \Sigma \supset \mu t. a \leq a$, $\vdash_R \Sigma \supset b \leq \mu s. b$, and (trans_R) .

Subcase (\perp_A) . $\Sigma \cup \{t \leq s\}, \varepsilon \supset \perp \leq \varepsilon(s)$, where $\varepsilon(t) = \perp$. Then $\langle t, \varepsilon \rangle = \mu t. \perp = \perp$, and we have $\vdash_R \Sigma \supset \perp \leq \langle s, \varepsilon \rangle$. Conclude by (eq_R) and (trans_R) .

Subcase (\top_A) . Similar.

Subcase (var_A). $\Sigma \cup \{t \leq s\}$, $\varepsilon \supset a \leq a$, where $\varepsilon(t) = \varepsilon(s) = a$ and $a \notin \text{Dom}(\varepsilon)$. Then $\langle t, \varepsilon \rangle = \mu t.a = a = \mu s.a = \langle s, \varepsilon \rangle$. We can apply (eq_R): $\Sigma \supset a \leq a$, then conclude with (eq_R) and (trans_R).

Case $k > 0$. The tree has a (μ_A) root with a (\rightarrow_A) child, hence $\varepsilon(t) = t_1 \rightarrow t_2$, $\varepsilon(s) = s_1 \rightarrow s_2$, where by definition of Tenv $t_1, t_2, s_1, s_2 \in \text{Dom}(\varepsilon)$:

$$\begin{aligned} & \Sigma \cup \{t \leq s\}, \varepsilon \supset s_1 \leq t_1, \Sigma \cup \{t \leq s\}, \varepsilon \supset t_2 \leq s_2 \\ & \Rightarrow \Sigma \cup \{t \leq s\}, \varepsilon \supset t_1 \rightarrow t_2 \leq s_1 \rightarrow s_2 \\ & \Rightarrow \Sigma, \varepsilon \supset t \leq s. \end{aligned}$$

We initially focus on one of the subgoals of depth $k - 1$:

$$(A) \Sigma \cup \{t \leq s\}, \varepsilon \supset t_2 \leq s_2.$$

Let us consider the following goal (B) , which we intend to subject, instead of (A) , to the induction hypothesis:

$$(B) \Sigma \cup \{t \leq s\}, \varepsilon' \supset \sigma(t_2) \leq \sigma(s_2)$$

where $\sigma \triangleq [t'/t, s'/s]$ is a substitution with fresh variables t' and s' , and $\varepsilon' \triangleq \sigma(\varepsilon \setminus t \setminus s) \cup \{t' = t, s' = s\}$.

First, we show that the goal (B) is initial. Since $\vdash_A \Sigma, \varepsilon \supset t \leq s$ is initial we have:

$$\begin{aligned} & \text{Vars}(\Sigma) \cap \text{Dom}(\varepsilon) = \emptyset \\ & \varepsilon = \varepsilon_1 \cup \varepsilon_2 \text{ with } \text{Dom}(\varepsilon_1) \cap \text{Dom}(\varepsilon_2) = \emptyset, \text{ such that } t \in \text{Dom}(\varepsilon_1), \\ & s \in \text{Dom}(\varepsilon_2). \end{aligned}$$

Hence we also have:

$$\begin{aligned} & t_1, t_2 \in \text{Dom}(\varepsilon_1) \text{ (only)}; s_1, s_2 \in \text{Dom}(\varepsilon_2) \text{ (only)} \\ & \varepsilon' = \varepsilon'_1 \cup \varepsilon'_2 \text{ where } \varepsilon'_1 \triangleq \sigma(\varepsilon_1 \setminus t) \cup \{t' = t\}, \quad \varepsilon'_2 \triangleq \sigma(\varepsilon_2 \setminus s) \cup \{s' = s\}. \end{aligned}$$

From which we conclude:

$$\begin{aligned} & \text{Vars}(\Sigma \cup \{t \leq s\}) \cap \text{Dom}(\varepsilon') = \emptyset \\ & \text{Dom}(\varepsilon'_1) \cap \text{Dom}(\varepsilon'_2) = \emptyset \\ & \sigma(t_2) \in \text{Dom}(\varepsilon'_1), \text{ because:} \\ & \quad \text{if } t_2 \equiv t \text{ then } \sigma(t_2) \equiv t' \text{ and } t' \in \text{Dom}(\varepsilon'_1); (t_2 \equiv s \text{ is not possible}) \\ & \quad \text{if } t_2 \not\equiv t \text{ then } \sigma(t_2) \equiv t_2; \text{ since } t_2 \in \text{Dom}(\varepsilon_1), \text{ we have } \sigma(t_2) \in \text{Dom}(\varepsilon'_1) \\ & \sigma(s_2) \in \text{Dom}(\varepsilon'_2), \text{ similarly.} \end{aligned}$$

Second, let $\text{Tree}(A)$ be the execution subtree of root (A) and $\text{Tree}(B)$ be the execution tree of root (B) . We show, by induction on the *length* of the longest path in $\text{Tree}(A)$, that we can build a tree T such that: (1) T has the same *depth* as $\text{Tree}(A)$; (2) T succeeds; (3) T expands the same variables as $\text{Tree}(A)$ in (μ_A) nodes, with the exception of t', s' ; (4) T has the one-expansion property; and (5) $T = \text{Tree}(B)$. (Hence, we also have $\vdash_A (B)$.)

We proceed by induction on each subgoal $\underline{A} = \underline{\Sigma} \cup \{t \leq s\}$, $\varepsilon \supset \underline{\alpha} \leq \underline{\beta}$ of $\text{Tree}(A)$, for which we build a subtree \underline{T} of the shape $\underline{\Sigma} \cup \{t \leq s\}$, $\varepsilon' \supset \sigma(\underline{\alpha}) \leq \sigma(\underline{\beta})$.

For the case (assmp_A) we have $\underline{\Sigma} \cup \{t \leq s\}$, $\varepsilon \supset t \leq s$ with $\underline{t} \leq \underline{s} \in \underline{\Sigma} \cup \{t \leq s\}$. By the properties of one-expansion noted in 5.4, we only need to consider the cases when either $\underline{t} = t$ and $\underline{s} = s$, or $\underline{t}, \underline{s}, t, s$ are pairwise distinct.

If $\underline{t} = t$, $\underline{s} = s$ then $\text{Tree}(\underline{A})$ is $\underline{\Sigma} \cup \{t \leq s\}$, $\varepsilon \supset t \leq s$, and \underline{T} is taken to be $\underline{\Sigma} \cup \{t \leq s, t' \leq s'\}$, $\varepsilon' \supset t \leq s \Rightarrow \underline{\Sigma} \cup \{t \leq s\}$, $\varepsilon' \supset t' \leq s'$, which is successful by (assmp_A) and (μ_A) , and has one-expansion. This \underline{T} is longer, but it still has depth 0.

If $\underline{t}, t, \underline{s}, s$ are pairwise distinct then $\text{Tree}(\underline{A})$ is $\underline{\Sigma} \cup \{t \leq s, t \leq s\}$, $\varepsilon \supset t \leq s$, and \underline{T} is taken to be $\underline{\Sigma} \cup \{t \leq s, t \leq s\}$, $\varepsilon' \supset t \leq s$, which is successful by (assmp_A), has one-expansion, and has depth 0.

For the case (μ_A) we must have, by one-expansion, $\underline{t}, t, \underline{s}, s$ pairwise distinct. Then $\text{Tree}(\underline{A})$ has the shape:

$$\underline{\Sigma} \cup \{t \leq s, t \leq s\}, \varepsilon \supset \varepsilon(\underline{t}) \leq \varepsilon(\underline{s}) \Rightarrow \underline{\Sigma} \cup \{t \leq s\}, \varepsilon \supset t \leq s$$

with $\underline{t}, \underline{s} \in \text{Dom}(\varepsilon)$.

