Fixed Point Calculus

Roland Backhouse*

January 24, 2000

Abstract

Fixed point calculus is about the solution of recursive equations defined by a monotonic endofunction on a partially ordered set. This tutorial discusses applications of fixed point calculus in the construction of computer programs, beginning with standard applications and progressing to recent research.

The basic properties of least and greatest fixed points are presented. Well-foundedness and inductive properties of relations are expressed in terms of fixed points. A class of fixed point equations, called "hylo" equations, is introduced. A methodology of recursive program design based on the use of hylo equations is presented. Current research on generalisations of well-foundedness and inductive properties of relations, making these properties relative to a datatype, is introduced.

*School of Computer Science and Information Technology, University of Nottingham, Nottingham NG8 1BB, England

1 Introduction

1.1 Fixed Point Equations

Formulating and solving equations is a fundamental mathematical activity and whole areas of mathematics are devoted to the methods for solving particular classes of equations — think, for example, of the differential calculus developed in order to solve differential equations. These lectures are about a class of equations called *fixed point* equations that is particularly important in computing.

Fixed point equations have a very simple form, namely

$$x = f.x \quad ,$$

where f is a given function and x is the unknown. A solution to the equation is called a "fixed point" of the function f because applying f to a solution leaves the value "fixed", i.e. unchanged.

1.2 Languages

In the literature on computing science, fixed point equations are most often called *recursive* equations because the unknown x "recurs" on the right side of the equation. Recursion was first used extensively in computing science in the now-classic Algol 60 report [Nau63] which defined the programming language Algol 60. The Algol 60 report introduced so-called "Backus-Naur Form" to define the syntax of the language. Here is a small, simplified extract from the language definition.

$$\langle \mathrm{Expression} \rangle \quad ::= \quad \langle \mathrm{Expression} \rangle + \langle \mathrm{Expression} \rangle \;\mid\; (\langle \mathrm{Expression} \rangle) \;\mid\; \langle \mathit{Variable} \rangle$$

This defines (arithmetic) expressions by recursion. The symbol "$::=$" can be read simply as "is" or, indeed, as an equality symbol. The symbol "$\mid$" is read as "or". Terms enclosed within angle brackets (thus $\langle \mathrm{Expression} \rangle$ and $\langle \mathit{Variable} \rangle$) are called "non-terminals" —these are the entities being defined— and the remaining symbols (here "$+$" and the two parentheses "(" and ")") are called the "terminal" symbols —these are the symbols that form words in the language being defined— . In this way the definition of $\langle \mathrm{Expression} \rangle$ is read as:

> An expression is either an expression followed by the symbol "$+$" followed by an expression, or a parenthesised expression, or a variable.

Assuming that x and y are variables, the following are examples of expressions:

$$x+y \quad , \quad (x) \quad , \quad x \quad , \quad x+x+y \quad , \quad y+(x+y) \quad .$$

The definition of $\langle Expression \rangle$ is an example of a fixed point equation in which the unknown x is $\langle Expression \rangle$ and the function f is a function from languages (sets of words) to languages. Language theory provides many good examples of fixed point calculus; we use it frequently as a source of examples.

1.3 Functions

Recursion is used extensively in mathematics for the definition of functions. An elementary example is the factorial function, which is specified by the equation in fac,

(1) $$\begin{aligned} fac.0 &= 1 \\ fac.n &= n*fac.(n-1), \text{ for } n>0. \end{aligned}$$

The definition of the factorial function also has the form of a *fixed point* equation although this is not so easy to see as in the previous example. To verify that this is indeed the case we need to rewrite the definition in the form

$$fac = \cdots \quad .$$

Using the notation $\langle x{:}x{\in}\mathrm{Type}{:}\mathrm{Exp}\rangle$ for a function that maps a value x of type Type to the value given by expression Exp, we have:

$$fac = \langle n{:}n{\in}\mathbb{N}{:}\textbf{if } n=0 \textbf{ then } 1 \textbf{ else } n*fac.(n-1)\rangle \quad .$$

Now, abstracting from fac on the right side of this equation, define the function $\mathcal{F}$ by

$$\mathcal{F} = \langle f{:}f{\in}\mathbb{N}{\leftarrow}\mathbb{N}{:}\langle n{:}n{\in}\mathbb{N}{:}\textbf{if } n=0 \textbf{ then } 1 \textbf{ else } n*f.(n-1)\rangle\rangle \quad .$$

This defines $\mathcal{F}$ to be an endofunction[1] on functions of type $\mathbb{N}{\leftarrow}\mathbb{N}$, the type of the factorial function. That is, the function $\mathcal{F}$ maps a function to natural numbers from natural numbers to a function to natural numbers from natural numbers. For example, applying $\mathcal{F}$ to the successor function $\langle n::n+1\rangle$ we get

$$\mathcal{F}.\langle n::n+1\rangle = \langle n::\textbf{if } n=0 \textbf{ then } 1 \textbf{ else } n*((n-1)+1)\rangle \quad .$$

(For brevity we have omitted the type information on n.) Simplifying,

$$\mathcal{F}.\langle n::n+1\rangle = \langle n::\textbf{if } n=0 \textbf{ then } 1 \textbf{ else } n^2\rangle \quad .$$

The characteristic feature of the factorial function is that its definition demands that

$$fac = \mathcal{F}.fac \quad .$$

That is, fac is a *fixed point* of $\mathcal{F}$.

[1] An *endofunction* is a function whose target and source are the same.

1.4 Datatypes

Recursion is also used in the definition of datatypes in programming languages. A definition of, for example, lists in a functional programming language looks like the following:

$$\mathsf{List}\ \alpha\ =\ \mathsf{Nil}\ \mid\ \mathsf{Cons}\ \alpha\ (\mathsf{List}\ \alpha)$$

This states that a list of α's (where α stands for an arbitrary type) is either Nil or the operator Cons applied to a value of type α and a list of α's.

The definition of List is not strictly a fixed-point equation. Rather than expressing the equality of List α and Nil | Cons α (List α) (misleadingly suggested by the equals sign) the declaration expresses an *isomorphism* between the type List α and the type $\mathbb{1}+\alpha\times(\mathsf{List}\ \alpha)$, the disjoint sum ("$+$") of the unit type (a set containing exactly one element, here denoted by "$\mathbb{1}$") and the cartesian product ("$\times$") of the type α and the type List α. We will return to this later.

1.5 Basic Assumptions

The very simple form of a fixed-point equation means that it is a very general notion which captures a very broad range of computational problems. But a theory that is too general is usually also too weak to be of practical value. In developing a useful fixed-point calculus we need to impose some non-trivial mathematical properties on the fixed point equations we consider. In these lectures we require two properties. The first property we demand is that the domain of solutions be a *complete lattice*. That is the domain of solutions is a partially ordered set such that suprema and infima of any shape always exist.

This requirement clearly holds in the case of languages. When a language is being defined it is required that the "alphabet" of the language is clearly stated. The *alphabet* of a language is just a finite set of symbols, and *words* in the language are finite sequences of symbols in the alphabet. If Σ denotes an alphabet then Σ^* is the notation used to denote the set of all words in that alphabet (that is, the set of all finite sequences — including the empty sequence— of symbols in the alpahabet). A language is thus a subset of Σ^* and, as is well known, the set of all subsets of a given set is a complete lattice under the subset ordering.

Although there is no ordering relation explicitly mentioned in the definition of $\mathtt{fac}$ (other than the equality relation, which is pretty uninteresting as an ordering relation) the theory we develop still applies by the simple device of viewing functions as special cases of binary relations. A binary relation between two sets A and B is a subset of the cartesian product $A\times B$ and the set of all subsets of $A\times B$ forms a complete lattice.

This will be our solution domain when solving a fixed-point equation that is intended to define a function to set A from set B.

The second requirement we impose is that the function f in the given fixed point equation be *monotonic* in the ordering on the solution domain. This requirement is also clearly met in the case of language definitions. The function $\langle X:: \{b\} \cup \{a\} \cdot X \cdot \{c\} \rangle$ is typical of the sort of functions that are used in language definitions. (A raised infix dot denotes the concatenation operator extended to sets.) It is the composition of three functions, the function $\langle X:: \{b\} \cup X \rangle$, which adds the word b to a language, the function $\langle X:: \{a\} \cdot X \rangle$, which concatenates a at the beginning of every word in a given language, and $\langle X:: X \cdot \{c\} \rangle$, which concatenates c at the end of every word in a given language. These functions are all clearly monotonic, and thus so is their composition.

1.6 Issues and Applications

An arbitrary fixed point equation may have no solutions, it may have exactly one solution, or it may have more than one solution. An important consequence of our two assumptions is that the fixed point equations we consider always have at least one solution. Indeed, the set of solutions is itself a complete lattice. In particular, a monotonic function on a complete lattice has a *least* and a *greatest* fixed point. Moreover, it has a unique fixed point if (and only if) the least and great fixed points coincide.

A major element of these lecture notes is how to manipulate fixed point equations. Finding out how to express the same quantity by different equations is important to understanding. Another important element is providing general, practical, sufficient conditions for a fixed point equation to have a unique solution. These foundational elements of fixed point calculus are discussed in depth in section 4.

An example of the sort of issue tackled in section 4 is the use of recursion to define a number of values simultaneously. The terminology "mutual recursion" is used. In the definition of Algol 60, for example, this is what the definition of $\langle Expression \rangle$ really looks like:

$$\begin{array}{lll} \langle Expression \rangle & ::= & \langle Expression \rangle + \langle Term \rangle \mid \langle Term \rangle \\ \langle Term \rangle & ::= & \langle Term \rangle \times \langle Factor \rangle \mid \langle Factor \rangle \\ \langle Factor \rangle & ::= & (\langle Expression \rangle) \mid \langle Variable \rangle \end{array}$$

This is still a fixed point equation, but it is an equation in the triple of languages

$$(\langle Expression \rangle, \langle Term \rangle, \langle Factor \rangle) \quad .$$

An important issue is whether a fixed point equation in a vector of values can be solved piecewise and, if so, how. This is the content of theorem 43.

The applications we present are all relevant to programming. The two main applications are developed in sections 5 and 6. Section 5 is about how well-foundedness and

inductivity properties are expressed in terms of fixed points, with application to the design of iterative, sequential programs. Section 6 extends the discussion of section 5 to a general class of recursive equations. The equations we consider are called "hylo" equations. A hylo equation has three elements, a so-called "relator" F, an F-algebra and an F-coalgebra. It is thus in this section that the theme of this summer school is addressed most directly. Section 2 is introductory; section 3 is also introductory but, even so, considers an important application of fixed point calculus, namely Kleene algebra, which is the algebra of choice, composition and iteration, three indispensible ingredients of any programming language.

2 Prefix Points

We begin our study of fixed points by introducing the notion of a "prefix" point. As an instance of where the notion commonly occurs consider the following example from language theory. The set $L = \{n: 0 \leq n: a^n b c^n\}$ over the alphabet $\{a,b,c\}$ is sometimes specified in the following way.

- The word b is in the set L.
- If w is in the set L then so is the word awc.
- Nothing else is in the set L.

Expressing the first two clauses in terms of set inclusion we see that L is required to satisfy the equation:

$$X:: \{b\} \subseteq X \wedge \{a\} \cdot X \cdot \{c\} \subseteq X \quad ,$$

which is equivalent to the equation:

$$X:: \{b\} \cup \{a\} \cdot X \cdot \{c\} \subseteq X \quad .$$

Now consider the function f mapping sets of words to sets of words defined by:

$$f.X = \{b\} \cup \{a\} \cdot X \cdot \{c\} \quad .$$

Then the requirement is that L be a so-called "prefix point" of f, i.e. L should satisfy the equation:

$$X:: f.X \subseteq X \quad .$$

What about the third clause: "Nothing else is in the set L"? Note that this clause is necessary to specify L since, for example, the set of all words over the alphabet $\{a,b,c\}$ is

a prefix point of the function f. One way to understand this clause is as the requirement that L be the *least* prefix point of the function f. Thus the complete specification of L is the equation

$$X :: f.X \subseteq X \wedge \forall\langle Y : f.Y \subseteq Y : X \subseteq Y\rangle \quad .$$

Here, the first conjunct is the requirement that X be a prefix point of f, and the second conjunct that f be at most all prefix points Y of f.

In this section we define the notions of least prefix point and least fixed point in a general setting, and we establish the most basic properties of these notions. You are recommended to instantiate the properties we establish on the example just discussed as you read in order to confirm your understanding.

Suppose $\mathcal{A} = (A, \sqsubseteq)$ is a partially ordered set and suppose f is a monotonic endo-function on $\mathcal{A}$. Then a *prefix point* of f is an element x of the carrier set A such that $f.x \sqsubseteq x$. A *least prefix point* of f is a solution of the equation

$$x :: f.x \sqsubseteq x \wedge \forall\langle y : f.y \sqsubseteq y : x \sqsubseteq y\rangle \quad .$$

A least prefix point of f is thus a prefix point of f that is smaller than all other prefix points of f. A *least fixed point* of f is a solution of the equation

$$x :: f.x = x \wedge \forall\langle y : f.y = y : x \sqsubseteq y\rangle \quad . \tag{2}$$

Rather than study *fixed* points of f we are going to study *prefix* points of f since the latter are mathematically more manageable than the former.

Exercise 3 Relations are often defined by just saying which pairs are in the relation. Implicitly the relation in question is defined to be the least solution of a certain equation. (Recall that relations are sets of pairs and thus ordered by set inclusion.)

An example is the following definition of the "at-most" relation on natural numbers (denoted as usual by the infix operator $\leq$). The "at-most" relation is the least relation satisfying, first, that for all natural numbers n,

$$0 \leq n$$

and, second, for all natural numbers m and n,

$$m+1 \leq n+1 \Leftarrow m \leq n \quad .$$

This defines the "at-most" relation as a least prefix point of a function from relations to relations. *What is this function?*

□

It is easy to see that a least prefix point of f is unique. Let x and x' be least prefix points of f. Then,

$$
\begin{array}{ll}
& x = x' \\
\equiv & \quad \{ \quad \text{anti-symmetry} \quad \} \\
& x \sqsubseteq x' \wedge x' \sqsubseteq x \\
\Leftarrow & \quad \{ \quad \forall\langle y : f.y \sqsubseteq y : x \sqsubseteq y\rangle \text{ with the instantiation } y := x' \\
& \qquad\quad \forall\langle y : f.y \sqsubseteq y : x' \sqsubseteq y\rangle \text{ with the instantiation } y := x \quad \} \\
& f.x' \sqsubseteq x' \wedge f.x \sqsubseteq x \\
\equiv & \quad \{ \quad x \text{ and } x' \text{ are prefix points of } f \quad \} \\
& \mathsf{true} \quad .
\end{array}
$$

A similar argument establishes that a least fixed point of f is unique.

Nothing is lost by studying least prefix points rather than least fixed points since, by the following calculation, a least prefix point of a monotonic function f is a fixed point of f.

$$
\begin{array}{ll}
& x = f.x \\
\equiv & \quad \{ \quad \bullet \quad f.x \sqsubseteq x \quad \} \\
& x \sqsubseteq f.x \\
\Leftarrow & \quad \{ \quad \bullet \quad \forall\langle y : f.y \sqsubseteq y : x \sqsubseteq y\rangle\,,\ y := f.x \quad \} \\
& f.(f.x) \sqsubseteq f.x \\
\Leftarrow & \quad \{ \quad f \text{ is monotonic} \quad \} \\
& f.x \sqsubseteq x \\
\equiv & \quad \{ \quad \bullet \quad f.x \sqsubseteq x \quad \} \\
& \mathsf{true} \quad .
\end{array}
$$

Since all fixed points are also prefix points it follows that a least prefix point of f is also a least fixed point of f.

Dual to the notion of least prefix point we can define the notion of "greatest postfix point": a *greatest postfix point* of endofunction $f \in (A, \sqsubseteq) \leftarrow (A, \sqsubseteq)$ is a least prefix point of the function $f \in (A, \sqsupseteq) \leftarrow (A, \sqsupseteq)$. Spelling this out in detail, a *postfix point* of f is an element x of the carrier set A such that $x \sqsubseteq f.x$ and a *greatest postfix point* of f is a solution of the equation

$$(4) \qquad x :: x \sqsubseteq f.x \wedge \forall\langle y : y \sqsubseteq f.y : y \sqsubseteq x\rangle \quad .$$

A *greatest fixed point* of f is a solution of the equation

$$x :: f.x = x \wedge \forall\langle y : f.y = y : y \sqsubseteq x\rangle \quad .$$

Since a least prefix point is a least fixed point, we immediately have the dual result that a greatest postfix point is also a greatest fixed point. A simple but important corollary is that function f has a unique fixed point if and only if the least prefix point of f equals the greatest postfix point of f.

Let us summarise what has been learnt in this section. Introducing the notation μf for a solution of the equation (2) and νf for a solution of the equation (4) we have:

Theorem 5 (Least Prefix Point) Suppose $(A, \sqsubseteq)$ is an ordered set and the function f of type $(A, \sqsubseteq) \leftarrow (A, \sqsubseteq)$ is monotonic. Then f has at most one least prefix point, μf, characterised by the two properties:

(6) $\quad f.\mu f \sqsubseteq \mu f$

and, for all $x \in A$,

(7) $\quad \mu f \sqsubseteq x \Leftarrow f.x \sqsubseteq x \quad .$

Moreover, the least prefix point of f is a fixed point of f:

(8) $\quad f.\mu f = \mu f \quad .$

□

Note that least fixed points are characterised by the combination of (6) and (7) or (8) and (7). Because (6) is weaker than (8) it is usual to use the former characterisation when it is required to *establish* that a particular value is a least prefix point, and the latter characterisation when it is required to *exploit* the fact that a particular value is a least prefix point.

Theorem 9 (Greatest Postfix Point) Suppose $(A, \sqsubseteq)$ is an ordered set. Suppose, also, that the function f of type $(A, \sqsubseteq) \leftarrow (A, \sqsubseteq)$ is monotonic. Then f has at most one greatest postfix point, νf, characterised by the two properties:

(10) $\quad \nu f \sqsubseteq f.\nu f$

and, for all $x \in A$,

(11) $\quad x \sqsubseteq \nu f \Leftarrow x \sqsubseteq f.x \quad .$

Moreover, the greatest postfix point of f is a fixed point of f:

(12) $\quad f.\nu f = \nu f \quad .$

□

Use of the rules (7) or (11) corresponds to using induction. We therefore refer to their use in calculations as *fixed point induction*. (Strictly we should say *least* fixed point induction or *greatest* fixed point induction but it will invariably be clear which of the two rules is meant.) Rules (8) and (12) are called *computation* rules because they are often used in programs as left-right rewrite rules, repeated application of which is used to compute a fixed point.