Now $\sigma(\underline{t}) = t$, $\sigma(\underline{s}) = s$, and since $\underline{t}, \underline{s} \in \text{Dom}(\varepsilon')$ we have $\varepsilon'(\underline{t}) = \sigma(\varepsilon(\underline{t}))$, $\varepsilon'(\underline{s}) = \sigma(\varepsilon(\underline{s}))$. The tree \underline{T} is then chosen with the shape:

$$\underline{\Sigma} \cup \{t \leq s, t \leq s\}, \varepsilon' \supset \sigma(\varepsilon(\underline{t})) \leq \sigma(\varepsilon(\underline{s})) \Rightarrow \underline{\Sigma} \cup \{t \leq s\}, \varepsilon' \supset t \leq s$$

hence preserving success and depth by (μ_A) and the induction hypothesis. One-expansion is preserved because, by induction hypothesis, \underline{T} expands the same variables in (μ_A) node as $\text{Tree}(\underline{A})$, which has one-expansion; except that t', s' are expanded in the (assmp_A) case, but in the present situation $\underline{t}, \underline{s}, t', s'$ are distinct.

The other cases do not pose difficulties. One-expansion for the (\rightarrow_A) case follows from one-expansion of the two branches, since one-expansion is defined pathwise.

Hence, we can apply the induction hypothesis to (B) , obtaining:

$$\vdash_R \Sigma \cup \{t \leq s\} \supset \langle \sigma t_2, \varepsilon' \rangle \leq \langle \sigma s_2, \varepsilon' \rangle.$$

Then, by the equivalences $\langle \sigma t_2, \varepsilon' \rangle =_R \langle t_2, \varepsilon \setminus t \rangle$, $\langle \sigma s_2, \varepsilon' \rangle =_R \langle s_2, \varepsilon \setminus s \rangle$, and (eq_R), (trans_R):

$$\vdash_R \Sigma \cup \{t \leq s\} \supset \langle t_2, \varepsilon \setminus t \rangle \leq \langle s_2, \varepsilon \setminus s \rangle.$$

By a similar argument on $\Sigma \cup \{t \leq s\}$, $\varepsilon \supset s_1 \leq t_1$ we obtain:

$$\vdash_R \Sigma \cup \{t \leq s\} \supset \langle s_1, \varepsilon \setminus s \rangle \leq \langle t_1, \varepsilon \setminus t \rangle.$$

Finally:

$$\begin{aligned} &\Rightarrow \Sigma \cup \{t \leq s\} \supset \langle t_1, \varepsilon \setminus t \rangle \rightarrow \langle t_2, \varepsilon \setminus t \rangle \leq \langle s_1, \varepsilon \setminus s \rangle \rightarrow \langle s_2, \varepsilon \setminus s \rangle \quad (\rightarrow_R) \\ &\Leftrightarrow \Sigma \cup \{t \leq s\} \supset \langle \varepsilon(t), \varepsilon \setminus t \rangle \leq \langle \varepsilon(s), \varepsilon \setminus s \rangle \\ &\Rightarrow \Sigma \supset \mu t. \langle \varepsilon(t), \varepsilon \setminus t \rangle \leq \mu s. \langle \varepsilon(s), \varepsilon \setminus s \rangle \quad (\mu_R) \\ &\Leftrightarrow \Sigma \cup \langle t, \varepsilon \rangle \leq \langle s, \varepsilon \rangle \quad \square \end{aligned}$$

5.4.4 Example. We describe how to associate a proof tree to the execution tree with one-expansion property built in 5.4.2, by repeatedly applying the inductive proof just presented. For convenience we rewrite here the execution

tree:

$$\begin{array}{ll}
(\top_A) & (\text{assmp}_A) \\
\Rightarrow \{r_1 \leq u_1, r_3 \leq u_3\}, \theta \supset u_6 \leq r_6 & \Rightarrow \{r_1 \leq u_1, r_3 \leq u_3\}, \theta \supset r_3 \leq u_3 \\
& \Rightarrow \{r_1 \leq u_1, r_3 \leq u_3\}, \theta \supset r_6 \rightarrow r_3 \leq u_6 \rightarrow u_3 \\
& \Rightarrow \{r_1 \leq u_1\}, \theta \supset r_3 \leq u_3 \\
(\perp_A) & \vdots \\
\Rightarrow \{r_1 \leq u_1\}, \theta \supset u_4 \leq r_4 & \\
& \Rightarrow \{r_1 \leq u_1\}, \theta \supset r_4 \rightarrow r_3 \leq u_4 \rightarrow u_3 \\
& \Rightarrow \emptyset, \theta \supset r_1 \leq u_1
\end{array}$$

where $\theta = \theta_1 \cup \theta_2$, with:

$$\begin{aligned}
\theta_1 &= \{r_1 = r_4 \rightarrow r_3, r_4 = \top, r_3 = r_6 \rightarrow r_3, r_6 = \top\}, \\
\theta_2 &= \{u_1 = u_4 \rightarrow u_3, u_4 = \perp, u_3 = u_6 \rightarrow u_3, u_6 = u_6 \rightarrow u_6\}.
\end{aligned}$$

Proceeding from the root, we first fall on an inductive case. Hence, we reapply the procedure to the modified subgoals:

$$\{r_1 \leq u_1\}, \theta' \supset u_4 \leq r_4 \quad \{r_1 \leq u_1\}, \theta' \supset r_3 \leq u_3$$

where: $\theta' = \theta'_1 \cup \theta'_2$, $\sigma = [r'/r_1, u'/u_1]$,

$$\begin{aligned}
\theta'_1 &= \sigma(\theta_1 \setminus r_1) \cup \{r' = r_1\} = \{r_4 = \top, r_3 = r_6 \rightarrow r_3, r_6 = \top, r' = r_1\}, \\
\theta'_2 &= \sigma(\theta_2 \setminus u_1) \cup \{u' = u_1\} \\
&= \{u_4 = \perp, u_3 = u_6 \rightarrow u_3, u_6 = u_6 \rightarrow u_6, u' = u_1\}.
\end{aligned}$$

The first modified subgoal, $\{r_1 \leq u_1\}$, $\theta' \supset u_4 \leq r_4$, leads to a subcase (\perp_A) . Hence we have:

$$\vdash_R \{r_1 \leq u_1\} \supset \langle u_4, \theta' \rangle \leq \langle r_4, \theta' \rangle, \langle u_4, \theta' \rangle = \perp, \langle r_4, \theta' \rangle = \top. \quad (\text{a})$$

The second modified subgoal, $\{r_1 \leq u_1\}$, $\theta' \supset r_3 \leq u_3$, leads again to an inductive case. Hence, we generate two new modified subgoals:

$$\{r_1 \leq u_1, r_3 \leq u_3\}, \theta'' \supset u_6 \leq r_6 \quad \{r_1 \leq u_1, r_3 \leq u_3\}, \theta'' \supset r_3 \leq u_3$$

where: $\theta'' = \theta''_1 \cup \theta''_2$, $\sigma = [r''/r_3, u''/u_3]$,

$$\begin{aligned}
\theta''_1 &= \sigma(\theta'_1 \setminus r_3) \cup \{r'' = r_3\} = \{r_4 = \top, r_6 = \top, r' = r_1, r'' = r_3\}, \\
\theta''_2 &= \sigma(\theta'_2 \setminus u_3) \cup \{u'' = u_3\} = \{u_4 = \perp, u_6 = u_6 \rightarrow u_6, u' = u_1, u'' = u_3\}.
\end{aligned}$$

The first modified subgoal, $\{r_1 \leq u_1, r_3 \leq u_3\}$, $\theta'' \supset u_6 \leq r_6$, leads to subcase (\top_A) . Hence we have:

$$\begin{aligned}
\vdash_R \{r_1 \leq u_1, r_3 \leq u_3\} \supset \langle u_6, \theta'' \rangle \leq \langle r_6, \theta'' \rangle, \\
\langle u_6, \theta'' \rangle = \mu u_6. u_6 \rightarrow u_6, \langle r_6, \theta'' \rangle = \top. \quad (\text{b})
\end{aligned}$$

The second modified subgoal, $\{r_1 \leq u_1, r_3 \leq u_3\}$, $\theta'' \supset r_3 \leq u_3$, leads to a subcase (assmp_A) . Hence we have:

$$\begin{aligned}
\vdash_R \{r_1 \leq u_1, r_3 \leq u_3\} \supset \langle r_3, \theta'' \rangle \leq \langle u_3, \theta'' \rangle, \\
\langle r_3, \theta'' \rangle = r_3, \langle u_3, \theta'' \rangle = u_3. \quad (\text{c})
\end{aligned}$$

We can now build the proof tree, bottom up, using the proofs (a), (b), (c) as leaves:

$$\begin{array}{l}
\{r_1 \leq u_1, r_3 \leq u_3\} \supset \mu u_6. u_6 \rightarrow u_6 \leq \top \\
\{r_1 \leq u_1, r_3 \leq u_3\} \supset r_3 \leq u_3 \\
\{r_1 \leq u_1, r_3 \leq u_3\} \supset \top \rightarrow r_3 \leq (\mu u_6. u_6 \rightarrow u_6) \rightarrow u_3 \\
\{r_1 \leq u_1\} \supset \mu r_3. (\top \rightarrow r_3) \leq \mu u_3. ((\mu u_6. u_6 \rightarrow u_6) \rightarrow u_3) \\
\{r_1 \leq u_1\} \supset \perp \leq \top \quad \vdots \\
\{r_1 \leq u_1\} \supset (\top \rightarrow \mu r_3. (\top \rightarrow r_3)) \leq (\perp \rightarrow \mu u_3. ((\mu u_6. u_6 \rightarrow u_6) \rightarrow u_3)) \\
\emptyset \supset \mu r_1. (\top \rightarrow \mu r_3. (\top \rightarrow r_3)) \leq \mu u_1. (\perp \rightarrow \mu u_3. ((\mu u_6. u_6 \rightarrow u_6) \rightarrow u_3))
\end{array}$$

It just remains to observe the following equivalences to get back to the types we started with in 5.4.2:

$$\begin{aligned}
\mu r_1. (\top \rightarrow \mu r_3. (\top \rightarrow r_3)) &=_R \top \rightarrow \mu r_3. (\top \rightarrow r_3) =_R \mu r_3. (\top \rightarrow r_3), \text{ and} \\
\mu u_1. (\perp \rightarrow \mu u_3. ((\mu u_6. u_6 \rightarrow u_6) \rightarrow u_3)) &=_R \perp \rightarrow \mu u_3. ((\mu u_6. u_6 \rightarrow u_6) \rightarrow u_3) \\
&=_R \perp \rightarrow (\mu u_6. u_6 \rightarrow u_6). \quad \square
\end{aligned}$$

5.4.5 THEOREM (Completeness of the subtyping rules). *If $\alpha \leq_T \beta$ then $\alpha \leq_R \beta$.*

PROOF. If $\alpha \leq_T \beta$ then $\alpha \leq_A \beta$ by completeness of the algorithm (4.3.7). Consider the corresponding successful execution tree and apply the lockstep recursion Lemma 5.4.1, obtaining a tree for $\alpha' \leq_A \beta'$ with $\alpha =_T \alpha'$ and $\beta =_T \beta'$. By Lemma 5.4.3 we can now extract from the new execution tree a proof of $\alpha' \leq_R \beta'$. Applying the completeness of the rules for type equivalence we conclude $\alpha =_R \alpha'$ and $\beta =_R \beta'$. Finally we derive $\alpha \leq_R \beta$ by (eq_R) and (trans_R). \square

6. A PER MODEL

We sketch the main features of a model described in [3] (see also [14] for a related work) based on *complete uniform pers* over a $D_\infty \lambda$ -model [26].

Per (partial equivalence relation) models provide an interpretation of subtyping as set-theoretic containment of the relations [7]. In addition, these structures have very interesting categorical properties (in particular, cartesian closure and interpretation of second-order quantification as intersection; see [19]) that entail a satisfying interpretation of higher-order typed λ -calculi. The particular class of pers considered here preserves the previous properties while providing a solution of recursive domain equations up to equality. This result is obtained by an application of Banach's theorem on the uniqueness of the fixpoint of a contractive operator over a complete metric space.

6.1 Realizability Structure

Consider the functor $G(D) \triangleq D^D + D \times D + \text{At}$ defined in the category of complete partial orders (cpo's) and projection pairs. The cpo At is a collection of atomic values, and $+$ is the coalesced sum. The morphism part of G is standard.

The cpo D_x is the initial fixpoint of the functor G , that is, the colimit of the following ω -diagram:

$$D_0 \triangleq O \quad (O \text{ is the initial object; the cpo with one element})$$

$$D_{n+1} \triangleq D_n^{D_n} + D_n \times D_n + \text{At} = G(D_n)$$

with uniquely determined projection pairs $(i_{n,n+1}, j_{n+1,n}): D_n \rightarrow D_{n+1}$.

Let (i_n, j_n) be the projection pair between D_n and D_x . Let $e_n \triangleq i_n(j_n(e))$ for $e \in D_x$. We have $\sqcup_{n < \omega} \{e_n\} = e$, where “ \sqcup ” denotes, as usual, the join. The cpo's $D_x^{D_x}$ and $D_x \times D_x$ are projected into D_x by means of the projection pairs: (i, j) and $([_, _], p)$. The operation of application on D_x is defined as usual as: $fd \triangleq j(f)(d)$.

6.2 Complete Uniform Pers

A per A over D_x is complete and uniform² (henceforth, *cuper*) iff

- (1) $(\perp_{D_x}, \perp_{D_x}) \in A$ (\perp_{D_x} is the least element of the cpo D_x).
- (2) If $X \subseteq A$ is directed in $D_x \times D_x$ then $\sqcup X \in A$.
- (3) If $(e, e') \in A$ then $\forall n. (e_n, e'_n) \in A$.

We will consider the *full* subcategory of complete and uniform pers; therefore the *morphisms* are defined as usual as:

$$\text{cuper}[A, B] \triangleq \{f: D_x/A \rightarrow D_x/B \mid \exists \phi \in D_x. \forall d \in D_x.$$

$$(d, d) \in A \Rightarrow \phi d \in f([d]_A)\}$$

$$\text{where } [d]_A \triangleq \{e \in D_x \mid (d, e) \in A\}, \text{ and } D_x/A \triangleq \{[d]_A \mid (d, d) \in A\}.$$

Let $A_{|n} \triangleq A \cap i_n(D_n) \times i_n(D_n)$. Given A, B cupers we can define as for ideals (see [20]):

$$\text{closeness: } c(A, B) \triangleq \infty, \text{ if } A = B; \max\{n \mid A_{|n} = B_{|n}\}, \text{ o.w.}$$

$$\text{distance: } d(A, B) \triangleq 0, \text{ if } c(A, B) = \infty; 2^{-c(A, B)}, \text{ o.w.}$$

6.2.1 Subtype Interpretation. Following [9] and [7] we say that the *cuper* A is a subtype of the *cuper* B iff $A \subseteq B$. This is easily shown to correspond to the existence of a unique map in the category that is *realized* by the identity. Such maps play the role of *coercions* from A to B .

6.2.2 Type Interpretation. A *type environment* η is a map from type variables to cupers: $\eta: \text{Tvar} \rightarrow \text{cuper}$. A *type interpretation* of a type α in an environment η is written as $\llbracket \alpha \rrbracket \eta$.

In view of the interpretation of subtyping, the interpretation of type variables and type constants is naturally given as follows:

$$\llbracket \perp \rrbracket \eta \triangleq \{(\perp_{D_x}, \perp_{D_x})\} \quad \llbracket \top \rrbracket \eta \triangleq D_x \times D_x \equiv \text{Top} \quad \llbracket t \rrbracket \eta \triangleq \eta(t).$$

As we already mentioned, *cuper* is a cartesian closed category. In particular, given A, B cupers the *exponent* B^A is defined as follows:

$$(f, g) \in B^A \Leftrightarrow \forall d, e. (d, e) \in A \Rightarrow (fd, ge) \in B.$$

²A term suggested by M. Abadi and G. Plotkin.

This interpretation of the arrow is sometimes referred to as *simple*.

In general, every object $\exp(A, B)$ is *isomorphic* to the simple interpretation will enjoy the same categorical properties. Therefore, we assume \exp is a binary operator on cupers satisfying:

$$\exp(A, B) \cong B^A.$$

However, not any choice will be satisfying from our point of view. In order to complete the interpretation we need two more properties of the operator \exp , namely, *contractiveness* and (*anti*-)*monotonicity*.