Both induction rules can of course be written as an implication rather than a follows-from. For example, (7) can be written as

$$f.x \sqsubseteq x \Rightarrow \mu f \sqsubseteq x \quad .$$

This is the way many would write the rule. We deviate from this practice because in almost all our calculations we will apply the rule from left to right in the form given in (7). This is independent of whether our goal is to *verify* a statement of the form $\mu f \sqsubseteq x$ or, given x, to *construct* an f satisfying $\mu f \sqsubseteq x$, or, given f, to *construct* an x satisfying $\mu f \sqsubseteq x$. The benefit of proceeding in this way is the replacement of complification steps by simplification steps.

Exercise 13 *Show that* $\mu f \sqsubseteq \nu f$.

□

2.1 A First Example

In this section we present a simple example of the Least Prefix Point theorem (theorem 5). The example illustrates why (7) really is an induction rule. It also illustrates the use of the computation rule (8) in a simple example of algorithm design.

Let L be the least solution of the equation

$$X :: \{a\} \cup \{b\} \cdot X \cdot X \subseteq X \quad .$$

This corresponds to a grammar having two productions $X ::= a$ and $X ::= bXX$. We first use (7) to prove that, for all words w in L, the number of a's in w is one more than the number of b's in w.

Let M be the set of all words w such that the number of a's in w is one more than the number of b's in w. Let $\#_a w$ denote the number of a's in w, and $\#_b w$ denote the number of b's in w. We have to show that $L \subseteq M$. We use fixed point induction:

$$\begin{array}{ll} & L \subseteq M \\ \Leftarrow & \quad \{ \quad \text{by definition } L = \mu f \text{ where } f = \langle X :: \{a\} \cup \{b\} \cdot X \cdot X \rangle \text{,} \end{array}$$

$$
\begin{array}{ll}
 & \qquad \text{induction: (7)} \quad \} \\
 & \{a\}\cup\{b\}\cdot M\cdot M \subseteq M \\
\equiv & \quad \{ \quad \text{set theory, definition of concatenation} \quad \} \\
 & a\in M \wedge \forall\langle x, y: x\in M \wedge y\in M: bxy\in M\rangle \\
\equiv & \quad \{ \quad \text{definition of } M \quad \} \\
 & \#_a a = \#_b a + 1 \\
 & \wedge\ \forall\langle x, y: \#_a x = \#_b x + 1 \wedge \#_a y = \#_b y + 1: \#_a(bxy) = \#_b(bxy) + 1\rangle \\
\equiv & \quad \{ \quad \text{definition of } \#_a \text{ and } \#_b \quad \} \\
 & \text{true} \\
 & \wedge\ \forall\langle x, y: \#_a x = \#_b x + 1 \wedge \#_a y = \#_b y + 1: \#_a x + \#_a y = 1 + \#_b x + \#_b y + 1\rangle \\
\equiv & \quad \{ \quad \text{arithmetic} \quad \} \\
 & \text{true} \quad .
\end{array}
$$

Note that this proof is not a proof by induction on the length of words in L. Introducing the length of words in L into the proof is spurious and only obscures the properties that are truly vital to the proof.

Now let us illustrate the use of the computation rule. The problem we consider is how to write a simple loop that determines whether a given word W is in the language L. At the same time we illustrate one of the most basic elements of the mathematics of program construction, namely the use of loop invariants.

For the purposes of the discussion we need two functions on words, fst and rest. The fst function returns the first symbol of a word of length at least one, and the rest function returns the remainder of the word. For example, $\text{fst}.(aba) = a$ and $\text{rest}.(aba) = ba$. We also use ε to denote the empty word (the word of length zero).

The problem is to determine the truth or falsity of $W\in L$. As L is a *fixed* point of its defining equation, we can expand this requirement as follows:

$$
\begin{array}{ll}
 & W\in L \\
\equiv & \quad \{ \quad \text{by (8), } L = \{a\}\cup\{b\}\cdot L\cdot L \quad \} \\
 & W \in \{a\}\cup\{b\}\cdot L\cdot L \\
\equiv & \quad \{ \quad \text{set calculus} \quad \} \\
 & W = a \vee (\text{fst}.W = b \wedge \text{rest}.W \in L\cdot L) \quad .
\end{array}
$$

The two tests $W = a$ and $\text{fst}.W = b$ can, of course, easily be implemented. This then leaves us with the task of implementing the test $\text{rest}.W \in L\cdot L$. Were we to repeat the

same small calculation beginning with the latter then we would end up with the task of implementing the test $\mathsf{rest}.(\mathsf{rest}.W) \in L{\cdot}L{\cdot}L$. This clearly suggests a generalisation of the original problem, namely, given natural number n and word w determine the truth or falsity of the statement $w \in L^n$ where, by definition, $L^0 = \{\varepsilon\}$ and $L^{n+1} = L{\cdot}L^n$.

With this generalised problem we begin our calculation again. First, by definition of L^0,

(14) $\quad w \in L^0 \equiv w = \varepsilon \quad .$

Second,

$$
\begin{array}{ll}
 & w \in L^{n+1} \\
\equiv & \quad \{ \quad L^{n+1} = L{\cdot}L^n \text{ and, by (8), } L = \{a\} \cup \{b\}{\cdot}L{\cdot}L \quad \} \\
 & w \in (\{a\} \cup \{b\}{\cdot}L{\cdot}L){\cdot}L^n \\
\equiv & \quad \{ \quad \text{concatenation distributes through union,} \\
 & \qquad \text{concatenation is associative} \quad \} \\
 & w \in \{a\}{\cdot}L^n \cup \{b\}{\cdot}L{\cdot}L{\cdot}L^n \\
\equiv & \quad \{ \quad \text{set calculus} \quad \} \\
 & (\mathsf{fst}.w = a \wedge \mathsf{rest}.w \in L^n) \vee (\mathsf{fst}.w = b \wedge \mathsf{rest}.w \in L^{n+2}) \quad .
\end{array}
$$

This calculation provides enough information to construct the required loop. The loop uses two variables, a word w and a natural number k, satisfying the loop invariant

$$W \in L \equiv w \in L^k \quad .$$

This invariant is established by the assignment:

$$w, k := W, 1 \quad .$$

Also, by the above calculation, it is maintained when $k = n+1$ for some number n by the assignment

$$
\begin{array}{ll}
\mathsf{if} & \mathsf{fst}.w = a \ \rightarrow \ w, k := \mathsf{rest}.w, k-1 \\
\Box & \mathsf{fst}.w = b \ \rightarrow \ w, k := \mathsf{rest}.w, k+1 \\
\mathsf{fi} & .
\end{array}
$$

(Here we make the implicit assumption that the given word W is a string of a's and b's and contains no other characters, thus avoiding stipulating the semantics of the statement when neither of the two tests evaluates to true.) This statement also makes progress by reducing the length of w by 1 at each iteration. Finally, if we impose the

termination condition $w=\varepsilon \vee k=0$, then the conjunction of the termination condition and the invariant

$$(w=\varepsilon \vee k=0) \wedge (W \in L \equiv w \in L^k)$$

implies, using (14) and the calculation immediately following it, that

$$W \in L \equiv w=\varepsilon \wedge k=0 \quad .$$

This then is the complete program:

$w, k := W, 1$

$\{$ Invariant: $W \in L \equiv w \in L^k$.

Bound function: length of w $\}$

; do $\neg(w=\varepsilon \vee k=0) \rightarrow$ if $\mathsf{fst}.w=a \rightarrow w, k := \mathsf{rest}.w, k-1$

$\square$ $\mathsf{fst}.w=b \rightarrow w, k := \mathsf{rest}.w, k+1$

fi

od

$\{$ $(w=\varepsilon \vee k=0) \wedge (W \in L \equiv w \in L^k)$ $\}$

$\{$ $W \in L \equiv w=\varepsilon \wedge k=0$ $\}$

3 Kleene Algebra

The purpose of this section is to provide some practice in the use of the characterising properties of least fixed points, in particular the induction rule. For this purpose we consider an algebraic structure called a "Kleene algebra" by Kozen [Koz94].

A Kleene algebra has two constants 0 and 1, two binary operators $+$ and $\cdot$, and a unary operator *. These operators satisfy a number of axioms which we shall give shortly. The name "Kleene algebra" is a tribute to S. C. Kleene [Kle56] who postulated an algebraic structure as a basis for studying the behaviour of finite state machines. He called it the "algebra of regular events", which nowadays is most commonly abbreviated to "regular algebra".

Regular algebra is central to the mathematics of program construction because the three operators capture the essential properties of the three main ingredients of programming languages, choice ($+$), sequencing ($\cdot$) and iteration (*). The axioms of a Kleene algebra are not sufficient to capture all the properties of choice, sequencing and iteration in programming languages and we shall need to add to the axioms later on. Any Kleene algebra is thus also a regular algebra but it is not necessarily the case that any regular algebra is a Kleene algebra. See sections 4.2 and 4.3 for further discussion.

3.1 The Axioms

It is convenient to discuss the axioms of a Kleene algebra in two stages, first the axiomatisation of the $+$ and $\cdot$ operators, and then the axiomatisation of the * operator. In presenting the axioms we use variables a, b, x, y and z to range over the carrier set of the algebra. These variables are universally quantified in the usual way.

The notation "$+$" and "$\cdot$" is chosen to suggest a connection with addition and multiplication in ordinary arithmetic. Indeed, the $+$ operator of a Kleene algebra is required to be associative and symmetric and to have 0 as unit element:

$$(x+y)+z = x+(y+z) \quad ,$$

$$x+y = y+x \quad ,$$

and $$x+0 = x = 0+x \quad .$$

Also, just as in arithmetic, multiplication is required to be associative, to distribute over addition, and to have 1 as unit and 0 as zero element:

$$x\cdot(y\cdot z) = (x\cdot y)\cdot z \quad ,$$

$$x\cdot(y+z) = (x\cdot y)+(x\cdot z) \quad ,$$

$$(y+z)\cdot x = (y\cdot x)+(z\cdot x) \quad ,$$

$$x\cdot 0 = 0 = 0\cdot x \quad ,$$

and $$1\cdot x = x = x\cdot 1 \quad .$$

Use of these rules in calculations will be referred to simply as "arithmetic".

The addition and multiplication operators of a Kleene algebra deviate from the operators of normal arithmetic in two respects. First, multiplication is *not* assumed to be commutative. Second, addition is assumed to be idempotent. That is,

$$x+x = x$$

for all x. (Equivalently, $1+1 = 1$.)

Because addition is idempotent, associative and symmetric, the relation $\leq$ defined by

$$x \leq y \equiv x+y = y$$

is reflexive, transitive and anti-symmetric. In short, it is a partial ordering relation. Moreover, because 0 is the unit of addition, it is the least element in the ordering:

$$0 \leq x$$

for all x. Finally, addition and multiplication are both monotonic with respect to the $\leq$ ordering.

Exercise 15 *Verify all the claims made in the last paragraph. Prove also that* $+$ *is the binary supremum operator.* That is,

$$x+y \leq z \equiv x \leq z \wedge y \leq z \quad .$$

□

The * operator will be the focus of our attention. The axioms for * require that $a^* \cdot b$ be the least prefix point of the (monotonic) function mapping x to $b+a \cdot x$ and that $b \cdot a^*$ be the least prefix point of the (monotonic) function mapping x to $b+x \cdot a$. Formally, $a^* \cdot b$ is a prefix point of the function mapping x to $b+a \cdot x$:

$$b+a \cdot (a^* \cdot b) \leq a^* \cdot b \quad , \tag{16}$$

and is the least among all such prefix points:

$$a^* \cdot b \leq x \Leftarrow b+a \cdot x \leq x \quad . \tag{17}$$

That is,

$$a^* \cdot b = \mu\langle x :: b+a \cdot x \rangle \quad . \tag{18}$$

Similarly, $b \cdot a^*$ is a prefix point of the function mapping x to $b+x \cdot a$:

$$b+(b \cdot a^*) \cdot a \leq b \cdot a^* \quad , \tag{19}$$

and is the least among all such prefix points:

$$b \cdot a^* \leq x \Leftarrow b+x \cdot a \leq x \quad . \tag{20}$$

That is,

$$b \cdot a^* = \mu\langle x :: b+x \cdot a \rangle \quad . \tag{21}$$

The axioms (16) and (19) are thus instances of (6) and the axioms (17) and (20) are instances of (7), the two properties characterising least prefix points. An immediate corollary is that $a^* \cdot b$ and $b \cdot a^*$ are *fixed* points of the relevant functions:

$$b+a \cdot (a^* \cdot b) = a^* \cdot b \quad , \tag{22}$$

$$b+(b \cdot a^*) \cdot a = b \cdot a^* \quad . \tag{23}$$

This concludes the axiomatisation of a Kleene algebra. There are several interpretations of the operators in a Kleene algebra that fulfill the axioms. Table 3.1 summarises a few. The first interpretation in the table, in which the carrier consists of sets of words

Table 1: Kleene algebras

	carrier	$+$	$\cdot$	0	1	a^*	$\leq$
Languages	sets of words	$\cup$	$\cdot$	ϕ	$\{\varepsilon\}$	a^*	$\subseteq$
Programming	binary relations	$\cup$	$\circ$	ϕ	id	a^*	$\subseteq$
Reachability	booleans	$\vee$	$\wedge$	false	true	true	$\Rightarrow$
Shortest paths	nonnegative reals	max	$+$	∞	0	0	$\geq$
Bottlenecks	nonnegative reals	max	min	0	∞	∞	$\leq$

over a given, fixed alphabet, is the original application of Kleene algebra. Here the addition operation is set union, the multiplication operation is concatenation of sets of words, 0 is the empty set, 1 is the set whose sole element is the empty word, and a^* is the set of words formed by concatenating together an arbitrary number of words in the set a.

The second interpretation is the one we will return to most often in these lecture notes. In this interpretation the carrier set of the algebra is the set of binary relations over some state space, the addition operator is set union, multiplication is the composition operator on relations, 0 is the empty relation, 1 is the identity relation, and the iteration operator is the reflexive, transitive closure operator on relations.

The remaining three interpretations all concern path-finding problems; in each case the iteration operator (*) is uninteresting. These interpretations only become interesting when one considers graphs with edge labels drawn from these primitive Kleene algebras, the crucial theorem being that it is possible to define a Kleene algebra of graphs given that the edge labels are elements of a Kleene algebra [BC75, Koz94].

3.2 Reflexive, Transitive Closure

It is common to call a^* *the reflexive, transitive closure* of a. This is because a^* is reflexive, in the sense that

$$1 \leq a^* \quad ,$$

transitive, in the sense that

$$a^* = a^* \cdot a^* \quad ,$$

and * is a closure operator, i.e.

$$a \leq b^* \equiv a^* \leq b^* \quad .$$

In this section we shall verify, from the axioms given above, that a^* does indeed have these and other related properties. The main tool will be fixed point induction, i.e. axioms (17) and (20).

The reflexivity of a^* is immediate from (22) or (23). Instantiating b to 1 in both these rules and using $1 \cdot x = x = x \cdot 1$ we get

$$1 + a \cdot a^* = a^* = 1 + a^* \cdot a \quad .$$

Thus, since $+$ is the binary supremum operator, $1 \leq a^*$. It then follows that $a \leq a^*$ since

$$a^* = 1 + a \cdot a^* = 1 + a \cdot (1 + a \cdot a^*) = 1 + a \cdot 1 + a \cdot (a \cdot a^*) = 1 + a + a \cdot a \cdot a^* \quad .$$

Now for the transitivity of a^* we use a ping-pong argument:

$$\begin{array}{ll} & a^* = a^* \cdot a^* \\ \equiv & \quad \{ \quad \text{antisymmetry} \quad \} \\ & a^* \leq a^* \cdot a^* \wedge a^* \cdot a^* \leq a^* \quad . \end{array}$$

The first inequality does not involve induction:

$$\begin{array}{ll} & a^* \leq a^* \cdot a^* \\ \Leftarrow & \quad \{ \quad \text{transitivity} \quad \} \\ & a^* \leq 1 \cdot a^* \leq a^* \cdot a^* \\ \equiv & \quad \{ \quad \text{arithmetic for the first inequality;} \\ & \qquad 1 \leq a^* \text{ and multiplication is monotonic for the second} \quad \} \\ & \mathsf{true} \quad . \end{array}$$

The second inequality is where we need to use induction:

$$\begin{array}{ll} & a^* \cdot a^* \leq a^* \\ \Leftarrow & \quad \{ \quad (17) \text{ with } a,b,x := a, a^*, a^* \quad \} \\ & a^* + a \cdot a^* \leq a^* \end{array}$$

$\equiv$ { $+$ is the binary supremum operator }

$a^* \leq a^* \wedge a \cdot a^* \leq a^*$

$\equiv$ { reflexivity of $\leq$; $a^* = 1 + a \cdot a^*$ }

true .

Thus we have established that $a^* = a^* \cdot a^*$.

We establish that * is a closure operator by the following ping-pong argument.

$a^* \leq b^*$

$\Leftarrow$ { (17) with $a,b,x := a, a, b^*$ }

$1 + a \cdot b^* \leq b^*$

$\equiv$ { $+$ is the binary supremum operator }

$1 \leq b^* \wedge a \cdot b^* \leq b^*$

$\Leftarrow$ { $b^* = 1 + b \cdot b^*$, $b^* = b^* \cdot b^*$, $\cdot$ is monotonic }

$a \leq b^*$

$\Leftarrow$ { $a \leq a^*$ }

$a^* \leq b^*$.

Exercise 24 *Prove the following properties*:

(a) $a \cdot b^* \leq c^* \cdot a \Leftarrow a \cdot b \leq c \cdot a$

(b) $c^* \cdot a \leq a \cdot b^* \Leftarrow c \cdot a \leq a \cdot b$

(c) $a \cdot (b \cdot a)^* = (a \cdot b)^* \cdot a$

(d) $(a+b)^* = b^* \cdot (a \cdot b^*)^* = (b^* \cdot a)^* \cdot b^*$

(e) $(a^*)^* = a^*$

Properties (a) and (b) are called *leapfrog* rules (because a "leapfrogs" from one side of a star term to the other). Both have the immediate corollary that * is monotonic (by taking a to be 1). Properties (c) and (d) are called the *mirror* rule and *star decomposition* rule, respectively. Property (e) states that * is idempotent.