6.2.3 Contractiveness. The set of cupers endowed with the metric d is a *complete metric space*. We require that the behavior of \exp at level $n + 1$ is determined by the value of the arguments up to level n :

$$\exp(A, B)_{|n+1} = \exp(A_{|n}, B_{|n})_{|n+1}.$$

Under this condition the exponentiation operator is *contractive* on the space (cuper, d) as it satisfies the following property:

$$A_{|n} = A'_{|n}, B_{|n} = B'_{|n} \Rightarrow \exp(A, B)_{|n+1} = \exp(A', B')_{|n+1}.$$

It turns out that every definable type operator is either contractive or the identity, and therefore admits a least fixpoint. The type-interpretation w.r.t a contractive exponent $\exp(A, B)$ is completed as follows:

$$\begin{aligned} \llbracket \alpha \rightarrow \beta \rrbracket \eta &\triangleq \exp(\llbracket \alpha \rrbracket \eta, \llbracket \beta \rrbracket \eta) & \llbracket \mu t. \alpha \rrbracket \eta &\triangleq Lfp(\lambda A. \llbracket \alpha \rrbracket \eta_{[A/t]}) \\ & & (Lfp &\equiv \text{least fixpoint}). \end{aligned}$$

6.2.4 Soundness of the (\rightarrow) Subtyping Rule. In order to have a sound interpretation of the (\rightarrow) rule in 3.1 it is convenient that the operator \exp satisfies the following additional condition:

$$A' \subseteq A, B \subseteq B' \Rightarrow \exp(A, B) \subseteq \exp(A', B').$$

Proviso. We can summarize our discussion as follows. We assume to have a binary operator, $\exp: \text{cuper} \times \text{cuper} \rightarrow \text{cuper}$, satisfying the following three properties, for any A, A', B, B' :

$$\begin{aligned} \exp(A, B) &\cong B^A \\ \exp(A, B)_{|n+1} &= \exp(A_{|n}, B_{|n})_{|n+1} \\ A' \subseteq A, B \subseteq B' &\Rightarrow \exp(A, B) \subseteq \exp(A', B'). \end{aligned}$$

The *simple* interpretation defined above provides an example of such operator. The *F*-interpretation discussed in 6.3 provides yet another example.

We can interpret the types parametrically in the operator \exp as follows:

$$\begin{aligned} \llbracket \perp \rrbracket \eta &\triangleq \{(\perp_{D_s}, \perp_{D_s})\} & \llbracket \top \rrbracket \eta &\triangleq D_\infty \times D_\infty \equiv \text{Top} & \llbracket t \rrbracket \eta &\triangleq \eta(t) \\ \llbracket \alpha \rightarrow \beta \rrbracket \eta &\triangleq \exp(\llbracket \alpha \rrbracket \eta, \llbracket \beta \rrbracket \eta) & \llbracket \mu t. \alpha \rrbracket \eta &\triangleq Lfp(\lambda A. \llbracket \alpha \rrbracket \eta_{[A/t]}) \\ & & (Lfp &\equiv \text{least fixpoint}). \end{aligned}$$

The three conditions above are also *sufficient* to obtain the following soundness theorem. We write $\models \alpha \leq \beta$ iff, given any operator \exp , with relative type-interpretation $[\]$, we have $[\alpha]\eta \subseteq [\beta]\eta$ for any $\eta: \text{Tvar} \rightarrow \text{cuper}$.

We also write $\models \Gamma \supset \alpha \leq \beta$. As usual this means: $\forall \eta. (\eta \models \Gamma \Rightarrow \eta \models \alpha \leq \beta)$.

6.2.5 THEOREM (Soundness of the tree ordering w.r.t. the model). *Given α, β types, if $\alpha \leq_T \beta$ then $\models \alpha \leq \beta$.*

PROOF (SKETCH). Given a per A we define its *completion* $\text{cmpl}(A)$ as the least cuper that contains A :

$$\text{cmpl}(A) \triangleq \bigcap \{B \text{ cuper} \mid A \subseteq B\}.$$

Given a tree A in $\text{Tree}(L)$ we define its interpretation as the completion of the set-theoretic union of the interpretations of its syntactic approximants:

$$[A]\eta \triangleq \text{cmpl}\left(\bigcup_{k < \omega} [A|_k]\eta\right).$$

It is easy to observe that $\{[A|_k]\eta \mid k < \omega\}$ is a growing chain of cupers.

Now we need the following fact (see [3]):

$$\forall n, \alpha. \exists N. \forall k \geq N. [(\alpha)_n]\eta = [(\alpha|_k)_n]\eta$$

where by definition $[(\beta)_n]\eta \triangleq [\beta]\eta \cap i_n(D_n) \times i_n(D_n)$.

In other words, if we are interested in the interpretation of the type α up to the n th level of the construction of D_∞ , it is enough to unfold α up to a certain level N and just consider the interpretation of this *finite* part of the associated tree expansion.

Next we use the fact that $[\alpha]\eta = \text{cmpl}(\bigcup_{n < \omega} [(\alpha)_n]\eta)$. From this we can conclude $[\alpha]\eta \subseteq [T\alpha]\eta$. Vice versa observe that $\forall k. [\alpha|_k]\eta \subseteq [\alpha]\eta$. Hence $[\alpha]\eta = [T\alpha]\eta$. Finally, $T\alpha \leq_\infty T\beta \Rightarrow \forall k. (\alpha|_k \leq \beta|_k) \Rightarrow \forall k. [\alpha|_k]\eta \subseteq [\beta|_k]\eta \Rightarrow [\alpha]\eta \subseteq [\beta]\eta$. \square

6.2.6 PROPOSITION (Soundness of the rule ordering w.r.t. the model). *If $\vdash_R \Gamma \supset \alpha \leq \beta$ then $\models \Gamma \supset \alpha \leq \beta$.*

PROOF. For the soundness of the type equivalence rules (5.1) one observes that the contractiveness of α in t is a sufficient (and necessary) condition to enforce the contractiveness of the following functional on the space cuper_{D_∞} (6.2):

$$G_{\alpha, \eta, t}(A) \triangleq [\alpha]\eta[A/t] \quad A \in \text{cuper}_{D_\infty}.$$

As for the subtyping rules (5.3) the problem is to check the soundness of (μ_R) . Suppose $\eta \models \Gamma$. By hypothesis we have:

$$\forall A, B \text{ cuper}. A \subseteq B \Rightarrow G_\alpha(A) \triangleq [\alpha]\eta[A/t] \subseteq [\beta]\eta[B/t] \triangleq G_\beta(B).$$

Therefore we have: $\forall n. G_\alpha^n(\text{Bot}) \subseteq G_\beta^n(\text{Bot})$, where $\text{Bot} = \{(\perp_{D_\infty}, \perp_{D_\infty})\}$.

It can be proved (see [3]) that for any type γ :

$$[(\mu t. \gamma)_n]\eta \triangleq [\mu t. \gamma]\eta \cap D_n \times D_n = G_\gamma^n(\text{Bot}) \cap D_n \times D_n.$$

And from $[\mu t. \gamma]\eta = \text{cmpl}(\bigcup_{n < \omega} [(\mu t. \gamma)_n]\eta)$ we have the thesis. \square

6.3 Completeness of an F -Interpretation

We now consider an F -interpretation of \rightarrow (see [27]) that is isomorphic to the simple interpretation and still satisfies the properties in 6.2.3 and 6.2.4. We will also use this interpretation for the completeness theorem 6.3.4.

Define: $(B^A)_F \triangleq B^A \cap F^2 \cup \{(\perp, f), (f, \perp), (f, f)\}$

where F is the embedding of the functional space $D_\infty^{D_\infty}$ into D_∞ and where f is the embedding of a distinct symbol of At into D_∞ .

Roughly speaking $(B^A)_F$ is built from B^A by selecting among these elements that are “functions” in the underlying λ -model D_∞ and by attaching to \perp a label f . We introduce the label f in order to distinguish the functional type $\top \rightarrow \perp$ from \perp (see Lemma 6.3.3). As an exercise one can try to give the complete rules for the “pure” version of the F -interpretation: $(B^A)_{F'} \triangleq B^A \cap F^2$. A more difficult exercise is to define a complete system for the simple semantics. In this case further identifications like $\mu t.t \rightarrow t = \top$ take place.

6.3.1 F -Theory of Subtyping. Rather than giving some abstract definition of a model that naively reflects the conditions for the soundness theorem and look for some ad hoc completeness result, we prefer to concentrate on a specific interpretation.

As a typical example, we *characterize the subtypings valid in every F -interpretation*. We write $\models_F \alpha \leq \beta$ iff for any type structure M constructed as just described we have $[\alpha]\eta \subseteq [\beta]\eta$ (or equivalently $\eta \models_F \alpha \leq \beta$) with respect to the induced F -interpretation and for any type environment η .

In order to prove the completeness of the theory it will be enough to use the *elementary substructure* of ideals. Ideals are cupers with just one equivalence class; they are closed w.r.t the standard operations over cupers.

Consider the type $\alpha \rightarrow \top$. Both in the simple interpretation and in the F -interpretation its meaning is essentially *independent from α* (this is not clearly the case for the tree equivalence).

In particular in the f -interpretation one has:

$$(\Phi) \quad \alpha \rightarrow \beta \leq \gamma \rightarrow \top$$

where $\gamma \rightarrow \top$ plays the role of *supertype of all the functional types* as:

$$[\gamma \rightarrow \top]\eta = [\perp \rightarrow \top]\eta = F^2 \cup \{(\perp, f), (f, \perp), (f, f)\}$$

Add to the subtyping system in 3.1 the axiom (Φ) . Denote with \vdash_Φ formal derivability in this new system. Write $\alpha \leq_\Phi \beta$ iff $\vdash_\Phi \alpha \leq \beta$.

By examining the 25 possible combinations of rules and axioms, it turns out that the relation \leq_Φ on the collection of nonrecursive types is a preorder (as in 3.1, one shows the transitive rule is derived by case analysis).

Next, extend the preorder \leq_Φ to recursive types by defining an ordering \leq_{Φ_∞} on trees as: $A \leq_{\Phi_\infty} B$ iff $\forall k. (A|_k \leq_\Phi B|_k)$. Also define $\alpha \leq_{\Phi_T} \beta$ iff $T\alpha \leq_{\Phi_\infty} T\beta$.

6.3.2 LEMMA. *Let α be a recursive type and η be a type environment. If $T\alpha \neq \top$ and for each type variable t free in α we have $\eta(t) \neq \text{Top}$ then $[\alpha]\eta \neq \text{Top}$.*

Table III

\leq	\perp	\top	s	$\perp \rightarrow \top$	$\alpha' \rightarrow \beta' (\beta' \neq \top)$
\perp	yes	yes	yes	yes	yes
\top	no	yes	no	no	no
t	no	yes	if $t = s$	no	no
$\perp \rightarrow \top$	no	yes	no	yes	no
$\alpha \rightarrow \beta (\beta \neq \top)$	no	yes	no	yes	$\alpha' \leq \alpha, \beta \leq \beta'$

PROOF. By induction on the structure of α . In particular, if $\alpha \equiv \mu t. \beta$ then either $T\alpha = \perp$ and the interpretation is the least cuper, or $T\alpha = t$, and we can use the hypothesis on η ; or the interpretation is a cuper A that solves the equation:

$$A = \left(G_1(A)^{G_2(A)} \right)_F \text{ for some definable operators } G_1 \text{ and } G_2.$$

This forces $A \subset \text{Top}$. \square

6.3.3 LEMMA (Separation). *Suppose D_∞ is an algebraic cpo. There is a type environment η such that whenever α (β) matches an element of the column (row) then $[\alpha]\eta \subseteq [\beta]\eta$ iff the situation described at the corresponding intersection occurs (see Table III).*

PROOF. As we already mentioned, it will be enough to consider ideals, that is, subsets of D_∞ with particular closure properties. Let us choose an environment η s.t. $\eta(t) = \{\perp, \lambda_t\}$ where λ_t is an element of the flat cpo At . Of course $t \neq s \Rightarrow \lambda_t \neq \lambda_s$ and $\lambda_t \neq f$.

The only interesting problem is to show that in the case $(\alpha \rightarrow \beta, \alpha' \rightarrow \beta')$ the condition $\alpha' \leq \alpha, \beta \leq \beta'$ is in fact necessary. First observe that $[\beta]\eta \subseteq [\beta']\eta$. Otherwise pick up $d \in [\beta]\eta \setminus [\beta']\eta$ and consider the constant map $\lambda x. d$ that belongs to $[\alpha \rightarrow \beta]\eta \setminus [\alpha' \rightarrow \beta']\eta$.

On the other hand, since $\beta' \neq \top$ by Lemma 6.3.2 $\exists e \in \text{Top} \setminus [\beta']\eta$. If the set $[\alpha']\eta \setminus [\alpha]\eta$ is not empty then it contains a compact element d_0 . Consider the continuous function $\text{step}_{d_0, e}$ that evaluates to e for elements greater than or equal to d_0 , and to \perp otherwise. Then such a function belongs to $[\alpha \rightarrow \beta]\eta \setminus [\alpha' \rightarrow \beta']\eta$. Note that we use the downward closure property of ideals to prove that the elements greater than or equal to d_0 do not belong to $[\alpha]\eta$. \square

6.3.4 PROPOSITION (Completeness for \leq_{Φ_T}). *Given recursive types α and β , $\models_F \alpha \leq \beta$ iff $\alpha \leq_{\Phi_T} \beta$.*

PROOF. The soundness follows from the discussion in 6.3.1 and the more general soundness result presented in 6.2.5.

For showing completeness, consider the type structure and the type environment η in Lemma 6.3.3. Given α, β , we want to show $\forall k. \alpha|_k \leq_{\Phi} \beta|_k$ whenever $\eta \models \alpha \leq \beta$.

Observe that the relation \leq_{Φ_T} is invariant under unfolding and under transformations of types of the shape $\alpha \rightarrow \top$ to $\perp \rightarrow \top$.

Fix k and unfold the types so that no μ appears before the k th level. Transform all the subtypes of the shape $\alpha \rightarrow \top$ in $\perp \rightarrow \top$.

Proceed by induction on k to show that the conditions in Table III force $\alpha_{|k} \leq_{\Phi} \beta_{|k}$. \square

In the remaining part of the paper we will be concerned with the tree ordering as it is very simple to analyze, and it is valid in every interpretation satisfying the conditions of Theorem 6.2.5. However, the previous study suggests that the tree ordering is very close to the model ordering so that, for example, the decision algorithm that we discuss in Section 4 for the former can be easily adapted to the latter.

7. COERCIONS

Coercions and subtyping are closely related topics; see for example [2] and [6]. We now show that the standard coercions $c_{\alpha, \beta}$ between two types $\alpha \leq \beta$ are definable in an extension of the basic calculus. This can be interpreted as saying that subtyping does not add any expressive power to such calculus (only convenience).

Then we show that the coercions implicit in a calculus with subsumption can be automatically synthesized. This fact is related to an algorithm for inferring the minimum type of a term.

7.1 Definability

In this section we show how to associate with each successful execution tree a λ -term whose denotation in the model is a coercion, that is, the unique map between the corresponding types that is realized by the identity.

7.1.1 Building the λ -Term. We can show that if we consider types up to tree equivalence, $=_T$, then for every initial goal $\Sigma, \varepsilon \supset t \leq s$ such that $\vdash_A \Sigma, \varepsilon \supset t \leq s$ there is a term $M(x_1, \dots, x_n): \langle t \rightarrow s, \varepsilon \rangle$ where $\Sigma = \{t_1 \leq s_1, \dots, t_n \leq s_n\}$ and $x_i (i = 1, \dots, n)$ are the free variables of M of type $\langle t_i \rightarrow s_i, \varepsilon \rangle$.

For the sake of readability the type labels on bound variables and on the fold and unfold constants are often omitted.

We recall that it is possible to define a *fixpoint combinator* as follows:

$$\mathbf{Y} \equiv \lambda f^{\alpha \rightarrow \alpha}. (\lambda x^{\mu t \rightarrow \alpha}. f((\text{unfold } x)x)) (\text{fold}(\lambda x^{\mu t \rightarrow \alpha}. f((\text{unfold } x)x))): \\ (\alpha \rightarrow \alpha) \rightarrow \alpha.$$

Proceed by induction on the structure of the execution tree (see 4.2.1). We refer to 4.1.7 for the properties 1. ... 6., of the translation $\langle -, - \rangle$:

Case (assmp) $x^{\langle t \rightarrow s, \varepsilon \rangle}$.