□

Exercise 25 The goal of this exercise is to show that a^* is the least reflexive and transitive element of the algebra that is at least a.

Let a^+ denote $a \cdot a^*$. We call a^+ the *transitive closure* of a. *Prove the following properties of* a^+.

(a) $a \leq a^+$

(b) $a^+ = a \cdot a^* = a^* \cdot a$

(c) $a^+ \cdot a^+ \leq a^+$

(d) $a^+ \leq x \Leftarrow a + x \cdot x \leq x$

Note that the combination of (a), (c) and (d) expresses the fact that a^+ is the least prefix point, and hence the least fixed point, of the function $\langle x :: a + x \cdot x \rangle$. That is, a^+ is the least transitive element that is at least a.

(e) *Show that* a^* *is the least fixed point of the function* $\langle x :: 1 + a + x \cdot x \rangle$.

□

4 Fixed Point Calculus

The least fixed point of a monotonic function is, as we have seen in theorem 5, characterised by two properties. It is a fixed point, and it is least among all prefix points of the functions. This gives us two calculational rules for reasoning about the least fixed point μf of monotonic function f: the *computation rule*

$$\mu f = f.\mu f$$

and the *induction rule*: for all x,

$$\mu f \sqsubseteq x \Leftarrow f.x \sqsubseteq x \quad .$$

In principle these are the only rules one needs to know about least fixed points. Every calculation can be reduced to one involving just these two rules (together with the rules pertaining to the specific problem in hand). This indeed is commonly what happens. Many postulates are verified by induction. But this is often not the most effective way of reasoning about fixed points.

The problem is that this basic characterisation of least fixed points typically invites proof by mutual inclusion. That is, given two expressions, E and F, involving fixed points (or prefix points) the obvious strategy to determine conditions under which they are equal is to determine conditions under which the two inclusions $E \sqsubseteq F$ and $F \sqsubseteq E$ hold. This can lead to long and clumsy calculations. To avoid this we state a number of equational properties of least fixed points. Their proofs, which are mostly straightforward, are left as exercises. We conclude the section with the mutual recursion theorem, which expresses the solution of mutually recursive fixed point equations in terms of the

successive solution of individual equations. The proof, which we do include, is a fine illustration of the earlier properties.

Unless otherwise stated, we assume throughout this section that f and g are monotonic endofunctions on the complete lattice $\mathcal{A}=(A, \sqsubseteq)$. (The assumption of completeness can be avoided but is convenient and practical.) The set of such functions also forms a complete lattice under the *pointwise ordering* $\dot{\sqsubseteq}$ defined by

$$f \dot{\sqsubseteq} g \equiv \forall\langle x :: f.x \sqsubseteq g.x\rangle \quad .$$

The set of prefix points of f is denoted by $\mathsf{Pre}.f$ and its set of postfix points is denoted by $\mathsf{Post}.f$. The set $\mathsf{Pre}.f$ is also a complete lattice, as is $\mathsf{Post}.f$.

4.1 Basic Rules

The most basic fixed point rule is that the least-fixed-point operator is monotonic:

(26) $\quad \mu f \sqsubseteq \mu g \Leftarrow f \dot{\sqsubseteq} g \quad .$

The second rule is called the *rolling rule*:

(27) $\quad g.\mu(f \bullet g) = \mu(g \bullet f)$

(where $f \in (A, \sqsubseteq) \leftarrow (B, \preceq)$ and $g \in (B, \preceq) \leftarrow (A, \sqsubseteq)$ for some posets $(A, \sqsubseteq)$ and $(B, \preceq)$). Note that the rolling rule subsumes the rule that a least prefix point is also a fixed point. Just instantiate f to the identity function. The proof we gave of this fact in the previous section can be adapted to give a proof of the rolling rule.

An immediate corollary of the rolling role is that, for a monotonic endofunction f, $\mu(f^2) = f.\mu(f^2)$ where f^2 denotes $f \bullet f$. This suggests that $\mu(f^2)$ and μf are equal. This is indeed the case, and is called the *square rule*:

(28) $\quad \mu f = \mu(f^2) \quad .$

Consider now a monotonic, binary function $\oplus$ on the poset $\mathcal{A}$. Then, for each x, the function $(x \oplus)$ (which maps y to $x \oplus y$) is monotonic and so has a least fixed point $\mu(x \oplus)$. Abstracting from x, the function $\langle x :: \mu(x \oplus)\rangle$ is a monotonic endofunction on $\mathcal{A}$ and its least fixed point is $\mu\langle x :: \mu\langle y :: x \oplus y\rangle\rangle$. The *diagonal rule* removes the nesting in this expression:

(29) $\quad \mu\langle x :: x \oplus x\rangle = \mu\langle x :: \mu\langle y :: x \oplus y\rangle\rangle \quad .$

The diagonal rule is a very important rule because it is the basis of methods for finding fixed points step by step. Suppose one wants to solve a fixed point equation $x :: x = x \oplus x$ where "$x \oplus x$" is a large expression involving several occurrences of the identifier x. The

diagonal rule allows one to eliminate the x's one-by-one. This is what happens when one solves a number of equations defining a set of values by mutual recursion (sometimes called a set of "simultaneous equations"). Such a set of equations can be viewed as one equation of the form $\underline{x} = f.\underline{x}$ where $\underline{x}$ is a vector of values. Each occurrence of an element x_i in the set of equations is in fact an occurrence of $\Pi_i.\underline{x}$, where Π_i denotes the function that "projects" the vector $\underline{x}$ onto its ith component. It is thus an occurrence of "$\underline{x}$" itself and the standard practice is to eliminate such occurrences individually. Mutual recursion is discussed further in section 4.4.

Exercise 30 Verify each of the rules given in this subsection.

□

Exercise 31 *Prove that in a Kleene algebra*

$$\mu\langle X :: 1 + a{\cdot}X{\cdot}b{\cdot}X + b{\cdot}X{\cdot}a{\cdot}X\rangle \;=\; \mu\langle X :: 1 + a{\cdot}X{\cdot}b + b{\cdot}X{\cdot}a + X{\cdot}X\rangle \quad .$$

□

4.2 Fusion

The most powerful of the two rules characterising least prefix points is the induction rule. Its power is, however, somewhat limited because it only allows one to calculate with orderings in which the μ operator is the *principal* operator on the *lower* side of the ordering (i.e. orderings of the form $\mu f \sqsubseteq \cdots$). A frequently used rule, which overcomes this restriction on the induction rule, can be obtained by combining the calculation properties of Galois connections with those of fixed points.

Theorem 32 (μ-fusion) Suppose $f \in A \leftarrow B$ is the lower adjoint in a Galois connection between the posets $(A, \sqsubseteq)$ and $(B, \preceq)$. Suppose also that $g \in (B, \preceq) \leftarrow (B, \preceq)$ and $h \in (A, \sqsubseteq) \leftarrow (A, \sqsubseteq)$ are monotonic functions.

(a) $\quad f.\mu g \sqsubseteq \mu h \;\Leftarrow\; f{\bullet}g \mathrel{\dot{\sqsubseteq}} h{\bullet}f \quad ,$

(b) $\quad f.\mu g = \mu h \;\Leftarrow\; f{\bullet}g = h{\bullet}f \quad .$

Indeed, if the condition

$$f{\bullet}g = h{\bullet}f$$

holds, f is the lower adjoint in a Galois connection between the posets $(\mathsf{Pre}.h, \sqsubseteq)$ and $(\mathsf{Pre}.g, \preceq)$.

□

We call this theorem μ-"fusion" because it states when application of function, f, can be "fused" with a fixed point, μg, to form a fixed point, μh. (The rule is also used, of course, to "defuse" a fixed point into the application of a function to another fixed point.) The fusion rule is the basis of so-called "loop fusion" techniques in programming: the combination of two loops, one executed after the other, into a single loop. More generally (as we shall see) the rule is the basis of techniques to combine two recursively defined computations into one.

Note that in order to apply the μ-fusion theorem we do not need to *know* the upper adjoint of function f, we only need to know that it exists. The fundamental theorem of Galois connections is the most commonly used tool to establish the existence of an upper adjoint.

As an aid to memorising the μ-fusion theorems note that the order of f, g, $\sqsubseteq$ or $=$, and h is the same in the consequent and the antecedent, be it that in the antecedent the lower adjoint occurs twice, in different argument positions of $\bullet$.

The conclusion of μ-fusion — $\mu h = f.\mu g$ — involves *two* premises, that f be a lower adjoint, and that $f \bullet g = h \bullet f$. The rule is nevertheless very versatile since being a lower adjoint is far from being uncommon, and many algebraic properties take the form $f \bullet g = h \bullet f$ for some functions f, g and h. The next lemma also combines fixed points with Galois connections. We call it the *exchange rule* because it is used to exchange one lower adjoint for another. Its proof is left as an exercise. (Hint: for part (a) use induction followed by the rolling rule.)

Lemma 33 (Exchange Rule) Suppose f has type $(A, \sqsubseteq) \leftarrow (B, \preceq)$, and g and h both have type $(B, \preceq) \leftarrow (A, \sqsubseteq)$. Then, if g is a lower adjoint in a Galois connection,

(a) $\quad \mu(g \bullet f) \sqsubseteq \mu(h \bullet f) \wedge \mu(f \bullet g) \sqsubseteq \mu(f \bullet h) \Leftarrow g \bullet f \bullet h \mathrel{\dot{\sqsubseteq}} h \bullet f \bullet g$.

Furthermore, if both g and h are lower adjoints, then

(b) $\quad \mu(g \bullet f) = \mu(h \bullet f) \wedge \mu(f \bullet g) = \mu(f \bullet h) \Leftarrow g \bullet f \bullet h = h \bullet f \bullet g$.

□

The two conclusions of (b), $\mu(g \bullet f) = \mu(h \bullet f)$ and $\mu(f \bullet g) = \mu(f \bullet h)$, say that g and h can be exchanged within a μ term in which they are composed after or before an arbitrary function f.

Exercise 34 *Prove the exchange rule.*

The two conjuncts in the consequent of 33(b) can be combined into one equation. The resulting lemma states when lower adjoint g can be replaced by lower adjoint h in an arbitrary context. The lemma has four variables rather than three. *Find the lemma and prove its validity.*

□

Uniqueness An important issue when confronted with a fixed point equation is whether or not the equation has a *unique* solution.

Uniqueness is a combination of their being at least one solution and at most one solution. In a complete lattice every monotonic function has a least and a greatest fixed point, so the question is whether the least and greatest fixed points are equal.

A very important special case is when the equation

$$x :: x = a + b \cdot x$$

has a unique solution in a Kleene algebra.

In order to give a general answer to this question, it is necessary to make a further assumption about the partial ordering relation on elements in the algebra. The assumption we make is that the elements in the algebra form a complete lattice under the partial ordering relation and, furthermore, this ordering is such that, for each a, the section $(a+)$ is the upper adoint in a Galois connection[2]. This important property is one that distinguishes a *regular algebra* from a Kleene algebra. The other additional property of a regular algebra (which we do not need to consider until later) is that each section $(b\cdot)$ is a lower adjoint in a Galois connection.

An ordering relation that is complete and is such that sections of the binary supremum operator are upper adjoints in a Galois connection is called a *complete, completely distributive lattice*. (It is called completely distributive because being an upper adjoint is equivalent to being universally distributive over all infima.) A *regular algebra* is thus a Kleene algebra for which underlying poset is a complete, completely distributive lattice and such that each section $(b\cdot)$ is a lower adjoint in a Galois connection.

The assumption of being completely distributive clearly holds for languages and relations, since in both cases the ordering relation is the subset relation and addition is set union and we have the shunting rule:

$$b \subseteq a \cup c \;\equiv\; \neg a \cap b \subseteq c \quad .$$

The assumption that $(a+)$ is the upper adoint in a Galois connection allows us to use fusion as follows:

Theorem 35 If $(y+)$ is an upper adjoint, then we have, for all a and b,

$$\nu\langle x :: a + x \cdot b\rangle = y + \nu\langle x :: x \cdot b\rangle \;\Leftarrow\; y = a + y \cdot b \quad .$$

Proof

[2]See the appendix for a definition of a Galois connection. Galois connections are discussed in detail elsewhere in this summer school.

$$
\begin{array}{ll}
 & \nu\langle x::\ a+x\cdot b\rangle = y+\nu\langle x::x\cdot b\rangle \\
\Leftarrow & \qquad \{ \quad (y+) \text{ is upper adjoint: } \nu\text{-fusion} \quad \} \\
 & \forall\langle x::\ a+(y+x)\cdot b = y+x\cdot b\rangle \\
\Leftarrow & \qquad \{ \quad (\cdot b) \text{ distributes over } +, \text{ associativity of } + \quad \} \\
 & a+y\cdot b = y \quad .
\end{array}
$$

□

As a consequence, in a regular algebra, the *largest* solution of the equation $x::\ x = a+x\cdot b$ is the sum (i.e. supremum) of an arbitrary solution and the largest solution of the equation $x::\ x = x\cdot b$. Note that a special choice for y in theorem 35 is $y = a\cdot b^*$.

An immediate corollary of theorem 35 is that if $\nu\langle x::x\cdot b\rangle = 0$, function $\langle x::\ a+x\cdot b\rangle$ has a unique fixed point. This is the rule we call the *unique extension property* (UEP) of regular algebra.

Theorem 36 (The unique extension property (UEP)) Suppose b is an element of a regular algebra. If $\nu\langle x::x\cdot b\rangle = 0$, then, for all a and x,

$$a+x\cdot b = x \quad \equiv \quad x = a\cdot b^* \quad .$$

□

The UEP draws attention to the importance of property $\nu\langle x::x\cdot b\rangle = 0$. In language theory it is equivalent to $\varepsilon\notin b$ since if, on the contrary, x is a non-empty set such that $x = x\cdot b$ then the length of the shortest word in x must be equal to the length of the shortest word in b plus the length of the shortest word in x. That is, the length of the shortest word in b is zero. The terminology that is often used is "b does not possess the *empty-word property*". In relation algebra we say "b is well-founded": the property expresses that there are no infinite sequences of b-related elements (thus, if relation b represents a finite directed graph, $\nu\langle x::x\cdot b\rangle = 0$ means that the graph is acyclic). This is discussed in detail in section 5.

Exercise 37 Consider the final three instances of a Kleene algebra shown in table 3.1. In each of these instances, suppose b is a square matrix with entries drawn from the carrier set of the relevant instance. (Equivalently, suppose b is a graph with edge labels drawn from the carrier set. If the (i,j)th matrix entry is 0 then there is no edge from i to j in the graph.) *What is the interpretation of* $\nu\langle x::x\cdot b\rangle = 0$ *in each case?*

□

4.3 Parameterised Prefix Points

Often fixed point equations are parameterised, sometimes explicitly, sometimes implicitly. An example of implicit parameters is the definition of the star operator discussed in section 3. In the defining equations given there (see for example equation (16)) a and b are parameters; the equations do not depend on how a and b are instantiated. The equations define the *function* mapping a to a^*. The rules we present in this section are concerned with such parameterised fixed point equations.

The first rule we present, the abstraction rule, is rarely given explicitly although it is very fundamental. The second rule, called the "beautiful theorem" because of its power and generality, is well-known in one particular form —its application to proving continuity properties of functions defined by fixed point equations— but is less well known at the level of generality given here.

The Abstraction Rule The definition of a^* in section 3 is a very good example of the "abstraction rule". Recall that a^* is defined to be the least prefix point of the function mapping x to $1 + a \cdot x$. Thus a is a *parameter* in the definition of the star operator. The remarkable, and very fundamental, property is that the star operator is itself a least prefix point. Indeed, suppose we consider the function

$$\langle g :: \langle a :: 1 + a \cdot g.a \rangle \rangle \quad .$$

(Note that this function maps an endofunction g on the Kleene algebra to an endofunction on the Kleene algebra and so is a (higher order) endofunction.) Then the star operator is the least fixed point of this function. That is,

$$\langle a :: a^* \rangle \;=\; \mu \langle g :: \langle a :: 1 + a \cdot g.a \rangle \rangle \quad .$$

We leave the verification of this fact as an exercise. (We prove a more general theorem shortly.)

The general theorem that this illustrates is obtained in the following way. We begin by making explicit the fact that a is a parameter in the definition of the star operator. This we do by considering the binary operator $\oplus$ defined by

$$a \oplus x = 1 + a \cdot x \quad .$$

Then, by definition, $a^* = \mu \langle x :: 1 + a \cdot x \rangle$. So

$$\langle a :: a^* \rangle \;=\; \langle a :: \mu \langle x :: a \oplus x \rangle \rangle \quad .$$

This is just a repetition of the definition in section 3 but slightly more roundabout. But now we consider the function F from functions to functions defined by

$$F \;=\; \langle g :: \langle a :: a \oplus g.a \rangle \rangle \quad .$$

(That is, $(F.g).a \;=\; a \oplus g.a$.) So for the particular definition of $\oplus$ given above,

$$F \;=\; \langle g:: \langle a:: \; 1 + a \cdot g.a \rangle \rangle \quad .$$

Then the theorem is that

$$\langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle \;=\; \mu F$$

for all (monotonic) binary operators $\oplus$ and all F defined as above. That is, for all monotonic binary operators $\oplus$,

(38) $$\langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle \;=\; \mu\langle g:: \langle a:: \; a \oplus g.a \rangle \rangle \quad .$$

This rule we call the *abstraction rule*.

In order to verify the validity of the rule we have to verify that $\langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle$ is a prefix point of the function F, and that it is at most any prefix point of F. This turns out to be straightforward. First, it is a prefix point by the following argument:

$$\begin{array}{ll} & F.\langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle \;\dot{\sqsubseteq}\; \langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle \\ \equiv & \quad \{ \quad \text{pointwise ordering, definition of } (F.g).a \quad \} \\ & \forall \langle a:: \; a \oplus \langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle.a \sqsubseteq \mu\langle x:: \; a \oplus x \rangle \rangle \\ \equiv & \quad \{ \quad \mu\langle x:: \; a \oplus x \rangle \text{ is a prefix point of } (\, a \oplus \,) \quad \} \\ & \text{true} \quad . \end{array}$$

Second, it is at most any prefix point. Suppose g is a prefix point of F, i.e. $F.g \;\dot{\sqsubseteq}\; g$. Then

$$\begin{array}{ll} & \langle a:: \; \mu\langle x:: \; a \oplus x \rangle \rangle \;\dot{\sqsubseteq}\; g \\ \equiv & \quad \{ \quad \text{pointwise ordering} \quad \} \\ & \forall \langle a:: \; \mu\langle x:: \; a \oplus x \rangle \sqsubseteq g.a \rangle \\ \Leftarrow & \quad \{ \quad \text{fixed point induction: (7)} \quad \} \\ & \forall \langle a:: \; a \oplus g.a \sqsubseteq g.a \rangle \\ \equiv & \quad \{ \quad \text{pointwise ordering, definition of } (F.g).a \quad \} \\ & F.g \;\dot{\sqsubseteq}\; g \\ \equiv & \quad \{ \quad g \text{ is a prefix point of } F \quad \} \\ & \text{true} \quad . \end{array}$$

This concludes the proof.