Case (\perp) $\lambda x^{\perp}. \mathbf{Y}(\lambda x^{\langle \beta, \varepsilon \rangle}. x): \langle \perp \rightarrow \beta, \varepsilon \rangle =_T \perp \rightarrow \langle \beta, \varepsilon \rangle$ by 1, 5.

Case (\top) $\lambda x^{\langle \alpha, \varepsilon \rangle}. \mathbf{Y}(\lambda x^{\top}. x): \langle \alpha \rightarrow \top, \varepsilon \rangle =_T \langle \alpha, \varepsilon \rangle \rightarrow \top$ by 2, 5.

Case (var) $\lambda x^a. x: \langle a \rightarrow a, \varepsilon \rangle =_T a \rightarrow a$ by 3, 5.

Case (\rightarrow) $\lambda f^{\langle \alpha \rightarrow \beta, \varepsilon \rangle}. \lambda x^{\langle \alpha', \varepsilon \rangle}. M_2(f(M_1(x))): \langle (\alpha \rightarrow \beta) \rightarrow (\alpha' \rightarrow \beta'), \varepsilon \rangle$ by 5.

Where by induction hypothesis $M_2: \langle \beta \rightarrow \beta', \varepsilon \rangle$ and $M_1: \langle \alpha' \rightarrow \alpha, \varepsilon \rangle$.

Case (μ) . By induction hypothesis we have $M(x^{\langle t \rightarrow s, \varepsilon \rangle}): \langle \varepsilon(t) \rightarrow \varepsilon(s), \varepsilon \rangle$. By 4, 5 $\langle t \rightarrow s, \varepsilon \rangle =_T \langle \varepsilon(t) \rightarrow \varepsilon(s), \varepsilon \rangle$; therefore we can type a term:

$$\mathbf{Y}(\lambda y^{\langle \varepsilon(t) \rightarrow \varepsilon(s), \varepsilon \rangle}. M(y)): \langle \varepsilon(t) \rightarrow \varepsilon(s), \varepsilon \rangle =_T \langle t \rightarrow s, \varepsilon \rangle, \text{ by 4.}$$

Remark. In a similar fashion one can associate a λ -term with a proof of the judgment $\Gamma \supset \alpha \leq \beta$ in the system in 5.3. The only difficulty arises for the rule (μ_R) . Suppose we have inductively built a term $M(x^{t \rightarrow s}): \alpha(t) \rightarrow \beta(s)$; then it is possible to transform it into a term $M'(x^{\mu t \alpha \rightarrow \mu s \beta}): [\mu t. \alpha / t] \alpha \rightarrow [\mu s. \beta / s] \beta$. The term associated with the conclusion of the (μ_R) rule can be defined as:

$$\mathbf{Y}(\lambda x^{\mu t \alpha \rightarrow \mu s \beta}. \lambda y^{\mu t \alpha}. (\text{fold } (M''(x)(\text{unfold } y))))).$$

7.1.2 PROPOSITION (Coercions are definable). *Let $\alpha, \beta \in \text{Type}$ and suppose $\alpha \leq_A \beta$. Let M be the term associated in 7.1.1 with the execution tree of $\emptyset, E\alpha \cup E\beta \supset \alpha^* \leq \beta^*$. Then the denotation of the term in the model is the unique coercion map from the interpretation of α to the interpretation of β .*

PROOF. Since we have not given the term interpretation explicitly (see [3]), we can only sketch an idea of the proof.

In the first place we need some facts about the interpretation of terms:

- (a) By erasing the type information and the constants fold, unfold from a typed term M , we obtain an untyped λ -term $er(M)$. We denote these untyped λ -terms with P, Q, \dots . It is a basic property of these interpretations that the interpretation of $er(M)$ gives a representative for the equivalence class that corresponds to the interpretation of M . We shortly refer to this fact by saying that $er(M)$ is a realizer for M .
- (b) Showing that the interpretation of M is a coercion from α to β means proving that the identity map, id, is a realizer for M . Equivalently id and $er(M)$ are equivalent in $\alpha \rightarrow \beta$. Note that here and in the following for the sake of readability we simply refer to syntactic objects, but we really intend to speak of their denotations in the model.
- (c) The realizer for \mathbf{Y} is an element Fix with functionality: $\lambda g. \sqcup g^n(\perp_{D_\lambda})$.

In order to prove the theorem by induction on the structure of the execution tree we need to generalize somewhat.

In the first place we observe that if in the execution tree of $\emptyset, \varepsilon \supset t \leq s$ we never use (assmp) then the interpretation of the associated term $M: \langle t \rightarrow s, \varepsilon \rangle$ is a coercion in $\langle t \rightarrow s, \varepsilon \rangle$.

However, this is not enough to make the induction go through in the case where the term $M(x^\Sigma)$ really depends on the assumption variable. We have to observe that $M(x^\Sigma)$ also enjoys a property of *contractiveness*.

Let us suppose that (μ) is the last rule applied. By construction assume we have a term $M(x)$ that is a functional from coercions to coercions. We would like to show that $\mathbf{Y}(\lambda x. M(x))$ is still a coercion.

Observe that after a (μ) rule we always have a (\rightarrow) rule. Therefore the term $M(x)$ has the structure $\lambda f. \lambda y. M_2(x)(f(M_1(x)y))$.

Now observe that a realizer for $\mathbf{Y}(\lambda x.M(x))$ will be something like $\sqcup g^n(\perp_{D_x})$ for $g = \lambda x.\lambda f.\lambda y.P_2(x)(f(P_1(x)y))$ where P_i is a realizer for M_i ($i = 1, 2$). We have to show that this realizer is equivalent to id in a type with the structure $C \equiv (A \Rightarrow B) \Rightarrow (A' \Rightarrow B')$, where $A \Rightarrow B \equiv \text{exp}(A, B)$. Since the type is a complete per, it will be enough to show that for each n $g^n(\perp_{D_x})$ is equivalent to id in the appropriate type.

To do this we need a last remark. Denote with $A_{|n}$ the approximation at the n th level of the cuper A as in 6.2. We observe that if $(P, \text{id}) \in C_n$ then $g(P), \text{id} \in C_{|n+1}$. This follows easily from the structure of g and the assumption (6.2.3). Hence we have $\forall n.(g^n(\perp_{D_x}), \text{id}) \in C_n$ that implies $(\sqcup g^n(\perp_{D_x}), \text{id}) \in C$. \square

7.2 Inference

Let $\lambda \rightarrow^\mu$ be the calculus in Section 2. Given a term in $\lambda \rightarrow^\mu$, possibly not typeable, we are interested in the problem of determining if it can be *well typed modulo the insertion of appropriate coercions*.

We refer to this problem as *coercion inference*. We will define a simple algorithm that, given a term M , succeeds exactly when M is typeable modulo the insertion of coercions. In this case the algorithm returns the *least type* among the types that can be assigned to M .

A similar problem was solved in [1] for a second-order lambda calculus with records and in [18] for a second-order lambda calculus including a form of bounded quantification.

All these results rely on the *structural properties of the subtype relation* that are stated, in this case, as Proposition 7.2.4.

Notation. In this section $\alpha \lesssim \beta$ and $\alpha \simeq \beta$ are shorthands for $T\alpha \leq_z T\beta$ and $T\alpha = T\beta$.

7.2.1 Typing Modulo Coercions. We can formalize the idea of *typing modulo coercions* in two ways:

- (a) **Subsumption.** Add to the typing system in 2.2 and 3.1 the following rule based on the tree order \leq_z . The version based on \leq_{fin} is often referred to as Subsumption:

$$(\text{Sub}_z) \quad M: \alpha, \alpha \lesssim \beta \Rightarrow M: \beta.$$

We denote formal derivability in this new system with \vdash_{Sub} .

- (b) **Explicit Coercions.** Extend the term language with a collection of constants $\{c_{\alpha, \beta} \mid \alpha, \beta \text{ types}\}$ and add to the typing system in 3.1 the following rule:

$$(\text{ExpCoer}_z) \quad M: \alpha, \alpha \lesssim \beta \Rightarrow (c_{\alpha, \beta} M): \beta.$$

Denote formal derivability in this new system with \vdash_c and denote the corresponding term language with $\lambda \rightarrow^{\mu c}$. Moreover, denote with er_c (mnemonic for erase coercions) the obvious function that takes a term in $\lambda \rightarrow^{\mu c}$, erases all the constants $c_{\alpha, \beta}$, and returns a term in $\lambda \rightarrow^\mu$.

The use of these rules is justified by the finitary axiomatization of \lesssim

given in Section 5. Note that in both these systems the $(\text{fold}_{\mu t \alpha} M)$ and $(\text{unfold}_{\mu t \alpha} M)$ terms become redundant.

7.2.2 Definition (coercion inference). We define inductively on the structure of the term M in $\lambda^{\rightarrow \mu}$ a function³

$$\text{CI}: (\lambda^{\rightarrow \mu}) \rightarrow (\lambda^{\rightarrow \mu^c} \cup \{\text{FAIL}\}) \quad (\text{CI for coercion inference})$$

that either fails or returns a well-typed term N in $\lambda^{\rightarrow \mu^c}$ such that $er_c(N) \equiv M$. The clauses (fold), (unfold) have priority on the clause (apl).

$$\begin{aligned} (\text{var}) \quad & \text{CI}(x^\alpha) \triangleq x^\alpha \\ (\text{abs}) \quad & \text{CI}(\lambda x^\alpha. M) \triangleq \text{if CI}(M): \beta \text{ then } \lambda x^\alpha. \text{CI}(M) \text{ else FAIL} \\ (\text{apl}) \quad & \text{CI}(MN) \triangleq \\ & \quad \text{if CI}(M): \alpha' \text{ and CI}(N): \gamma \text{ then} \\ & \quad \quad \text{if } \alpha' \simeq \alpha \rightarrow \beta \text{ and } \gamma \lesssim \alpha \text{ then } (c_{\alpha', \alpha \rightarrow \beta} \text{CI}(M))(c_{\gamma, \alpha} \text{CI}(N)) \\ & \quad \quad \text{else if } \alpha' \simeq \perp \text{ then } (c_{\alpha', \alpha \rightarrow \perp} \text{CI}(M)) \text{CI}(N) \text{ else FAIL} \\ & \quad \text{else FAIL} \\ (\text{fold}) \quad & \text{CI}(\text{fold}_{\mu t \alpha} M) \triangleq \\ & \quad \text{if CI}(M): \beta \text{ and } \beta \lesssim \mu t. \alpha \text{ then} \\ & \quad \quad \text{fold}_{\mu t \alpha}(c_{\beta, [\mu t \alpha / t] \alpha} \text{CI}(M)) \\ & \quad \text{else FAIL} \\ (\text{unfold}) \quad & \text{CI}(\text{unfold}_{\mu t \alpha} M) \triangleq \\ & \quad \text{if CI}(M): \beta \text{ and } \beta \lesssim \mu t. \alpha \text{ then} \\ & \quad \quad \text{unfold}_{\mu t \alpha}(c_{\beta, \mu t \alpha} \text{CI}(M)) \\ & \quad \text{else FAIL} \quad \square \end{aligned}$$

Clearly, CI can also be used to define an inference algorithm for \vdash_{Sub} ; just consider the type of the term synthesized by CI. We prove in 7.2.5 that this algorithm computes the minimal type of a term (if any). To achieve this result we need the following simple properties.

7.2.3 PROPOSITION. *Let M be a term in $\lambda^{\rightarrow \mu}$; then:*

- (1) $\vdash_{\text{Sub}} M: \alpha$ iff for some $N: \vdash_c N: \alpha$ and $er_c(N) \equiv M$.
- (2) If $\text{CI}(M): \beta$ then $er_c(\text{CI}(M)) \equiv M$.

PROOF. (1) Every introduction of an explicit coercion corresponds to an application of subsumption and vice versa. (2) By induction on the definition of CI. \square

7.2.4 PROPOSITION. *Let α, β, \dots be recursive types; then:*

- (1) If $\alpha \lesssim \beta_1 \rightarrow \beta_2$ then either $\alpha \simeq \perp$ or $\alpha \simeq \alpha_1 \rightarrow \alpha_2$, $\beta_1 \lesssim \alpha_1$, and $\alpha_2 \lesssim \beta_2$.
- (2) If $\alpha_1 \rightarrow \alpha_2 \lesssim \beta$ then either $\beta \simeq \top$ or $\beta \simeq \beta_1 \rightarrow \beta_2$, $\beta_1 \lesssim \alpha_1$, and $\alpha_2 \lesssim \beta_2$.

PROOF.

- (1) α can be rewritten, by unfolding, to an equivalent type of the shape \perp , \top , t , or $\alpha_1 \rightarrow \alpha_2$. The definition of the tree ordering and the hypothesis $\alpha \lesssim \beta_1 \rightarrow \beta_2$ lead to the conclusion by a simple case analysis.

³Actually the following specification determines a *class* of algorithms that suffices for our purposes.

(2) Analogous. \square

7.2.5 THEOREM (Terms have a least type). *Let M be a term in $\lambda^{\rightarrow \mu}$; then $\vdash_{Sub} M: \alpha$ implies $CI(M): \beta$ and $\beta \leq \alpha$.*

PROOF. By induction on the structure of M . $C(N)$ is a metanotation for $c\alpha_{n-1}, \alpha_n \dots (c\alpha_1, \alpha_2 N) \dots$, where: $N: \alpha_1, n \geq 1, \alpha_i \leq \alpha_{i+1}$. By virtue of 7.2.3.(1) we may equivalently assume the existence of a well-typed term N in $\lambda^{\rightarrow \mu^c}$ such that $er_c(N) \equiv M$. Observe the crucial role of property 7.2.4 in proving the rather surprising fact that the algorithm is *complete* in the sense just stated above.

Case $M \equiv x^\beta$. If $er_c(N) \equiv x^\beta$ then $N \equiv Cx^\beta: \alpha$ and $\beta \leq \alpha$. On the other hand $CI(x^\beta) \equiv x^\beta: \beta$.

Case $M \equiv (\lambda x^\alpha. M')$. If $er_c(N) \equiv \lambda x^\alpha. M'$ then $N \equiv C(\lambda x^\alpha. N'): \gamma$, $er_c(N') \equiv M'$, and $N': \beta'$. By induction hypothesis $CI(M'): \beta$ and $\beta \leq \beta'$; hence by definition $CI(\lambda x^\alpha. M'): \alpha \rightarrow \beta$. Note that $\alpha \rightarrow \beta' \leq \gamma$ by definition of N , and this implies (by 7.2.4) either $\gamma \simeq \top$ (and in this case we are done as $\alpha \rightarrow \beta \leq \top$) or $\gamma \simeq \gamma_1 \rightarrow \gamma_2$, $\gamma_1 \leq \alpha$ and $\beta' \leq \gamma_2$. In the latter case $\beta \leq \beta' \leq \gamma_2$ implies $\alpha \rightarrow \beta \leq \gamma$.

Case $M \equiv (M_1 M_2)$. If $er_c(N) \equiv M_1 M_2$ then $N \equiv C(N_1 N_2): \gamma$, $er_c(N_i) \equiv M_i$, $i = 1, 2$, $N_1: \gamma_1 \rightarrow \gamma_2$, $N_2: \gamma_1$. By induction hypothesis $CI(M_i): \beta_i$, $i = 1, 2$, $\beta_1 \leq \gamma_1 \rightarrow \gamma_2$ and $\beta_2 \leq \gamma_1$. From (7.2.4) follows that $\beta_1 \simeq \perp$ or $\beta_1 \simeq \beta'_1 \rightarrow \beta''_1$, $\gamma_1 \leq \beta'_1$, $\beta''_1 \leq \gamma_2$. In the first case $CI(M_1 M_2): \perp$ and we are done. In the second $CI(M_1 M_2): \beta''_1$ as $\beta_2 \leq \gamma_1 \leq \beta'_1$. Finally observe: $\beta''_1 \leq \gamma_2 \leq \gamma$.

Case $M \equiv (\text{fold}_{\mu t. \alpha} M')$. If $er_c(N) \equiv \text{fold}_{\mu t. \alpha} M'$ then $N \equiv C(\text{fold } N'): \gamma$, $er_c(N') \equiv M'$, $N': [\mu t. \alpha / t] \alpha \equiv \gamma'$, $\gamma' \leq \gamma$. By induction hypothesis $CI(M'): \beta'$, $\beta' \leq \gamma'$. Hence by definition, $CI(\text{fold } M'): \mu t. \alpha$, and we have $\mu t. \alpha \simeq \gamma' \leq \gamma$.