Exercise 39 Suppose the operator $\oplus$ is defined by $x \oplus y = f.y$. *What property does one then obtain by instantiating the abstraction rule?*

□

The Beautiful Theorem A very important application of the abstraction rule in combination with fixed point fusion is a theorem that has been dubbed "beautiful" by Dijkstra and Scholten [DS90, p. 159].

As observed precisely above, a parameterised least-fixed-point equation defines a function in terms of a binary operator. The beautiful theorem is that this function enjoys any kind of supremum-preserving property enjoyed by the binary operator. Thus supremum-preserving properties are preserved in the process of constructing least fixed points.

Dijkstra and Scholten formulate the "beautiful theorem" in the context of the predicate calculus. In terms of poset theory the theorem is stated as follows. Suppose that $\mathcal{A}=(A,\sqsubseteq)$ and $\mathcal{B}=(B,\preceq)$ are ordered sets and suppose $\oplus\in(\mathcal{A}\leftarrow\mathcal{A})\leftarrow\mathcal{B}$. (Thus $\oplus$ is a monotonic function mapping elements of the set B into monotonic endofunctions on the set A.) As in section 4.3 denote application of $\oplus$ to x by $x\oplus$. Assume $\mu(x\oplus)$ exists for each x and consider the function $\langle x::\mu(x\oplus)\rangle$, which we denote by $\dagger$. The "beautiful theorem" is that $\dagger$ enjoys any type of supremum-preserving property that is enjoyed by the (uncurried binary) function $\oplus$. More precisely,

$$\forall\langle f::\sqcup(\dagger\bullet f)=\dagger(\sqcup f)\rangle \Leftarrow \forall\langle f,g::\sqcup(f\dot{\oplus}g)=(\sqcup f)\oplus(\sqcup g)\rangle \quad ,$$

where, by definition, $f\dot{\oplus}g$ is the function mapping x to $f.x\oplus g.x$, $\bullet$ denotes function composition and $\sqcup$ denotes the supremum operator on functions of the shape of f and g.

The supremum operator, if it exists for a class of functions of a certain shape, is the lower adjoint in a Galois connection. This is the key to the proof of the beautiful theorem. In order to make this explicit we formulate the theorem yet more abstractly. The theorem is that for arbitrary function G that is the lower adjoint in a Galois connection, and arbitrary function H (G and H having of course appropriate types),

$$\forall\langle f::\ G.(\dagger\bullet f)=\dagger(H.f) \Leftarrow \forall\langle g::\ G.(f\dot{\oplus}g)=H.f\oplus G.g\rangle \quad .$$

The proof is very short. We first note that $(\dagger\bullet f).x=\dagger(f.x)=\mu((f.x)\oplus)$, by definition of $\dagger$. Thus, by the abstraction rule, $\dagger\bullet f=\mu(f\dot{\oplus})$. This is just what we need in order to apply the fusion theorem.

$$\begin{array}{ll} & G.(\dagger\bullet f)=\dagger.(H.f)\\ \equiv & \quad\{\quad \text{above}\quad\}\\ & G.\mu(f\dot{\oplus})=\mu((H.f)\oplus)\\ \Leftarrow & \quad\{\quad \text{by assumption, } G \text{ is a lower adjoint.}\\ & \qquad\quad \text{fusion: theorem 32}\quad\}\\ & \forall\langle g::\ G.(f\dot{\oplus}g)=H.f\oplus G.g\rangle \quad . \end{array}$$

Cocontinuity of Iteration We conclude the discussion of parameterised fixed points by considering the implications of the beautiful theorem for the iteration operator. Specifically, we show that in a regular algebra the iteration operator is *cocontinuous*[3]. Specifically, this is the property that for an ascending sequence $f.0 \leq f.1 \leq \ldots \leq f.n \leq \ldots$

$$(\Sigma\langle n{:}\ 0 \leq n{:}\ f.n\rangle)^* \ =\ \Sigma\langle n{:}\ 0 \leq n{:}\ (f.n)^*\rangle \quad .$$

A property of a regular algebra that we need to complete the proof is that, for all such sequences f and for all z,

$$z \cdot \Sigma\langle n{:}\ 0 \leq n{:}\ f.n\rangle \ =\ \Sigma\langle n{:}\ 0 \leq n{:}\ z \cdot f.n\rangle$$

and $$\Sigma\langle n{:}\ 0 \leq n{:}\ f.n\rangle \cdot z \ =\ \Sigma\langle n{:}\ 0 \leq n{:}\ f.n \cdot z\rangle$$

The property of iteration is thus that in a Kleene algebra in which the sections $(\cdot z)$ and $(z\cdot)$ are cocontinuous, for all z, the iteration operator * is also cocontinuous. (This proviso is satisfied in a regular algebra and thus in all Kleene algebras of interest to us.)

Recalling the discussion in the introduction to this section, the instantiation of the beautiful theorem we wish to consider is as follows: the function $\dagger$ is the operator * (i.e. the function $\langle a{::}\ a^*\rangle$; correspondingly the binary operator $\oplus$ is defined by $a \oplus x = 1 + a \cdot x$. The functions G and H are both taken to be the supremum operator of ascending functions on the natural numbers to the carrier of a regular algebra, which we denote by Σ.

The instantiation of the beautiful theorem (with $G,H,\dagger := \Sigma,\Sigma,^*$) gives us, for all ascending sequences f:

$$\Sigma\langle n{::}\ (f.n)^*\rangle = (\Sigma f)^* \quad \Leftarrow \quad \forall\langle f, g{::}\ \Sigma\langle n{::}\ \ 1 + f.n \cdot g.n\rangle \ =\ 1 + (\Sigma f)\cdot(\Sigma g)\rangle$$

where the dummies n range over the natural numbers, and the dummies f and g range over ascending sequences. We try to simplify the antecedent of this theorem. First, we have:

$$\begin{array}{ll} & \Sigma\langle n{::}\ 1 + f.n \cdot g.n\rangle \leq 1 + (\Sigma f)\cdot(\Sigma g) \\ \equiv & \qquad \{ \quad \text{definition of supremum} \quad \} \\ & \forall\langle n{::}\ 1 + f.n \cdot g.n \ \leq\ 1 + (\Sigma f)\cdot(\Sigma g)\rangle \\ \Leftarrow & \qquad \{ \quad + \text{ is the binary supremum operator} \quad \} \\ & \forall\langle n{::}\ f.n \cdot g.n \leq (\Sigma f)\cdot(\Sigma g)\rangle \end{array}$$

[3]There is a disturbing confusion between the standard terminology in the computing science literature and the literature on category theory in this respect. What is called a "continuity" property in the computing science literature is called a "*co*continuity" property in the category theory literature. We use the category theory terminology here.

$\equiv$ { property of suprema }

true .

So it is the opposite inclusion that is crucial. Now,

$(\Sigma f)\cdot(\Sigma g)$

$=$ { • assume, for all z, $(\cdot z)$ is cocontinuous }

$\Sigma\langle n::\ f.n\cdot(\Sigma g)\rangle$

$=$ { • assume, for all z, $(z\cdot)$ is cocontinuous }

$\Sigma\langle n::\Sigma\langle m::\ f.n\cdot g.m\rangle\rangle$.

So,

$\Sigma\langle n::\ 1+f.n\cdot g.n\rangle \geq 1+(\Sigma f)\cdot(\Sigma g)$

$\equiv$ { $+$ is the binary supremum operator }

$\Sigma\langle n::\ 1+f.n\cdot g.n\rangle \geq 1 \ \wedge\ \Sigma\langle n::\ 1+f.n\cdot g.n\rangle \geq (\Sigma f)\cdot(\Sigma g)$

$\Leftarrow$ { $\Sigma\langle n::1+c\rangle \geq 1$, for all c, and $1+f.n\cdot g.n \geq f.n\cdot g.n$ }

$\Sigma\langle n::\ f.n\cdot g.n\rangle \geq (\Sigma f)\cdot(\Sigma g)$

$\equiv$ { above }

$\Sigma\langle n::\ f.n\cdot g.n\rangle \geq \Sigma\langle n::\Sigma\langle m::\ f.n\cdot g.m\rangle\rangle$

$\equiv$ { suprema, dummy change }

$\forall\langle n,m::\Sigma\langle p::\ f.p\cdot g.p\rangle \geq f.n\cdot g.m\rangle$

$\Leftarrow$ { $\Sigma\langle p::\ f.p\cdot g.p\rangle \geq f.(n\uparrow m)\cdot g.(n\uparrow m)$
where $(n\uparrow m)$ denotes the maximum of m and n }

$\forall\langle n,m::\ f.(n\uparrow m)\cdot g.(n\uparrow m) \geq f.n\cdot g.m\rangle$

$\equiv$ { f and g are ascending sequences, product is monotonic }

true .

This completes the proof.

4.4 Mutual Recursion

In all but the simplest cases, recursive functions are defined by *mutual recursion*. For example, the BNF definition of the syntax of a programming language uses mutual recursion to simultaneously define the syntax of expressions, statements, declarations

etc. A definition by mutual recursion can be seen as a *single* fixed point equation where there is one unknown —a vector of values— and the function defining the unknown maps a vector to a vector. On the other hand, one also wishes to view a definition by mutual recursion as a collection of individual equations, defining several unknown values, and which may be solved on an individual basis. In this section we show that these two views are compatible with each other. We present one theorem to the effect that least prefix points can be calculated by solving individual equations and back-substituting.

Without loss of generality we can confine ourselves to fixed point equations involving two unknowns: a fixed point equation with $n+1$ unknowns can always be viewed as an equation on two unknowns, one of which is a vector of length n.

Any fixed point of a function mapping pairs of values to pairs of values is, of course, a pair. We show how to calculate the individual components of the least fixed point of such a function, given that it is possible to calculate least fixed points of functions mapping single elements to single elements. The fixed-point equation to be dealt with is the following.

(40) $\quad x, y :: \; x = x \odot y \;\wedge\; y = x \otimes y \quad .$

We first consider two special cases in which $\odot$ and $\otimes$ depend on one of their arguments only.

Lemma 41

(a) $\quad (\; \mu f \, , \mu g \;) = \mu \langle x, y :: (f.x, g.y) \rangle$

(b) $\quad (\; \mu(f \bullet g) \, , \mu(g \bullet f) \;) = \mu \langle x, y :: (f.y, g.x) \rangle \quad .$

Proof of (a). With Π_1 denoting projection on the first component, we have to prove

(42) $\quad \Pi_1 \, . \, \mu \langle x, y :: (f.x, g.y) \rangle \;=\; \mu f \quad .$

Since Π_1 distributes over all suprema, this is an occasion for applying μ-fusion: doing so, we see that (42) follows from

$$\Pi_1 . (f.x \, , g.y) = f . (\Pi_1 . (x, y)) \quad \text{for all } x \text{ and } y,$$

which is true.

The second component is dealt with similarly.

Proof of (b)

$$\begin{array}{ll} & (\; \mu(f \bullet g) \, , \mu(g \bullet f) \;) = \mu \langle x, y :: (f.y, g.x) \rangle \\ \Leftarrow & \qquad \{ \quad \text{(a) on lhs} \quad \} \\ & \mu \langle x, y :: (f.g.x \, , g.f.y) \rangle = \mu \langle x, y :: (f.y, g.x) \rangle \end{array}$$

$\equiv$ { define ϕ : $\phi(x,y) = (f.y\,,\,g.x)$ }

$\mu(\phi \bullet \phi) = \mu\phi$

$\equiv$ { square rule }

true .

□

With the aid of lemma (41), we now compute the least solution of (40), viz. we prove

Theorem 43

$$(\ \mu\langle x{::}\ x \odot p.x\rangle\ ,\ \mu\langle y{::}\ q.y \otimes y\rangle\ ) = \mu\langle x, y{::} (x \odot y, x \otimes y)\rangle$$

where $p.x = \mu\langle v{::} x \otimes v\rangle$ and $q.y = \mu\langle u{::} u \odot y\rangle$, i.e. $p.x$ and $q.y$ are the least fixed points of the individual equations.

Proof

$\mu\langle x, y{::}\ (x \odot y, x \otimes y)\rangle$

$=$ { diagonal rule (29) }

$\mu\langle x, y{::}\ \mu\langle u, v{::} (u \odot y, x \otimes v)\rangle\rangle$

$=$ { lemma (41a) and definition of p and q }

$\mu\langle x, y{::}\ (q.y, p.x)\rangle$

$=$ { lemma (41b) }

$(\ \mu\langle x{::}\ q.(p.x)\rangle\ ,\ \mu\langle y{::}\ p.(q.y)\rangle\ )$

$=$ { definition q , p }

$(\ \mu\langle x{::}\ \mu\langle u{::}\ u \odot p.x\rangle\rangle\ ,\ \mu\langle y{::}\ \mu\langle v{::}\ q.y \otimes v\rangle\rangle\ )$

$=$ { diagonal rule (29) twice: $u := x$, $v := y$ }

$(\ \mu\langle x{::}\ x \odot p.x\rangle\ ,\ \mu\langle y{::}\ q.y \otimes y\rangle\ )$.

□

This concludes the presentation of the calculus. The corresponding calculus of greatest fixed points is obtained by the interchanges $\mu \leftrightarrow \nu$, $\sqsubseteq \leftrightarrow \sqsupseteq$, and lower adjoint $\leftrightarrow$ upper adjoint.

5 Imperative Programming and Well-founded Relations

An important application of fixed point calculus is in the study of well-founded relations. Well-founded relations are fundamental to program termination and to inductive proofs of program properties.

Before we can begin to discuss how to express well-foundedness in terms of fixed points we need to introduce relation algebra. In relation algebra, relations are not viewed as sets of pairs but just as values in (the carrier of) an algebraic system. In this way, relation algebra expresses the properties of the most fundamental operations on relations without reference to the elements being related. For example, relation algebra encapsulates the properties of the converse $R^\cup$ of relation R without reference to the so-called *pointwise* definition of $R^\cup$:

$$(x,y) \in R^\cup \equiv (y,x) \in R \quad .$$

Instead the converse operation is given an axiomatic definition, which includes for example the property that it is its own inverse:

$$(R^\cup)^\cup = R \quad .$$

This *point-free* axiomatisation is the key to formulating the notions of well-foundedness and admitting induction in terms of fixed points.

A continuous motivational thread throughout this section will be the formulation of a methodology for the design of iterative programs in terms of relation algebra. We begin in section 5.1 with a brief introduction to relation algebra, just enough to be able to present a concise formulation of the use of invariant properties in section 5.2. This latter discussion raises several issues which motivates the introduction of the domain and division operators in section 5.3.

In section 5.4 we begin with the standard pointwise definition of well-foundedness which we then reformulate in a point-free definition. We then go on to derive an equivalent but more compact definition. The same process is followed for the notion of admitting induction. We recap the standard pointwise definition, reformulate this in a point-free manner and then derive a more compact but equivalent definition. We then discuss the equivalence of admitting induction and being well-founded.

5.1 Relation Algebra

For us, a (non-deterministic) program is an input-output relation. The convention we use when defining relations is that the input is on the right and the output on the left. The convention is thus that used in functional programming and not that used in

sequential programming. For example, the relation $<$ on numbers is a program that maps a number into one smaller than itself. The function succ is the relation between natural numbers such that $m\langle\mathrm{succ}\rangle n$ equivales $m = n+1$. It is thus the program that maps a natural number into its successor.

A relation is a set of ordered pairs. In discussions of the theory of imperative programming the "state space" from which the elements of each pair are drawn often remains anonymous. This reflects the fact that type structure is often not a significant parameter in the construction of imperative programs, in contrast to functional programs where it is pervasive. One goal here is to combine the functional and imperative programming paradigms. For this reason, we adopt a typed algebra of relations (formally an "allegory" [Fv90]). A relation is thus a triple consisting of a pair of types I and J, say, and a subset of the cartesian product $I \times J$. We write $R \in I \leftarrow J$ (read R has type I from J), the left-pointing arrow indicating that we view J as the set of all possible inputs and I as the set of possible outputs. I is called the *target* and J the *source* of the relation R, and $I \leftarrow J$ (read I from J) is called its *type*.

We write $x[\![R]\!]y$ if the pair (x, y) is an element of relation R. (As is usual in mathematics, we omit the brackets when R is denoted by a symbol and it is easy to parse the resulting expression — as in, for example, $x < y$.) We use a raised infix dot to denote relational composition. Thus $R \circ S$ denotes the composition of relations R and S (the relation defined by $x[\![R \circ S]\!]z$ equivales $\exists\langle y :: x[\![R]\!]y \wedge y[\![S]\!]z\rangle$). The composition $R \circ S$ is only defined when the source of R equals the target of S. Moreover, the target of $R \circ S$ is the target of R, and the source of $R \circ S$ is the source of S. Thus, $R \circ S \in I \leftarrow K$ if $R \in I \leftarrow J$ and $S \in J \leftarrow K$. The converse of relation R is denoted by $R^\cup$. Thus, $x[\![R^\cup]\!]y$ equivales $y[\![R]\!]x$. The type rule is that $R^\cup \in I \leftarrow J$ equivales $R \in J \leftarrow I$.

Relations of the same type are ordered by set inclusion denoted in the conventional way by the infix $\subseteq$ operator. The relations of a given type $I \leftarrow J$ form a complete lattice under this ordering. The smallest relation of type $I \leftarrow J$ is the empty relation, denoted here by $\bot\!\!\!\bot_{I \leftarrow J}$, and the largest relation of type $I \leftarrow J$ is the universal relation, which we denote by $\top\!\!\!\top_{I \leftarrow J}$. (We use this notation for the empty and universal relations because the conventional notation $\top$ for the universal relation is easily confused with T, a sans serif letter T, particularly in hand-written documents.) The union and intersection of two relations R and S of the same type are denoted by $R \cup S$ and $R \cap S$, respectively.

Because relations are sets, we have the shunting rule

$$R \cap S \subseteq T \;\equiv\; S \subseteq \neg R \cup T$$

where $\neg R$ denotes the complement of relation R. The only use we make of this rule here is the fact that relations of type $I \leftarrow I$ form a completely distributive lattice and thus a regular algebra (with operators as shown in table 3.1). We use this fact when we exploit

the unique extension property of regular algebras (theorem 36) in the identification of fixed point equations that define a relation uniquely.

For each set I there is an identity relation on I which we denote by id_I. Thus $id_I \in I \leftarrow I$. Relations of type $I \leftarrow I$ contained in the identity relation of that type will be called *coreflexives*. (The terminology *partial identity relation* and *monotype* is also used.) By convention, we use R, S, T to denote arbitrary relations and A, B and C to denote coreflexives. A coreflexive A thus has the property that if $x[\![A]\!]y$ then $x = y$. Clearly, the coreflexives of type $I \leftarrow I$ are in one to one correspondence with the subsets of I; we shall exploit this correspondence by identifying subsets of a set I with the coreflexives of type $I \leftarrow I$. Specifically, by an abuse of notation, we write $x \in A$ for $x[\![A]\!]x$ (on condition that A is a coreflexive). We also identify coreflexives with predicates, particularly when discussing induction principles (which are traditionally formulated in terms of predicates rather than sets). So we shall say "x has property A" meaning formally that $x[\![A]\!]x$. Continuing this abuse of notation, we use $\sim A$ to denote the coreflexive having the same type as A and containing just those elements not in A. Thus, $x[\![\sim A]\!]y$ equivales the conjunction of $x \in I$ (where A has type $I \leftarrow I$) and $x = y$ and not $x[\![A]\!]x$. We also sometimes write I where id_I is meant. (This fits in with the convention in category theory of giving the same name to that part of a functor which maps objects to objects and that part which maps arrows to arrows.) A final, important remark about coreflexives is that their composition coincides with their intersection. That is, for coreflexives A and B, $A \circ B = A \cap B$.

We use an infix dot to denote function application. Thus $f.x$ denotes application of function f to argument x. Functions are particular sorts of relations; a relation R is *functional* if $y[\![R]\!]x$ and $z[\![R]\!]x$ together imply that $y = z$. If this is the case we write $R.x$ for the unique y such that $y[\![R]\!]x$. Note that functionality of relation R is equivalent to the property $R \circ R^{\cup} \subseteq id_I$ where I is the target of R. We normally use f, g and h to denote functional relations.

Dual to the notion of functionality of a relation is the notion of injectivity. A relation R with source J is *injective* if $R^{\cup} \circ R \subseteq id_J$. Which of the properties $R \circ R^{\cup} \subseteq id_I$ or $R^{\cup} \circ R \subseteq id_J$ one calls "functional" and which "injective" is a matter of interpretation. The choice here fits in with the convention that input is on the right and output on the left. More importantly, it fits with the convention of writing $f.x$ rather than say x^f (that is the function to the left of its argument). A sensible consequence is that type arrows point from right to left.

5.2 Imperative Programming

In this section we introduce the derivation of repetitive statements using invariant relations. The section contains just an outline of the methodology expressed in relation

algebra. For extensive introductions see (for example) [Gri81, Bac86]. At the end of the section we identify a need to delve deeper into relation algebra, thus motivating the section which follows.

Given a (non-trivial) specification, X, the key to constructing a loop implementing X is the invention of an invariant, Inv. The invariant is chosen in such a way that it satisfies three properties. First, the invariant can be "established" by some initialisation $Init$. Second, the combination of the invariant and some termination $Term$ satisfies the specification X. Third, the invariant is "maintained by" some loop body $Body$ whilst making progress towards termination.

These informal requirements can be made precise in a very concise way. The three components Inv, $Init$ and $Term$ are all binary relations on the state space, just like the specification X. They are so-called input-output relations.

"Establishing" the invariant is the requirement that

$$Init \subseteq Inv \quad .$$

In words, any value w' related to input value w by $Init$ is also related by the invariant relation to w.

That the combination of the termination and invariant satisfies the specification X is the requirement that

$$Term \circ Inv \subseteq X \quad .$$

This is the requirement that for all output values w' and input values w,

$$\forall \langle v\colon w' [\![Term]\!] v \wedge v [\![Inv]\!] w\colon w' [\![X]\!] w \rangle$$

(Here we see the convention of placing input values on the right and output values on the left.)

Finally, that the invariant is maintained by the loop body is expressed by

$$Body \circ Inv \subseteq Inv$$

Pointwise this is

$$\forall \langle w', v, w\colon \ w' [\![Body]\!] v \wedge v [\![Inv]\!] w\colon \ w' [\![Inv]\!] w \rangle \quad .$$

So $Body$ maps values v related by the invariant Inv to w to values w' that are also related by Inv to w.

Together these three properties guarantee that

$$Term \circ Body^* \circ Init \ \subseteq \ X \quad .$$

That progress is made is the requirement that the relation $Body$ be well-founded. (This we will return to shortly.)

As an example, consider the classic problem of finding the greatest common divisor (abbreviated gcd) of two positive numbers x and y. The state space of the program is $Int \times Int$. The specification, invariant, initialisation and termination are thus binary relations on this set. The specification, X, is simply

$$x' = y' = gcd.(x,y) \quad .$$

Here priming x and y is a commonly used convention for abbreviating the definition of a relation between the pair of output values x' and y', and the pair of input values x and y. More formally, X is the relation

$$\{x, y, x', y' : x' = y' = gcd.(x,y) : ((x',y'), (x,y))\} \quad .$$

The convention is that the definition

$$\{x, y, x', y' : p.(x,y,x',y') : ((x',y'), (x,y))\}$$

is abbreviated to

$$p.(x,y,x',y') \quad ,$$

the primes indicating the correspondence between input and output variables. Using this convention, the invariant is the relation

$$gcd.(x',y') = gcd.(x,y)$$

and the initialisation is the identity relation

$$x' = x \wedge y' = y \quad .$$

(The initialisation is thus implemented by $skip$, the do-nothing statement.) The termination is a subset of the identity relation on the state space. It is the relation

$$x' = x = y' = y \quad .$$

The composition of the termination relation and the invariant is thus the relation

$$x' = y' \wedge gcd.(x',y') = gcd.(x,y)$$

which, since $gcd.(x',x')$ equals x', is identical to the specification X. The loop body in Dijkstra's well-known guarded command solution to this problem is the union of two relations, the relation

$$x < y \wedge x' = x \wedge y' = y - x$$

and the relation

$$y < x \wedge y' = y \wedge x' = x - y \quad .$$

Exercise 44 Identify X, Inv, $Init$, $Body$ and $Term$ in the language recognition program discussed in section 2.1.

□

5.3 Domains and Division

5.3.1 Domains

Our account of invariants is not yet complete. The relationship between the specification X and $Term \circ Body^* \circ Init$ is containment not equality, and may indeed be a proper superset relation. Not every subset of the specification will do, however. An additional requirement is that the input-output relation computed by the program is total on all input values. Formally this is a requirement on the so-called "right domain" of the computed input-output relation. Right domains are also relevant if we are to relate our account of invariants to the implementation of loops by a **while** statement. Recall that $Body$ is the body of the loop, and $Term$ terminates the computation. The implementation of $Term \circ Body^*$ by a **while** statement demands that both of these relations are partial and, more specifically, that their right domains are complementary.

The *right domain* of a relation R is, informally, the set of input values that are related by R to at least one output value. Formally, the right domain $R^>$ of a relation R of type $I \leftarrow J$ is a coreflexive of type $J \leftarrow J$ satisfying the property that

$$\forall \langle A: A \subseteq id_J: R \circ A = R \equiv R^> \subseteq A \rangle \quad .$$

Given a coreflexive A, $A \subseteq id_J$, the relation $R \circ A$ can be viewed as the relation R restricted to inputs in the set A. Thus, in words, the right domain of R is the least coreflexive A that maintains R when R is restricted to inputs in the set A.

Note that the right domain should not be confused with the source of the relation. The source expresses the set of input values of interest in the context of the application being considered whereas the right domain is the set of input values over which the relation is defined. In other words, we admit the possibility of *partial* relations. Formally, a relation R of type $I \leftarrow J$ is *total* if $R^>$ is id_J, otherwise it is partial. Similarly the target should not be confused with the left domain of a relation. A relation R of type $I \leftarrow J$ is *surjective* if $R^<$ is id_I.

Returning to loops, the requirement is that the right domain of $Term$ is the complement of the right domain of $Body$. Letting b denote the right domain of $Body$ and $\sim b$ its complement (thus $b \cup \sim b = id$ and $b \cap \sim b = \bot\!\!\!\bot$) we thus have

$$Term = Term \circ \sim b \quad \text{and} \quad Body = Body \circ b \quad .$$

As a consequence,

$$\mathrm{Term} \circ \mathrm{Body}^* \circ \mathrm{Init} = \mathrm{Term} \circ \sim \mathrm{b} \circ (\mathrm{Body} \circ \mathrm{b})^* \circ \mathrm{Init} \quad .$$

The statement **while** b **do** Body is the implementation of $\sim \mathrm{b} \circ (\mathrm{Body} \circ \mathrm{b})^*$ in that the latter is the least solution of the equation

$$X :: X = \sim \mathrm{b} \cup X \circ \mathrm{Body} \circ \mathrm{b}$$

and executing this equation is equivalent to executing the program

$$X = \textbf{if } \mathrm{b} \textbf{ then } \mathrm{Body}; X \quad .$$

We continue this discussion in section 5.5.

5.3.2 Division

The body of a loop should maintain the loop invariant. Formally, the requirement is that $\mathrm{Body} \circ \mathrm{Inv} \subseteq \mathrm{Inv}$. In general, for relations R of type $\mathrm{I} \leftarrow \mathrm{J}$ and T of type $\mathrm{I} \leftarrow \mathrm{K}$ there is a relation $\mathrm{R} \backslash \mathrm{T}$ of type $\mathrm{J} \leftarrow \mathrm{K}$ satisfying the Galois connection, for all relations S,

$$\mathrm{R} \circ \mathrm{S} \subseteq \mathrm{T} \equiv \mathrm{S} \subseteq \mathrm{R} \backslash \mathrm{T} \quad .$$

The operator $\backslash$ is called a *division* operator (because of the similarity of the above rule to the rule of division in ordinary arithmetic). The relation $\mathrm{R} \backslash \mathrm{T}$ is called a *residual* or a *factor* of the relation T. Relation $\mathrm{R} \backslash \mathrm{T}$ holds between output value w' and input value w if and only if

$$\forall \langle v :\ v [\![\mathrm{R}]\!] w' :\ v [\![\mathrm{T}]\!] w \rangle \quad .$$

Applying this Galois connection, the requirement on Body is thus equivalent to

$$\mathrm{Inv} \subseteq \mathrm{Body} \backslash \mathrm{Inv} \quad ,$$

the pointwise formulation of which is

$$\forall \langle w', w :\ w' [\![\mathrm{Inv}]\!] w :\ \forall \langle w'' : w'' [\![\mathrm{Body}]\!] w' : w'' [\![\mathrm{Inv}]\!] w \rangle \rangle \quad .$$

The relation $\mathrm{Body} \backslash \mathrm{Inv}$ corresponds to what is called the *weakest prespecification* of Inv with respect to Body in the more usual predicate calculus formulations of the methodology [HH86]. The *weakest liberal precondition* operator will be denoted here by the symbol "$\backslash\!\!\cdot$". Formally, if R is a relation of type $\mathrm{I} \leftarrow \mathrm{J}$ and A is a coreflexive of type $\mathrm{I} \leftarrow \mathrm{I}$ then $\mathrm{R} \backslash\!\!\cdot \mathrm{A}$ is a coreflexive of type $\mathrm{J} \leftarrow \mathrm{J}$ characterised by the property that, for all coreflexives B of type $\mathrm{J} \leftarrow \mathrm{J}$,

$$(45) \qquad (\mathrm{R} \circ \mathrm{B})^< \subseteq \mathrm{A} \equiv \mathrm{B} \subseteq \mathrm{R} \backslash\!\!\cdot \mathrm{A} \quad .$$

(If we interpret the coreflexive A as a predicate p on the type I, then $R\backslash\!\!\cdot A$ is the predicate q such that

$$q.w \equiv \forall\langle w' : w'[\![R]\!]w : p.w\rangle \quad .$$

It is the weakest condition q on input values w that guarantees that all output values w' that are R-related to w satisfy the predicate p.)

The operator $\backslash\!\!\cdot$ plays a very significant role in what is to follow. For this reason it is useful to have a full and intimate understanding of its algebraic properties. This, however, is not the place to develop that understanding and we make do with a summary of the most frequently used properties.

First note that the function $(R\backslash\!\!\cdot)$, being an upper adjoint, distributes over arbitrary meets of coreflexives. Because meet on coreflexives coincides with composition it follows that $R\backslash\!\!\cdot$ distributes over composition: $R\backslash\!\!\cdot(A\circ B) = (R\backslash\!\!\cdot A)\circ(R\backslash\!\!\cdot B)$. This corresponds to the fact that weakest liberal precondition operator associated with a statement R is universally conjunctive. From (45) we obtain the *cancellation* property:

(46) $$(R\circ R\backslash\!\!\cdot B)^{<} \subseteq B \quad .$$

Often this property is used in a slightly different form, namely:

(47) $$R\circ R\backslash\!\!\cdot B \subseteq B\circ R \quad .$$

Both (46) and (47) express that program R produces a result from set B when started in a state satisfying $R\backslash\!\!\cdot B$. If R is a function then $R\backslash\!\!\cdot A$ can be expressed without recourse to the left-domain operator. Specifically, we have for function f:

(48) $$f\backslash\!\!\cdot A = f^{\cup}\circ A\circ f \quad .$$

A full discussion, including all the properties used here, can be found in [BW93].

5.4 Well-Foundedness Defined

Expressed in terms of points, a relation R is said to be *well-founded* if there are no infinite chains x_0, x_1, ... such that $x_{i+1}[\![R]\!]x_i$ for all i, $i \geq 0$. A relation R is thus *not* well-founded if there is a set A such that

$$A \neq \phi \wedge \forall\langle x : x\in A : \exists\langle y : y\in A : y[\![R]\!]x\rangle\rangle \quad .$$

Noting that $\exists\langle y : y\in A : y[\![R]\!]x\rangle \equiv x \in (A\circ R)^{>}$ this definition converts directly into the following point-free form.

Definition 49 (Well-founded) Relation R is said to be *well-founded* if and only if it satisfies

$$\forall\langle A:\ A \subseteq I:\ A \subseteq \bot\!\!\bot \Leftarrow A \subseteq (A \circ R)^{>}\rangle \quad .$$

□

The connection between well-foundedness and fixed points is the following.

Theorem 50 Relation R is well-founded equivales

$$\nu\langle A: A \subseteq I: (A \circ R)^{>}\rangle = \bot\!\!\bot \quad .$$

Proof For arbitrary monotonic function f we have:

$$\begin{array}{ll} & \nu f = \bot\!\!\bot \\ \Leftarrow & \quad \{ \quad \text{reflexivity of } \subseteq,\ \bot\!\!\bot \text{ is the least element} \quad \} \\ & \forall\langle X::\ X \subseteq \bot\!\!\bot \Leftarrow X \subseteq \nu f\rangle \\ \Leftarrow & \quad \{ \quad \text{fixed point induction} \quad \} \\ & \forall\langle X::\ X \subseteq \bot\!\!\bot \Leftarrow X \subseteq f.X\rangle \quad . \end{array}$$

The corollary follows by instantiating f to $\langle A: A \subseteq I: (A \circ R)^{>}\rangle$.
□

Characteristic of definition 49 is that it is a rule for establishing when a set represented by a coreflexive A, $A \subseteq I$, is empty. In the following theorem we replace sets by arbitrary relations. This has the advantage that we can then immediately exploit the unique extension property of a regular algebra (theorem 36).

Theorem 51 For arbitrary relation $R \in I \leftarrow I$,

$$(\nu\langle X:: X \circ R\rangle)^{>} \ = \ \nu\langle A: A \subseteq I: (A \circ R)^{>}\rangle \quad .$$

Hence R is well-founded if and only if it satisfies

$$\nu\langle X:: X \circ R\rangle = \bot\!\!\bot \quad .$$

(Here the dummy X ranges over all relations of type $I \leftarrow I$.)

Proof We shall only sketch the proof since we have not discussed the algebraic properties of the right domain operator in sufficient detail to give a completely formal proof. (For such a proof see [DBvdW97].)

At first sight it would seem that a simple application of the fusion theorem would suffice. This is not the case, however, because the right domain operator is a lower adjoint in a Galois connection, not an upper adjoint as is required to apply fusion.

The key to the proof is to observe that $\nu\langle X :: X \circ R\rangle$ is a so-called *right condition*. That is,

$$\nu\langle X :: X \circ R\rangle \;=\; \top\!\!\top \circ \nu\langle X :: X \circ R\rangle \quad .$$

(This is easily proved by a mutual inclusion argument.) This suggests the use of the fusion theorem to prove that

$$\iota \,.\, \nu\langle p :\, p = \top\!\!\top \circ p :\, p \circ R\rangle = \nu\langle X :: X \circ R\rangle$$

where ι denotes the function that embeds the set of right conditions of type $I \leftarrow I$ into the set of relations of type $I \leftarrow I$. (Embedding functions between complete lattices are both upper and lower adjoints so there is no difficulty in applying the fusion theorem.) The proof is now completed by using the fact that the two functions $\langle p :\, p = \top\!\!\top \circ p :\, p_{>}\rangle$ and $\langle A :\, A \subseteq I :\, \top\!\!\top \circ A\rangle$ are inverse lattice isomorphisms between the lattice of right conditions and the lattice of coreflexives of type $I \leftarrow I$. (This is the only part of the proof that the reader cannot reproduce for themself.) Thus, using the fusion theorem once again,

$$\nu\langle A :: (A \circ R)_{>}\rangle = (\nu\langle p :: p \circ R\rangle)_{>}$$

and

$$\top\!\!\top \circ \nu\langle A :: (A \circ R)_{>}\rangle = \nu\langle p :: p \circ R\rangle \quad .$$

From this it follows that

$$(\nu\langle X :: X \circ R\rangle)_{>} \;=\; \nu\langle A :\, A \subseteq I :\, (A \circ R)_{>}\rangle \quad .$$

Now,

$$\begin{array}{ll} & R \text{ is well-founded} \\ \equiv & \quad \{ \quad \text{definition} \quad \} \\ & \nu\langle A :\, A \subseteq I :\, (A \circ R)_{>}\rangle = \bot\!\!\bot \\ \equiv & \quad \{ \quad \text{above} \quad \} \\ & (\nu\langle X :: X \circ R\rangle)_{>} = \bot\!\!\bot \\ \equiv & \quad \{ \quad \text{domains} \quad \} \\ & \nu\langle X :: X \circ R\rangle = \bot\!\!\bot \quad . \end{array}$$

□

Corollary 52 Relation R is well-founded equivales

$$\forall\langle S, T::\ T = S \cup T \circ R \ \equiv\ T = S \circ R^*\rangle$$

Proof A relation algebra is a regular algebra. So theorem 36 holds with the product operator instantiated to relational composition and the addition operator instantiated to set union.
□

Having expressed well-foundedness in terms of fixed points it is now possible to apply fixed-point calculus to deduce some of its properties. The following is elementary (even without the use of fixed points!) but needs to be stated because of its frequent use.