Case $M \equiv (\text{unfold}_{\mu t. \alpha} M')$. Analogous. \square

7.2.6 We can think of substituting the explicit coercions with the definable coercions constructed in Section 7.1 The resulting term is now typeable in an extension of the calculus in Section 2 including the rule: $M: \alpha, \alpha = \beta \Rightarrow M: \beta$. We recall that this rule is soundly interpreted by the model.

7.2.7 Observe that in general there are many possible well-typed terms of the same type to which the erase-coercions map assigns the *same term*, that is:

$$\vdash_c N_1: \alpha, \vdash_c N_2: \alpha \text{ and } er_c(N_1) \equiv er_c(N_2).$$

However, N_1 and N_2 receive the same interpretation in the model.

It is an appealing aspect of our semantic approach to the interpretation of subtyping that many hard coherence problems (see [18]) simply disappear, if we recall the uniqueness of the coercion in the model.

7.2.8 The following is a trivial example of a term that can be typed in the type system with subsumption but *not* in the system described in 2.2:

$$\vdash_{Sub} \lambda f^{t \rightarrow s}. \lambda x^t. (\lambda y^\top. x)(fx): (t \rightarrow s) \rightarrow t.$$

8. CONCLUSION

We have used a subtyping relation based on infinite trees as the central concept of our work. In our experience this relation has arisen naturally, giving insights about both the subtypings valid in certain per models and the behavior of the Amber implementation. In fact we have shown that this relation can be used to characterize sound and complete theories for a certain class of per models and that it can be simply and efficiently implemented. We have also shown the soundness and completeness of certain rules and the definability of coercions within the calculus (modulo a strengthening of the notion of type equality). Finally, we have observed that the whole process of inferring coercions and minimal types can be automated.

The notions of tree expansion and finite approximation (Section 3) can be easily adapted to larger languages, both with first-order type constructors like products, sums, records, and variants, and with higher-order type constructors like second-order universal quantification. The important point is that the tree resulting from the expansion is regular. Under this assumption it seems possible to adapt algorithms and rules to obtain results of soundness and completeness (Sections 4, 5). Caution is necessary in extensions to bounded quantification since some of those systems are undecidable [24].

About the relationship between the tree ordering and the model, we expect the extension of the soundness theorem (6.2) to be straightforward. On the other hand, we expect technical problems from the completeness theorem (6.3) when introducing higher-order type constructors like second-order universal quantification. In particular, in this case, it is not clear how to extend the separation lemma (6.3.3).

The result on the definability of the coercions has already been obtained for several calculi with records, variants, and bounded quantification (but without recursion). It is a reassuring result that shows that the subtyping theory is in good harmony with the calculus.

The fact that terms have a least type has a clear impact on the implementation of the type checker. This appears to be a very desirable property towards an automatic treatment of coercions. The result, at the present state of the art, clearly relies on the *structural properties* of the subtyping relation.

Finally, we observe that challenging extensions arise when dealing with nonground collections of subtyping assumptions (see [2]). In this case much work remains to be done.

ACKNOWLEDGMENTS

We would like to thank Martín Abadi for comments on an early draft, and the referees for their efforts in improving the presentation of the paper.

REFERENCES

1. AMADIO, R. Formal theories of inheritance for typed functional languages. Tech. Rep. TR 28/89, Dip. di Informatica, Univ. di Pisa, Pisa, Italy, 1989.
2. AMADIO, R. Typed equivalence, type assignment and type containment. In *Proceedings of Conditional and Typed Rewriting Systems 90. Lecture Notes in Computer Science*, 516. Springer-Verlag, New York, 1990.

3. AMADIO, R. Recursion over realizability structures. *Inf. Comput.* 91, 1 (1991), 55–85.
4. ARNOLD, A., AND NIVAT, M. The metric space of infinite trees. Algebraic and topological properties. *Fundamenta Inf.* III. (1980), 445–476.
5. BREAZU-TANNEN, V., COQUAND, T., GUNTER, C., AND SCEDROV, A. Inheritance as implicit coercion. *Inf. Comput.* 93, (1991), 172–221.
6. BREAZU-TANNEN, V., GUNTER, C., AND SCEDROV, A. Denotational semantics for subtyping between recursive types. Rep. MS-CIS 89 63, Logic of Computation 12, Dept. of Computer and Information Science, Univ. of Pennsylvania, Philadelphia, 1989.
7. BRUCE, K., AND LONGO, G. A modest model of records, inheritance and bounded quantification. *Inf. Comput.* 87, 1/2 (1990), 196–240.
8. CANNING, P., COOK, W., HILL, W., OLTHOFF, W., AND MITCHELL, J. C. F -bounded polymorphism for object-oriented programming. In *Proceedings of Functional Programming and Computer Architecture 89*. ACM Press, New York, 1989.
9. CARDELLI, L. A semantics of multiple inheritance. *Inf. Comput.* 76, (1988), 138–164.
10. CARDELLI, L. Amber, combinators and functional programming languages. In *Proceedings of the 13th Summer School of the LITP, Le Val D'Ajol* (Vosges, France, May 1985). *Lecture Notes in Computer Science*, 242. Springer-Verlag, New York, 1985.
11. CARDELLI, L. Typeful programming. In *Formal Description of Programming Concepts*, E. J. Neuhold and M. Paul, Eds. Springer-Verlag, New York, 1991, 431–507.
12. CARDELLI, L., AND LONGO, G. A semantic basis for Quest. *J. Functional Program.* 1, 4 (Oct. 1991), 417–458.
13. CARDELLI, L., AND WEGNER, P. On understanding types, data abstraction and polymorphism. *ACM Comput. Surv.* 17, 4 (Dec. 1985), 471–522.
14. CARDONE, F., AND COPPO, M. Type inference with recursive types: Syntax and semantics. *Inf. Comput.* 92, 1, 48–80.
15. COOK, W. A denotational semantics of inheritance. Ph.D. dissertation, Brown Univ., Providence, R.I., 1989.
16. COURCELLE, B. Fundamental properties of infinite trees. *Theor. Comput. Sci.* 25, (1983), 95–169.
17. COURCELLE, B. Equivalence and transformation of regular systems—Applications to recursive program schemes and grammars. *Theor. Comput. Sci.* 42, (1986), 1–122.
18. CURIEN, P.-L., AND GHELLI, G. Coherence of subsumption, minimum typing and type checking in F_{\leq} . *Math. Struc. Comput. Sci.* 2, (1992), 55–91.
19. HYLAND, M. A small complete category. *Ann. Pure Appl. Logic* 40, 2, (1989), 135–165.
20. MACQUEEN, D., PLOTKIN, G., AND SETHI, R. An ideal model for recursive polymorphic types. *Inf. Comput.* 71, (1986), 1–2.
21. MILNER, R. A complete inference system for a class of regular behaviors. *J. Comput. Syst. Sci.* 28 (1984), 439–466.
22. NELSON, G., ED. *Systems Programming with Modula-3*. Prentice-Hall, Englewood Cliffs, N.J., 1991.
23. PARK, D. M. R. Concurrency and automata on infinite sequences. In *Proceedings of the 5th GI conference. Lecture Notes in Computer Science*, 104. Springer-Verlag, New York, 1981, 167–183.
24. PIERCE, B. C. Bounded quantification is undecidable. In *Proceedings of the ACM Symposium on Principles of Programming Languages*. ACM, New York, 1992, 305–315.
25. SALOMAA, A. Two complete systems for the algebra of regular events. *J. ACM* 13, 1 (1966).
26. SCOTT, D. Continuous lattices. In *Toposes, Algebraic Geometry and Logic. Lecture Notes in Mathematics* 274. Springer-Verlag, New York, 1972, 97–136.
27. SCOTT, D. Data types as lattices. *SIAM J. Comput.* 5, (1976), 522–587.
28. VAN WIJNGAARDEN ET AL., EDS. *Revised Report on the Algorithmic Language Algol68*. Springer-Verlag, New York, 1976, 103–107.
29. WADSWORTH, C. The relation between computational and denotational properties for Scott's D_{ω} models of the lambda-calculus. *SIAM J. Comput.* 5, (1976), 488–521.

Received October 1990; revised December 1991 and August 1992; accepted September 1992