Lemma 53 If relation R is well-founded and $S \subseteq R$ then S is well-founded.
□

Fixed point calculus gives an easy proof of the following more interesting theorem.

Theorem 54 For all R, that R is well-founded equivales that R^+ is well-founded.

Proof We prove the stronger theorem that

$$\nu\langle X:: X \circ R\rangle = \nu\langle X:: X \circ R^+\rangle \quad .$$

The inclusion $\nu\langle X:: X \circ R\rangle \subseteq \nu\langle X:: X \circ R^+\rangle$ is immediate from monotonicity of ν and the fact that $R \subseteq R^+$. For the other inclusion, we calculate:

$$\begin{array}{ll} & \nu\langle X:: X \circ R^+\rangle \subseteq \nu\langle X:: X \circ R\rangle \\ \Leftarrow & \quad \{ \quad \text{fixed point induction} \quad \} \\ & \nu\langle X:: X \circ R^+\rangle \subseteq \nu\langle X:: X \circ R^+\rangle \circ R \\ \equiv & \quad \{ \quad R^+ = R \circ R^* \quad \} \\ & \nu\langle X:: X \circ R^+\rangle \subseteq \nu\langle X:: X \circ R \circ R^*\rangle \circ R \\ \equiv & \quad \{ \quad \text{rolling rule} \quad \} \\ & \nu\langle X:: X \circ R^+\rangle \subseteq \nu\langle X:: X \circ R^* \circ R\rangle \\ \equiv & \quad \{ \quad R^+ = R^* \circ R \quad \} \\ & \text{true} \quad . \end{array}$$

□

This concludes this section. With dummies A and p ranging over coreflexives and right conditions, respectively, and X, S and T over relations, we have established the equivalence of the properties:

- R is well-founded.
- $\nu\langle A :: (A \circ R)^{>} \rangle = \bot\!\!\bot$.
- $\nu\langle p :: p \circ R \rangle = \bot\!\!\bot$.
- $\nu\langle X :: X \circ R \rangle = \bot\!\!\bot$.
- R^+ is well-founded.
- $\forall\langle S, T ::\ T = S \cup T \circ R \ \equiv\ T = S \circ R^* \rangle$.

Exercise 55 We saw above that $\nu\langle X :: X \circ R \rangle$ is a right condition. *What is the interpretation of this right condition as a set?*

□

Exercise 56 The standard technique for proving termination of a loop statement or a recursive definition in a program is to use a *bound function*. That is, one defines a function from the state space to a set on which a well-founded relation is defined. Most commonly the set is the set of natural numbers, and one proves that the body of the loop statement or recursive definition strictly reduces the value of the bound function.

Suppose the body of the loop statement is given by relation S, the bound function is f and the well-founded relation is R. Then the technique amounts to proving that if $x[\![S]\!]y$ then $f.x\,[\![R]\!]\,f.y$. That is, $S \subseteq f\cup \circ R \circ f$. In view of lemma 53, the validity of the use of bound functions is justified by the following theorem: If R is a relation and f a functional relation such that R is well-founded, then relation $f\cup \circ R \circ f$ is well-founded as well.

Prove this theorem.

Hint: as in the proof of theorem 54 one can make a more general statement relating $\nu\langle X :: X \circ R \rangle$ and $\nu\langle X ::\ X \circ f\cup \circ R \circ f \rangle$.

□

Exercise 57 *Show that if* R *is well-founded then* $R^+ \cap I \subseteq \bot\!\!\bot$. (So no non-empty subset of a well-founded relation is reflexive.)

Under what conditions is the well-foundedness of R *equivalent to* $R^+ \cap I \subseteq \bot\!\!\bot$ *? Provide examples where possible.*

□

5.5 Totality of while statements

We are now in a position to complete our discussion of the construction of **while** statements using loop invariants.

Recall that $Body$ is the body of the loop, and $Term$ terminates the computation. The well-foundedness of $Body$ guarantees that the execution of the **while** statement will always terminate. It also guarantees that the implementation is total, provided that $Term$ and $Body$ have complementary right domains, and the initialisation $Init$ is total. Specifically, we have:

$$
\begin{array}{ll}
 & (Term \circ Body^* \circ Init)^> = I \\
\equiv & \quad \{ \quad \text{domain calculus} \quad \} \\
 & ((Term \circ Body^*)^> \circ Init)^> = I \\
\Leftarrow & \quad \{ \quad \text{by assumption, } Init \text{ is total, i.e. } Init^> = I \quad \} \\
 & (Term \circ Body^*)^> = I \\
\equiv & \quad \{ \quad (Term \circ Body^*)^> \text{ is the unique solution of the equation} \\
 & \qquad A :: A = Term^> \cup (A \circ Body)^> \quad \} \\
 & I = Term^> \cup (I \circ Body)^> \\
\equiv & \quad \{ \quad \text{by assumption,} \\
 & \qquad Term \text{ and } Body \text{ have complementary right domains.} \\
 & \qquad \text{In particular, } I = Term^> \cup Body^> \quad \} \\
 & \text{true} \quad .
\end{array}
$$

The penultimate step needs further justification. The claim is that the equation

$$A :: A = Term^> \cup (A \circ Body)^>$$

has a unique solution provided that $Body$ is well-founded. This is easily derived from (52). Indeed, for all coreflexives A,

$$
\begin{array}{ll}
 & A = Term^> \cup (A \circ Body)^> \\
\equiv & \quad \{ \quad \text{domain calculus.} \\
 & \qquad \text{Specifically, } (\top\!\!\top \circ A)^> = A \text{ and } \top\!\!\top \circ R = \top\!\!\top \circ R^> \quad \} \\
 & \top\!\!\top \circ A = \top\!\!\top \circ Term \cup \top\!\!\top \circ A \circ Body \\
\equiv & \quad \{ \quad Body \text{ is well-founded, (52)} \quad \} \\
 & \top\!\!\top \circ A = \top\!\!\top \circ Term \circ Body^*
\end{array}
$$

$$\equiv \quad \{ \quad \text{domain calculus (as above)} \quad \}$$
$$A = (Term \circ Body^*)_{>} \quad .$$

That is, $(Term \circ Body^*)_{>}$ is the unique solution of the above equation in A.

5.6 Induction Principles

Dual to the notion of well-foundedness is the notion of admitting induction. This section formulates the latter notion in terms of fixed points and then shows that well-foundedness and admitting induction are equivalent.

A relation R is said to *admit induction* if the following schema can be used to establish that property P holds everywhere: prove, for all y, that the induction hypothesis $\forall\langle x: x[\![R]\!]y: P.x\rangle$ implies $P.y$. That is, expressed in terms of points, R *admits induction* iff

$$\forall\langle y:: P.y\rangle \Leftarrow \forall\langle y:: \forall\langle x: x[\![R]\!]y: P.x\rangle \Rightarrow P.y\rangle \quad .$$

The subterm

$$\forall\langle x: x[\![R]\!]y: P.x\rangle$$

in this formula is called the *induction hypothesis* while the proof of the subterm

$$\forall\langle y:: \forall\langle x: x[\![R]\!]y: P.x\rangle \Rightarrow P.y\rangle$$

the *induction step*. For instance, replacing R by the less-than relation $(<)$ on natural numbers and the dummies x and y by m and n, the less-than relation admits induction iff

$$\forall\langle n:: P.n\rangle \Leftarrow \forall\langle n:: \forall\langle m: m<n: P.m\rangle \Rightarrow P.n\rangle \quad .$$

This is indeed the case since the above is the statement of the principle of strong mathematical induction. The induction hypothesis is $\forall\langle m: m<n: P.m\rangle$; the induction step is the proof that assuming the truth of the induction hypothesis one can prove $P.n$. That the less-than relation admits induction means that from the proof of the induction step one can infer that $P.n$ holds for all n.

Less directly but nevertheless straightforwardly, replacing R by the predecessor relation on natural numbers, i.e. $x[\![R]\!]y \equiv y = x+1$, and simplifying using the fact that no number is a predecessor of 0, one obtains the principle of simple mathematical induction: with n ranging over the natural numbers,

$$\forall\langle n:: P.n\rangle \;\Leftarrow\; P.0 \wedge \forall\langle n:: P.n \Rightarrow P.(n+1)\rangle \quad .$$

Thus the predecessor relation on natural numbers admits induction. Note that in this case the proof of $P.0$ is called the *basis* of the proof by induction and the proof of $P.n \Rightarrow P.(n+1)$, for all n, the induction step.

The pointwise definition of "admits induction" given above is in terms of predicates. Because we want to arrive at a definition in terms of relations we first reformulate it in terms of sets. So we define: relation R *admits induction* if and only if:

(58) $\quad \forall\langle y :: y \in A\rangle \Leftarrow \forall\langle y :: \forall\langle x : x[\![R]\!]y : x \in A\rangle \Rightarrow y \in A\rangle$.

To arrive at a definition without dummies we first notice that $\forall\langle y :: y \in A\rangle$, the (understood) domain of y being I, can be rewritten as $I \subseteq A$. Furthermore, we see that the expression in the domain of the antecedent, $\forall\langle x : x[\![R]\!]y : x \in A\rangle$, is just $y \in R\backslash\!\!\cdot A$. So (58) can be drastically simplified to

(59) $\quad I \subseteq A \Leftarrow R\backslash\!\!\cdot A \subseteq A$,

for all coreflexives A of type $I \leftarrow I$.

According to the terminology introduced above, $R\backslash\!\!\cdot A$ is the induction hypothesis, whilst a proof of $R\backslash\!\!\cdot A \subseteq A$ is the induction step.

This then is the definition of "admits induction".

Definition 60 The relation R is said to *admit induction* if and only if it satisfies

$$\forall\langle A : A \subseteq I : I \subseteq A \Leftarrow R\backslash\!\!\cdot A \subseteq A\rangle \ .$$

Equivalently, R is said to admit induction if and only if

$$\mu\langle A : A \subseteq I : R\backslash\!\!\cdot A\rangle = \top\!\!\top \ .$$

□

Just as for well-foundedness, we propose a definition in which the type difference between the variables is removed.

Theorem 61 That relation R admits induction equivales $\mu\langle X :: R\backslash X\rangle = \top\!\!\top$.

□

We omit the proof as it is essentially dual to the proof of theorem 51.

Exercise 62 Definition 60 draws attention to the coreflexive $\mu(R\backslash\!\!\cdot)$: if relation R admits induction then the set corresponding to $\mu(R\backslash\!\!\cdot)$ is the universe over which relation R is defined. By restricting the domain of any relation R it is always possible to obtain a relation that admits induction. Specifically, for any relation R, the relation $R \circ \mu(R\backslash\!\!\cdot)$ admits induction. *Prove this theorem.*

□

5.7 Admits-induction Implies Well-Founded

Now that we have seen several equivalent definitions of well-founded it is time to explore its relationship to admitting induction. The following lemma is the key insight.

Lemma 63 $\nu\langle T :: T \circ R\rangle \circ \mu\langle T :: R \backslash T\rangle = \bot\!\!\!\bot$.

Proof The form of the theorem suggests that we try to apply μ-fusion. Of course we then have to find a suitable function f such that $\mu f = \bot\!\!\!\bot$. The identity function is one possibility and, as it turns out, is a good choice. However, since we want to demonstrate how good use of the calculational technique can avoid the need to make guesses of this nature, we construct f. We have, for all X,

$$
\begin{array}{ll}
 & X \circ \mu\langle T :: R \backslash T\rangle = \bot\!\!\!\bot \\
\equiv & \quad \{ \quad \text{introduce } f \text{ such that } \bot\!\!\!\bot = \mu f \quad \} \\
 & X \circ \mu\langle T :: R \backslash T\rangle \subseteq \mu f \\
\Leftarrow & \quad \{ \quad \text{basic fusion theorem} \quad \} \\
 & \forall\langle T :: X \circ R \backslash T \subseteq X \circ f.T\rangle \\
\equiv & \quad \{ \quad \text{choose for } f,\ f.T = T, \text{ noting that indeed } \bot\!\!\!\bot = \mu\langle T :: T\rangle. \\
 & \qquad \text{factor cancellation: specifically, } R \circ R \backslash T \subseteq T \quad \} \\
 & X \subseteq X \circ R \\
\Leftarrow & \quad \{ \quad \text{definition of } \nu\langle T :: T \circ R\rangle \quad \} \\
 & X = \nu\langle T :: T \circ R\rangle \quad .
\end{array}
$$

□

Theorem 64 If R admits induction then R is well-founded.

Proof If R admits induction then, by definition, $\mu\langle T :: R \backslash T\rangle = \top\!\!\!\top$. So, by lemma 63, $\nu\langle T :: T \circ R\rangle \circ \top\!\!\!\top = \bot\!\!\!\bot$. But then, since $I \subseteq \top\!\!\!\top$, $\nu\langle T :: T \circ R\rangle \subseteq \bot\!\!\!\bot$. By theorem 51 we have thus established that R is well-founded.
□

Exercise 65 One might suppose that an argument dual to the above leads to a proof that well-foundedness implies admits-induction. Unfortunately this is not the case: a true inverse, viz. complementation, is needed to do that. To prove the theorem using the techniques developed here it suffices to know that $R \backslash S = \neg(R^{\cup} \circ \neg S)$. (We haven't given enough information about relation algebra for you to verify this fact within the

algebra. A pointwise verification can, of course, be given instead.) This fact can then be used to construct a function f such that $\nu\langle T :: T \circ R\rangle = f.\mu\langle T :: R \backslash T\rangle$. μ-fusion should be used bearing in mind the Galois connection $\neg R \subseteq S \equiv R \supseteq \neg S$ and being particularly careful about the reversal of the ordering relation. Having constructed f it is then straightforward to establish the equivalence between the two notions.

Prove that well-foundedness admits induction along the lines outlined above.

In general, the right condition $\nu\langle T :: T \circ R\rangle$ can be interpreted as the set of all points from which an infinite R-chain begins.

What is the interpretation of $\mu\langle T :: R \backslash T\rangle$?

□

6 Hylo Equations

A central issue of computing science is the development of practical programming methodologies. Characteristic of a programming methodology is that it involves a *discipline* designed to maximise confidence in the reliability of the end product. The discipline constrains the construction methods to those that are demonstrably simple and easy to use, whilst still allowing sufficient flexibility that the creative process of program construction is not impeded. An example of such a methodology is the combined use of invariant relations and well-founded relations in the design of iterative algorithms in sequential programming (see sections 5.2 and 5.5), the constraint being the restriction of the methodology to the design of **while** statements rather than allowing arbitrary use of **goto** statements.

In this section we introduce a methodology for the design of recursive programs. The methodology is based on constraining the recursion to a particular form of fixed point equation, called a "hylo" equation, rather than allowing arbitrary recursion (which has been called the **goto** of functional programming). The methodology generalises the methodology for designing **while** statements by introducing a datatype as an additional parameter in the design process. (In the case of **while** statements the datatype is just the set of natural numbers, in the case of a divide-and-conquer algorithm the datatype is a tree structure.)

A hylo equation comprises three elements, a "relator" F (which is a function from relations to relations), and two relations, one of which is an "F-algebra" and the other is an "F-coalgebra". The complete definition is given in section 6.1. Section 6.2 gives a number of examples of programs that take the form of a hylo equation. It is shown that programs defined by structural or primitive recursion are instances of hylo programs as well as several standard sorting algorithms and other programs based on a divide-and-conquer strategy. The goal in this section is, of course, to demonstrate that restricting the

design methodology to hylo programs still allows sufficient room for creativity. Sections 6.3 and 6.4 introduce an important fixed-point theorem which formally relates hylo equations with the use of an intermediate or "virtual" data structure. Understanding this theorem is crucial to understanding the methodology of designing hylo programs. The final section, section 6.5 is about generalising notions of well-foundedness and inductivity to take into account the intermediate data structure implicit in any hylo program.

6.1 Relators and Hylos

A hylo equation comprises three elements, a so-called "relator" and two relations. The notion of relator plays the same role in relation algebra as the notion of "functor" in the category of functions and sets.

Functors are relevant to functional programming because they correspond to type constructors. The canonical example is List, which is an endofunctor on the category Fun. The object part of the functor List is the mapping from types (sets) to types. (For example $\mathsf{List}.\mathbb{N}$, lists of natural numbers, is the result of applying List to $\mathbb{N}$.) The arrow part of the functor List is the function known as `map` to functional programmers. If $f \in I \leftarrow J$ then $\mathsf{map}.f \in \mathsf{List}.I \leftarrow \mathsf{List}.J$ is the function that applies function f to each element in a list of Js to create a list of Is of the same length. It is a general fact that parameterised datatypes (of which List is an example) define functors. The object part of the functor is the mapping from types to types and the arrow part is the "map" operation that applies a given function to every value stored in an instance of the datatype.

Rather than constrain ourselves to the design of functional programs, we consider programs involving relations as well. (The reasons are obvious: doing so means that we may allow non-determinism in our programs and do not have to make an arbitrary distinction between specifications —which typically involve an element of non-determinism— and implementations. Also, as the theory below shows, there is no good reason for not extending the discussion to include relations.) But the categorical notion of functor is too weak to describe type constructors in the context of a relational theory of datatypes. The notion of an "allegory" [Fv90] extends the notion of a category in order to better capture the essential properties of relations, and the notion of a "relator" [BBM+91, BW93] extends the notion of a functor in order to better capture the relational properties of datatype constructors.

Formally an *allegory* is a category such that, for each pair of objects A and B, the class of arrows of type $A \leftarrow B$ forms an ordered set. In addition there is a converse operation on arrows and a meet (intersection) operation on pairs of arrows of the same type. These are the minimum requirements in order to be able to state the algebraic properties of the converse operation. For practical purposes more is needed. A *locally-complete, tabulated, unitary, division allegory* is an allegory such that, for each pair

of objects A and B, the partial ordering on the set of arrows of type $A \leftarrow B$ is complete ("locally-complete"), the division operators introduced in section 5.3.2 are well-defined ("division allegory"), the allegory has a unit (which is a relational extension of the categorical notion of a unit — "unitary") and, finally, the allegory is "tabulated". We won't go into the details of what it means to be "tabulated" but, basically, it means that every arrow in the allegory can be represented by a pair of arrows in the underlying map category (i.e. by a pair of functions) and captures the fact that relations are subsets of the cartesian product of a pair of sets. (Tabularity is vital because it provides the link between categorical properties and their extensions to relations.)

A suitable extension to the notion of functor is the notion of a "relator". A *relator* is a functor whose source and target are both allegories —remember that an allegory is a category— that is monotonic with respect to the subset ordering on relations of the same type and commutes with converse. Thus, a *relator* F is a function to the objects of an allegory $\mathcal{C}$ from the objects of an allegory $\mathcal{D}$ together with a mapping to the arrows (relations) of $\mathcal{C}$ from the arrows of $\mathcal{D}$ satisfying the following properties:

(66) $\quad F.R \in F.I \xleftarrow{\mathcal{C}} F.J$ whenever $R \in I \xleftarrow{\mathcal{D}} J$.

(67) $\quad F.R \circ F.S = F.(R \circ S)$ for each R and S of composable type,

(68) $\quad F.id_A = id_{F.A}$ for each object A,

(69) $\quad F.R \subseteq F.S \;\Leftarrow\; R \subseteq S$ for each R and S of the same type,

(70) $\quad (F.R)^{\cup} = F.(R^{\cup})$ for each R.

Two examples of relators are List and product. List is a unary relator, and product is a binary relator. If R is a relation of type $I \leftarrow J$ then List.R relates a list of Is to a list of Js whenever the two lists have the same length and corresponding elements are related by R. The relation $R \times S$ relates two pairs if the first components are related by R and the second components are related by S. List is an example of an inductively-defined datatype; in [BBH+92] it was observed that all inductively-defined datatypes are relators.

Now that we have the definition of a relator we may also give the definition of a hylo equation.

Definition 71 (Hylos) Let R and S be relations and F a relator. An equation of the form

(72) $\quad X:: \quad X = R \circ F.X \circ S$

is said to be a *hylo equation* or *hylo program*.
□

(The name "hylomorphism" for the least solution of a (functional) equation of this form originated in [Fok92].)

Note that, on typing grounds, if the unknown X in equation (72) is to have type $A \leftarrow B$ then R must have type $A \leftarrow F.A$. We say that R is an F-*algebra with carrier* A. Also S must have type $F.B \leftarrow B$ (equivalently $S^\cup$ must be an F-algebra with carrier B). It is convenient to use the term *coalgebra* for a relation of type $F.B \leftarrow B$ for some B. So a *coalgebra with carrier* B is the converse of an algebra with carrier B.

6.2 Hylo Programs

In this section we show how frequently recursive programs can be rewritten in the form of hylo equations. We consider a variety of classes of recursion: structural recursion, primitive recursion, divide-and-conquer, and so on. In order to show that each of these classes is subsumed by the class of hylo equations some additional notation is introduced as and when necessary. It is not necessay to understand the notation in detail in order to be able to appreciate the examples, and the notation will not be used elsewhere.

Structural recursion The heart of functional programming is the declaration and use of datatypes. This is facilitated by the special purpose syntax that is used. A definition like that of the natural numbers in Haskell:

```
datatype Nat = Zero | Succ Nat
```

introduces two datatype constructors `Zero` and `Succ` of types `Nat` and `Nat -> Nat`, respectively. It also facilitates the definition of functions on natural numbers by pattern matching as in the definition of the function `even`:

```
even Zero = True

even (Succ n) = not (even n)
```

Category theory enables one to gain a proper understanding of such definitions and to lift the level of discussion from particular instances of datatypes to the general case, thus improving the effectiveness of program construction.

Category theory encourages us to focus on function composition rather than function application and to combine the two equations above into one equation, namely:

(73) $\quad \mathsf{even} \circ (\mathsf{zero} \triangledown \mathsf{succ}) \;=\; (\mathsf{true} \triangledown \mathsf{not}) \circ (\mathbb{1} + \mathsf{even})$.

In this form various important elements are more readily recognised. First, the two datatype constructors `Zero` and `Succ` have been combined into one *algebra* $\mathsf{zero} \triangledown \mathsf{succ}$. Similarly, `True` and `not` have been combined into the algebra $\mathsf{true} \triangledown \mathsf{not}$. The general

mechanism being used here is the disjoint sum type constructor $(+)$ and the case operator $(\triangledown)$. Specifically, given types A and B, their disjoint sum $A+B$ comprises elements of A together with elements of B but tagged to say in which component of the disjoint sum they belong. Application of the function $f\triangledown g$ to a value of type $A+B$ involves inspecting the tag to see whether the value is in the left component of the sum or in the right. In the former case the function f is applied (after stripping off the tag); in the latter case the function g is applied. Thus for $f\triangledown g$ to be correctly typed, f and g must have the same target type. Then, if f has type $A\leftarrow B$ and g has type $A\leftarrow C$, the type of $f\triangledown g$ is $A\leftarrow B+C$.

Another important element of (73) is the *unit type* $\mathbb{1}$ and the term $\mathbb{1}+\mathsf{even}$. The unit type is a type with exactly one element. The term $\mathbb{1}+\mathsf{even}$ is read as the *functor* $\mathbb{1}+$ applied to the function even. As explained earlier, if f has type $A\leftarrow B$ the function $\mathbb{1}+f$ has type $\mathbb{1}+A\leftarrow\mathbb{1}+B$. It is the function that inspects the tag on a value of type $\mathbb{1}+B$ to see if it belongs to the left component, $\mathbb{1}$, or the right component, B. In the former case the value is left unaltered (complete with tag), and in the latter case the function f is applied to the untagged value, and then the tag is replaced. The functor $\mathbb{1}+$ is called the *pattern functor* of the datatype $\mathbb{N}$ (`Nat` in Haskell-speak) [BJJM99].

The final aspect of (73) that is crucial is that it uniquely defines the function even. (To be precise, the equation

$$X::\quad X\circ(\mathsf{zero}\triangledown\mathsf{succ}) = (\mathsf{true}\triangledown\mathsf{not})\circ(\mathbb{1}+X)$$

has a unique solution.) This is the concept of *initiality* in category theory. Specifically, $\mathsf{zero}\triangledown\mathsf{succ}$ is an *initial* $(\mathbb{1}+)$*-algebra* which means that for all $(\mathbb{1}+)$-algebras f the equation

$$X::\quad X\circ(\mathsf{zero}\triangledown\mathsf{succ}) = f\circ(\mathbb{1}+X)$$

has exactly one solution.

In summary, category theory identifies three vital ingredients in the definition (73) of the function even, namely, the functor $\mathbb{1}+$, the initial $(\mathbb{1}+)$-algebra $\mathsf{zero}\triangledown\mathsf{succ}$ and the $(\mathbb{1}+)$-algebra $\mathsf{true}\triangledown\mathsf{not}$.

The general form exemplified by (73) is

$$(74)\qquad X\circ\mathsf{in} = f\circ F.X$$

where F is a functor, in is an initial F-algebra and f is an F-algebra. This general form embodies the use of structural recursion in modern functional programming languages like Haskell. The left side embodies pattern matching since, typically, in embodies a case analysis as exemplified by $\mathsf{zero}\triangledown\mathsf{succ}$. The right side exhibits recursion over the structure of the datatype, which is represented by the "pattern" functor F.

Here is the formal definition of an initial algebra. The definition is standard —an initial object in the category of F-algebras— but we give it nonetheless in order to introduce some terminology.

Definition 75 Suppose F is an endofunctor on some category $\mathcal{C}$. An arrow f in $\mathcal{C}$ is an F-*algebra* if $f \in A \leftarrow FA$ for some A, the so-called *carrier* of the algebra. If f and g are both F-algebras with carriers A and B then arrow $\varphi \in A \leftarrow B$ is said to be an F-algebra *homomorphism* to f from g if $\varphi \circ f = g \circ F.\varphi$. The category $FAlg$ has objects all F-algebras and arrows all F-algebra homomorphisms. Composition and identity arrows are inherited from the base category $\mathcal{C}$. The arrow $in \in I \leftarrow FI$ is an *initial* F-algebra if for each $f \in A \leftarrow FA$ there exists an arrow $(\!|f|\!) \in A \leftarrow I$ such that for all $h \in A \leftarrow I$,

$$(76) \qquad h = (\!|f|\!) \;\equiv\; h \in f \xleftarrow{FAlg} in \quad .$$

So, $(\!|f|\!)$ is the unique homomorphism to algebra f from algebra in. We call $(\!|f|\!)$ the *catamorphism* of f.

□

The "banana bracket" notation for catamorphisms (as it is affectionately known) was introduced by Malcolm [Mal90a, Mal90b]. Malcolm was also the first to express the unicity property using an equivalence in this way. It is a mathematically trivial device but it helps enormously in reasoning about catamorphisms. Note that the functor F is also a parameter of $(\!|f|\!)$ but the notation does not make this explicit. This is because the functor F is usually fixed in the context of the discussion. Where disambiguation is necessary, the notation $(\!|F\,;f|\!)$ is sometimes used. The initial algebra is also a parameter that is not made explicit; this is less of a problem because initial F-algebras are isomorphic and thus catamorphisms are defined "up to isomorphism".

An important property of initial algebras, commonly referred to as Lambek's lemma [Lam68], is that an initial algebra is both injective and surjective. Thus, for example, $zero \triangledown succ$ is an isomorphism between $\mathbb{N}$ and $\mathbb{1}+\mathbb{N}$. Lambek's lemma has the consequence that, if in is an initial F-algebra,

$$h \in f \xleftarrow{FAlg} in \;\equiv\; h = f \circ F.h \circ in^{\cup}$$

where $in^{\cup}$ is the inverse of in. Thus, the characterising property (76) of catamorphisms is equivalent to, for all h and all F-algebras f,

$$(77) \qquad h = (\!|f|\!) \;\equiv\; h = f \circ F.h \circ in^{\cup} \quad .$$

That is, $(\!|f|\!)$ is the unique fixed point of the function mapping h to $f \circ F.h \circ in^{\cup}$. Equivalently, $(\!|f|\!)$ is the unique solution of the hylo equation:

$$h:: \quad h = f \circ F.h \circ in^{\cup} \quad .$$

In the context of functions on lists the catamorphism $(\!|f|\!)$ is known to functional programmers as a *fold* operation. Specifically, for lists of type I the relevant pattern functor F is the functor mapping X to $\mathbb{1}+(I \times X)$ (where $\times$ denotes the cartesian product functor) and an F-algebra is a function of type $A \leftarrow \mathbb{1}+(I \times A)$ for some A. Thus an F-algebra takes the form $c \triangledown (\oplus)$ for some function c of type $A \leftarrow \mathbb{1}$ and some function $\oplus$ of type $A \leftarrow I \times A$. The characterising property of the catamorphisms is thus

$$h = (\!|c \triangledown (\oplus)|\!) \equiv h = (c \triangledown (\oplus)) \circ (\mathbb{1} + (I \times h)) \circ (nil^{\cup} \blacktriangledown cons^{\cup}) \quad .$$

Here $nil^{\cup} \blacktriangledown cons^{\cup}$ is the inverse of $nil \triangledown cons$. (In general, $R \blacktriangledown S$ is the converse conjugate of $R \triangledown S$. That is, $(R \blacktriangledown S)^{\cup} = R^{\cup} \triangledown S^{\cup}$.) It can be read as the pattern matching operator: look to see whether the argument is an empty list or a non-empty list. In the former case $nil^{\cup}$ returns an element of the unit type, tagging it so that the result of the test is passed on to later stages; in the latter case $cons^{\cup}$ splits the list into a head and a tail, the resulting pair also being tagged for later identification. Using the algebraic properties of case analysis, the characterising property is equivalent to

$$h = (\!|c \triangledown (\oplus)|\!) \equiv h \circ nil = c \wedge h \circ cons = (\oplus) \circ (I \times h)$$

the right side of which is a point-free free formulation of the definition of a fold with seed the constant c and binary operator $(\oplus)$. As a concrete example, the function sum that sums the elements of a list is

$$(\!|zero \triangledown add|\!)$$

where `add` is the addition function. In Haskell this function would be written

```
fold 0 add .
```

Although catamorphisms (folds) are best known in the context of functional programming many relations are also catamorphisms. For example, the prefix relation on lists is uniquely characterised by the two equations

$$nil [\![prefix]\!] nil$$

and $\quad xs [\![prefix]\!] (y{:}ys) \equiv xs = nil \vee \exists (zs:: \ xs = y{:}zs \wedge zs [\![prefix]\!] ys) \quad .$

Expressed as one, point-free equation this is

$$(78) \qquad \mathsf{prefix} \circ (\mathsf{nil} \triangledown \mathsf{cons}) \;=\; (\mathsf{nil} \triangledown ((\mathsf{nil} \circ \top\!\!\top) \cup \mathsf{cons})) \circ (\mathbb{1} + (\mathrm{I} \times \mathsf{prefix}))$$

where I denotes the type of the list elements. Here we recognise a *relator* and two algebras: in this case the relator is $(\,\mathbb{1}+(\mathrm{I}\times)\,)$ and the two $(\,\mathbb{1}+(\mathrm{I}\times)\,)$-algebras are $\mathsf{nil} \triangledown \mathsf{cons}$ and $\mathsf{nil} \triangledown ((\mathsf{nil} \circ \top\!\!\top) \cup \mathsf{cons})$. (Note that the second algebra is not a function.) Equivalently, prefix is the unique solution of a hylo equation:

$$(79) \qquad \mathsf{prefix} \;=\; (\mathsf{nil} \triangledown ((\mathsf{nil} \circ \top\!\!\top) \cup \mathsf{cons})) \circ (\mathbb{1} + (\mathrm{I} \times \mathsf{prefix})) \circ (\mathsf{nil}^{\cup} \blacktriangledown \mathsf{cons}^{\cup}) \quad .$$

Primitive recursion Structural recursion is useful since many programs that arise in practice have this kind of recursion. However, just as structural induction is not enough to prove all facts that can be proved by induction, structural recursion is not enough to define all programs that can be defined by recursion. As an example of a program that is not structurally recursive, consider the factorial function, the function defined by the two equations

$$\mathsf{fact} \circ \mathsf{zero} = \mathsf{one} \quad \text{and} \quad \mathsf{fact} \circ \mathsf{succ} = \mathsf{times} \circ (\mathsf{fact} \vartriangle \mathsf{succ}) \quad ,$$

where one is the constant function returning the number 1 and times is the multiplication function. These equations can be combined into the single equation

$$(80) \qquad \mathsf{fact} \;=\; (\mathsf{one} \triangledown (\mathsf{times} \circ (\mathsf{succ} \times \mathbb{N}))) \circ (\mathbb{1} + (\mathbb{N} \vartriangle \mathsf{fact})) \circ (\mathsf{zero}^{\cup} \blacktriangledown \mathsf{succ}^{\cup}) \qquad .$$

Reading from the right, the factorial function first examines its argument to determine whether it is zero or the successor of another number; in the former case a tagged element of the unit type is returned, and in the latter case the predecessor of the input value is returned, suitably tagged. Subsequently, if the input value is $n+1$, the function $\mathbb{N} \vartriangle \mathsf{fact}$ constructs a pair consisting of the number n and the result of the factorial function applied to n. (As forewarned, $\mathbb{N}$ is used here to denote the identity function on natural numbers.) The calculation of $(n+1) \times n!$ is the result of applying the function $\mathsf{times} \circ (\mathsf{succ} \times \mathbb{N})$ to the (untagged) pair. On the other hand, if the input value is zero then one is returned as result.

To give an example of a relation defined by primitive recursion we need look no further than the suffix relation on lists. It satisfies

$$\mathsf{nil} \,[\![\mathsf{suffix}]\!]\, \mathsf{nil}$$

$$\text{and} \qquad xs \,[\![\mathsf{suffix}]\!]\, (y{:}ys) \quad \equiv \quad xs = y{:}ys \;\vee\; xs \,[\![\mathsf{suffix}]\!]\, ys \quad .$$

Expressed as a fixed point equation this is:

$$\mathsf{suffix} \;=\; (\mathsf{nil} \triangledown ((\mathsf{cons} \circ \mathsf{exl}) \cup (\mathsf{exr} \circ \mathsf{exr}))) \circ (\mathbb{1} + (\mathrm{I} \times (\mathsf{List}.\mathrm{I} \vartriangle \mathsf{suffix}))) \circ (\mathsf{nil}^{\cup} \blacktriangledown \mathsf{cons}^{\cup})$$

where I is the type of the list elements and exl and exr project a pair onto its left and right components, respectively. This is a definition by primitive recursion.

When we abstract from the particular functor and initial algebra in factorial program (80) a general recursion scheme is obtained.

$$X \;::\; X \;=\; \mathrm{R} \circ \mathrm{F}.(\mathrm{I} \times X) \circ \mathrm{F}.(\mathrm{I} \vartriangle \mathrm{I}) \circ \mathsf{in}^{\cup} \quad . \tag{81}$$

In the case of the factorial function R is $\mathsf{one} \triangledown (\mathsf{times} \circ \mathsf{succ} \times \mathsf{id})$, F is $(\mathbb{1} +)$, I is the (identity on) natural numbers and in is $\mathsf{zero} \triangledown \mathsf{succ}$. (Note that $(W \times X) \circ (\mathrm{I} \vartriangle \mathrm{I}) = W \vartriangle X$ for all W and X with source I. Hence $\mathrm{F}.(\mathrm{I} \times X) \circ \mathrm{F}.(\mathrm{I} \vartriangle \mathrm{I}) = \mathrm{F}.(\mathrm{I} \vartriangle X)$. We have applied this so-called $\times$-$\vartriangle$-fusion law in order to make the term $\mathrm{F}.(\mathrm{I} \vartriangle \mathrm{I})$ explicit.) A definition of this form is called *primitive recursive*.

This generic formulation of primitive recursion was introduced (for functions) by Meertens [Mee92]. He called such an equation a *para equation* and a solution to the equation a *paramorphism*.

Divide and Conquer As the name suggests, "primitive" recursion is also unsuitable as the basis for a practical methodology of recursive program construction. Divide-and-conquer is a well-known technique that is not easily expressed using primitive recursion.

An example of a divide-and-conquer program is the sorting algorithm known as "quicksort". Quicksort, here abbreviated to qs, is uniquely defined by the hylo equation:

$$\mathsf{qs} \;=\; (\mathsf{nil} \triangledown (\mathsf{join} \circ (\mathrm{I} \times \mathsf{cons}))) \circ (\mathbb{1} + (\mathsf{qs} \times (\mathrm{I} \times \mathsf{qs}))) \circ (\mathsf{nil}^{\cup} \blacktriangledown \mathsf{dnf}) \tag{82}$$

To see that this is the quicksort program one has to interpret dnf as the well-known "Dutch national flag" relation: the relation that splits a non-empty list into a tuple $(\mathrm{xs}, (\mathrm{x}, \mathrm{ys}))$ formed by a list, an element and a list such that all elements in the list xs are at most x and all elements in ys are greater than x. The results of the recursive calls are assembled to the output list by the operation $\mathsf{join} \circ (\mathrm{I} \times \mathsf{cons})$, where join produces the concatenation of two lists.

A typical divide and conquer program is of the form

$$X = (\mathrm{R} \triangledown \mathsf{conquer}) \circ (\mathrm{I} + (X \times X)) \circ (\mathrm{I} + \mathsf{divide}) \circ (A \blacktriangledown B) \quad . \tag{83}$$

Interpreting this program should not be difficult. A test is made to determine whether the input is a base case (if the input satisfies A), the output then being computed by R. If the input is not a base case (if the input satisfies B) the input is split into two

smaller "subproblems" by divide. Then the smaller problems are solved recursively and finally the two solutions of the subproblems are assembled into an output by conquer.

Of course there are more divide and conquer schemes. For example, the original problem can be split into more than two subproblems. It is also possible that the divide step produces, besides a number of subproblems, a value that is not "passed into the recursion"; then the middle relation of (83) has a form like $I \times (X \times X)$. Quicksort is an example of such a divide and conquer algorithm.

Repetition is an elementary and familiar example of divide and conquer in which the original problem is reduced to a single subproblem. A repetition is a solution of the equation in x:

$$(84) \qquad x \;=\; \mathsf{if}\ \neg b \to \mathsf{skip}\ [\!]\ b \to s;x\ \mathsf{fi} \quad .$$

Using the fact that skip (do nothing) corresponds to the identity function, I, on the state space and writing B for the coreflexive corresponding to predicate b and S for the relation corresponding to the statement s, we may express (84) using disjoint sum as:

$$(85) \qquad X \;=\; (I \triangledown I) \circ (I + X) \circ (\sim B \blacktriangledown (S \circ B)) \quad .$$

Here we see how **while** statements are expressed in terms of hylo equations, the relator being $(\ I+\ )$.

Parameterised recursion Often recursive programs conform to one of the schemes discussed above but this is obscured by the presence of an additional parameter. Elementary examples are the definitions of addition, multiplication and exponentiation on natural numbers, which are essentially, but not quite, definitions by structural recursion:

$$0+n \;=\; n \quad \text{and} \quad (m+1)+n \;=\; (m+n)+1 \ ,$$

$$0 \times n \;=\; 0 \quad \text{and} \quad (m+1) \times n \;=\; m \times n + n \ ,$$

$$n^0 \;=\; 1 \quad \text{and} \quad n^{m+1} \;=\; n^m \times n \ .$$

All these definitions have the form

$$X.(0, n) \;=\; f.n \quad \text{and} \quad X.(m+1\,,\, n) \;=\; g.(m\,,\, h.n)$$

where X is the function being defined and f, g and h. (We leave the reader to supply the instantiations for f, g and h.) In point-free form, we have yet again a hylo equation:

$$X \;=\; k \circ ((\mathbb{1} + X) \times \mathbb{N}) \circ (\mathsf{pass} \vartriangle \mathsf{exr}) \circ ((\mathsf{zero}^\cup \blacktriangledown \mathsf{succ}^\cup) \times \mathbb{N})$$

where $k \;=\; ((f \circ \mathsf{exr}) \triangledown (g \circ (\mathsf{id} \times h))) \circ \mathsf{distr}$.

Here distr is a function of type $(H \times K) + (J \times K) \leftarrow (H + J) \times K$ that is polymorphic in H, J and K, and pass is a function of type $\mathbb{1} + (I \times K) \leftarrow (\mathbb{1} + I) \times K$ that is polymorphic in I and K.

Despite the seeming complexity of the underlying algebra and coalgebra, the basic structure is thus a hylo equation.

Another example, with the same structure but defined on a datatype other than the natural numbers, is the program that appends two lists. The standard definition comprises the two equations

$$\mathsf{nil} +\!\!+ ys = ys \quad \text{and} \quad (x{:}xs) +\!\!+ ys = x{:}(xs +\!\!+ ys) \quad .$$

As a single equation (where we write join instead of $+\!\!+$):

$$\mathsf{join} \;=\; \mathsf{post} \circ ((\mathbb{1} + (I \times \mathsf{join})) \times \mathsf{List}.I) \circ (\mathsf{pass} \vartriangle \mathsf{exr}) \circ ((\mathsf{nil}^\cup \blacktriangledown \mathsf{cons}^\cup) \times \mathsf{List}.I) \quad .$$

where $\mathsf{post} \;=\; (\mathsf{exr} \triangledown \mathsf{cons}) \circ \mathsf{distr}$. Here distr is as before whereas in this case pass is a function of type $\mathbb{1} + (I \times (J \times K)) \leftarrow (\mathbb{1} + (I \times J)) \times K$ that is polymorphic in I, J and K. Once again we recognise a hylo equation.

6.3 Intermediate data structures

At the beginning of section 6.2 we discussed the use of recursion on the structure of a datatype; if R is an F-algebra with carrier A then the catamorphism $(\!|R|\!)$ can be seen as a program that *destructs* an element of an initial F-algebra in order to compute a value of type A. The converse $(\!|R|\!)^\cup$ is thus a program that *constructs* an element of the initial algebra from a value of type A.

Now suppose R and $S^\cup$ are both F-algebras with carriers A and B, respectively. Then the composition $(\!|R|\!) \circ (\!|S^\cup|\!)^\cup$ has type $A \leftarrow B$. It computes a value of type A from a value of type B by first building up an intermediate value which is an element of an initial F-algebra and then breaking the element down. The remarkable theorem is that

$(\!|R|\!) \circ (\!|S^\cup|\!)^\cup$ is the least solution of the hylo equation (72).

This theorem (which we formulate precisely below) gives much insight into the design of hylo programs. It says that executing a hylo program is equivalent to constructing an intermediate data structure, the form of which is specified by the relator F, and then breaking this structure down. The two phases are called the *anamorphism* phase and the *catamorphism* phase. Executing a hylo equation for a specific input value by

unfolding the recursion hides this process; it is as if the intermediate data structure is broken down as it is being built up. (A good comparison is with a Unix pipe in which the values in the pipe are consumed as soon as they are produced.) Execution of $(\!|R|\!) \circ (\!|S^\cup|\!)^\cup$ does make the process explicit. For this reason, the relator F is said to specify a *virtual* data structure [SdM93].

Two simple examples of virtual data structures are provided by do-statements and the factorial function. In the case of do-statements (see (85)) the virtual datatype is the carrier set of an initial $(I+)$-algebra, a type which is isomorphic to $I \times \mathbb{N}$ —thus an element of the virtual datatype can be seen as a pair consisting of an element of the state space and a natural number, the latter being a "virtual" count of the number of times the loop body is executed. In the case of the factorial function, definition (80) can be rewritten so as to make the relator F explicit:

$$\mathsf{fact} \;=\; (\mathsf{one} \triangledown (\mathsf{times} \circ (\mathsf{succ} \times \mathbb{N}))) \circ (\mathbb{1} + (\mathbb{N} \times \mathsf{fact})) \circ (\mathsf{zero}^\cup \blacktriangledown ((\mathbb{N} \vartriangle \mathbb{N}) \circ \mathsf{succ}^\cup)) \quad .$$

The "virtual" datatype is thus the type of lists of natural numbers, the carrier set of an initial $\mathbb{1}+(\mathbb{N}\times)$-algebra. The list that is constructed for a given input n is the list of natural numbers from $n-1$ down to 0 and the hylo theorem states that the factorial of n can be calculated by constructing this list (the anamorphism phase) and then multiplying the numbers together after adding 1 (the catamorphism phase).

Language recognition also illustrates the process well. Let us explain the process first with a concrete example following which we will sketch the generic process. Consider the following grammar:

$$S \;::=\; aSb \mid c$$

where, for our purposes, a, b and c denote some arbitrary sets of words over some fixed alphabet. Associated with this grammar is a data structure: the class of parse trees for strings in the language generated by the grammar. This data structure, Stree, satisfies the equation:

$$\mathsf{Stree} \;=\; (a \times \mathsf{Stree} \times b) + c \quad .$$

It is an initial F-algebra where F maps X to $(a \times X \times b) + c$. Now the process of *unparsing* a parse tree is very easy to describe since it is defined by induction on the structure of parse trees. Indeed the unparse function is the F-catamorphism $(\!|(\mathsf{concat3} \circ (a \times \mathsf{id} \times b)) \triangledown c|\!)$ where $\mathsf{concat3}$ concatenates three strings together, a, b and c are the identity functions on the sets a, b and c, and id is the identity function on all words. Moreover, its left domain is equal to the language generated by the grammar. Since in general the left domain of function f is $f \circ f^\cup$ the language generated satisfies

$$S \;=\; (\!|(\mathsf{concat3} \circ (a \times \mathsf{id} \times b)) \triangledown c|\!) \circ (\!|(\mathsf{concat3} \circ (a \times \mathsf{id} \times b)) \triangledown c|\!)^\cup \quad .$$

This equation defines a (nondeterministic) program to recognise strings in the language. The program is a partial identity on words. Words are recognised by first building a parse tree and then unparsing the tree. By the hylo theorem (or directly from the definition of S) we also have the hylo program

$$S \;=\; ((\mathsf{concat3} \circ (a \times \mathsf{id} \times b)) \triangledown c) \circ ((a \times S \times b) + c) \circ (((a \times \mathsf{id} \times b) \circ \mathsf{concat3}^{\cup}) \blacktriangledown c) \quad .$$

This is a program that works by (nondeterministically) choosing to split the input word into three segments (using $\mathsf{concat3}^{\cup}$) or to check whether the word is in the language c. In the former case the first segment is checked for membership in a, the third segment is checked for membership in b and the program is called recursively to check the middle segment. Subsequently the three segments are recombined into one. In the latter case the word is left unchanged.

The derivation of a language recogniser in this way can be generalised to an arbitrary context-free grammar. (This is only possible because we base our methodology on relation algebra. The non-determinism present in a typical context-free grammar prohibits the generalisation we are about to make in the context of functional programming.) A context-free grammar defines a type of parse trees in a fairly obvious way. Also an $\mathsf{unparse}$ function can always be defined mapping parse trees to strings. This function is a catamorphism. The language generated by the grammar is the left domain of the unparse function, which is $\mathsf{unparse} \circ \mathsf{unparse}^{\cup}$. This in turn is the composition of a catamorphism and the converse of a catamorphism, which can be expressed as a hylo program using the hylo theorem.

In practice the process is complicated by the fact that all practical context-free grammars have more than one nonterminal, and nonterminals are linked together via mutual recursion. But the theory we have developed covers this case too. Mutual recursion is modelled by endorelators on a product category.

6.4 The Hylo Theorem

We summarise the previous section with a formal statement of the hylo theorem. The theorem is rather deeper than just the statement that the least solution of a hylo equation is the composition of a catamorphism and an anamorphism. The proof of the theorem has been given in detail elsewhere [BH99][4].

[4] Actually [BH99] contains a proof of the dual theorem concerning final coalgebras and is more general than the theorem stated here. Unlike in a category, dualising between initiality and finality is not always straightforward in an allegory because of the lack of duality between intersection and union. However, dualising from a finality property to an initiality property is usually straightforward and it is the other direction that is difficult. That is one reason why [BH99] chose to present the theorem in terms of coalgebras rather than algebras. The extra generality offered by the theorem in [BH99] encompasses the relational

Recall that we defined the notion of an initial algebra in the context of a category. (See (75).) To all intents and purposes this amounts to defining the notion of an initial algebra in the context of functions between sets. What we need however is the notion of an initial algebra in the context of binary relations on sets, that is, in the context of an allegory. Definition 86 is such a definition. The hylo theorem states that the categorical notion of an initial algebra coincides with the allegorical notion if the allegory is locally complete and tabular.

Definition 86 Assume that F is an endorelator. Then (I, in) is a *relational initial* F-algebra iff $in \in I \leftarrow F.I$ is an F-algebra and there is a mapping $(\!|_|\!)$ defined on all F-algebras such that

(87) $\quad (\!|R|\!) \in A \leftarrow I \quad \text{if } R \in A \leftarrow F.A \quad ,$

(88) $\quad (\!|in|\!) = id_I \quad , \text{ and}$

(89) $\quad (\!|R|\!) \circ (\!|S|\!)^{\cup} \;=\; \mu\langle X :: R \circ F.X \circ S^{\cup} \rangle \quad .$

That is, $(\!|R|\!) \circ (\!|S|\!)^{\cup}$ is the smallest solution of the equation $X :: R \circ F.X \circ S^{\cup} \subseteq X$.
□

In order to state the hylo theorem we let $Map(\mathcal{A})$ denote the sub-category of functions in the allegory $\mathcal{A}$. For clarity we distinguish between the endorelator F and the corresponding endofunctor defined on $Map(\mathcal{A})$.

Theorem 90 (Hylo Theorem) Suppose F is an endorelator on a locally-complete, tabular allegory $\mathcal{A}$. Let F' denote the endofunctor obtained by restricting F to the objects and arrows of $Map(\mathcal{A})$. Then in is an initial F'-algebra if and only it is a relational initial F-algebra.

□

6.5 Reducing problem size

There are two elements in the design of the body of a **while** statements: it should maintain an invariant relation established by the initialisation procedure, and it should make progress to the termination condition. The latter is guaranteed if the loop body is a well-founded relation on the state space. There are also two elements in the design of hylo equations. The intermediate data structure plays the role of the invariant relation, whilst making progress is achieved by ensuring that each recursive call is "smaller" than

properties of disjoint sum and cartesian product but at the expense of requiring a more sophisticated understanding of allegory theory which we wanted to avoid in the current presentation.

the original argument. In this section we formalise this requirement. The notion we introduce, "F-reductivity" due to Henk Doornbos [Doo96], generalises the notion of admitting induction essentially by making the intermediate data structure a parameter. As we shall indicate in section 6.6 this has important ramifications for developing a calculus of program termination.

Informally, for hylo program $X = S \circ F.X \circ R$ we require that all values stored in an output F-structure of R have to be smaller than the corresponding input to R. More formally, with $x[\![\mathsf{mem}]\!]y$ standing for "x is a member of F-structure y" (or, x is a value stored in F-structure y"), we demand that for all x and z

$$\forall\langle y:: x[\![\mathsf{mem}]\!]y \wedge y[\![R]\!]z \Rightarrow x \prec z\rangle \quad ,$$

for some well-founded ordering $\prec$. If this is the case we say that R is F-*reductive*.

To make the definition of reductivity completely precise we actually want to avoid the concept of "values stored in an F-structure". (This is because its incorporation into the definition of F-reductivity limits the practicality of the resulting theory.) Fortunately, Hoogendijk and De Moor [HdM00, Hoo97] have shown how to characterise membership of a so-called "container" type in such a way that it can be extended to other types where the intuitive notion of "membership" is not so readily apparent.

Hoogendijk and De Moor's characterisation of the membership relation of a relator is the following:

Definition 91 (Membership) Relation $\mathsf{mem} \in I \leftarrow F.I$ is a membership relation of relator F if and only if it satisfies, for all coreflexives A, $A \subseteq I$:

$$F.A = \mathsf{mem} \backslash A \quad .$$

□

When this definition is expressed pointwise it reads:

$$x \in F.A \equiv \forall\langle i: i[\![\mathsf{mem}]\!]x: i \in A\rangle \quad .$$

Informally: an F-structure satisfies the property $F.A$ iff all the values stored in the structure satisfy property A. For example, for the list relator mem holds between a point and a list precisely when the point is in the list. For product the relation holds between x and (x,y) and also between y and (x,y).

This definition of membership leads to a definition of F-reductivity independent of the notion of values stored in an F-structure. To see this we observe that, for coalgebra R with carrier I and for coreflexive A below I, we have:

$$
\begin{array}{ll}
 & (\mathsf{mem} \circ R)\backslash\!\!\cdot A \\
= & \quad \{ \quad \text{factors (45)} \quad \} \\
 & R\backslash\!\!\cdot(\mathsf{mem}\backslash\!\!\cdot A) \\
= & \quad \{ \quad \text{definition 91} \quad \} \\
 & R\backslash\!\!\cdot \mathsf{F}.A \quad .
\end{array}
$$

Now, that $S \in I \leftarrow I$ admits induction is the condition that the least prefix point of the function $\langle A :: S\backslash\!\!\cdot A\rangle$ is I, and our informal notion of the reductivity of $R \in \mathsf{F}.I \leftarrow I$ is that $\mathsf{mem} \circ R$ should be well-founded. Since being well-founded is equivalent to admitting induction, the latter is equivalent to the requirement that the least prefix point of the function $\langle A :: R\backslash\!\!\cdot \mathsf{F}.A\rangle$ is I, which does not involve any appeal to notions of membership of a "container" type. This gives us a precise, generic definition of the notion of F-reductivity:

Definition 92 (F-reductivity) Relation $R \in \mathsf{F}.I \leftarrow I$ is said to be F-*reductive* if and only if it enjoys the property:

$$\mu((R\backslash\!\!\cdot)\bullet\mathsf{F}) = I \quad . \tag{93}$$

□

Obviously F-reductivity generalises the notion of admitting induction. (A relation R admits induction if and only if it is Id-reductive, where Id denotes the identity relator.) An immediate question is whether there is a similar generalisation of the notion of well-foundedness and a corresponding theorem that F-well-foundedness is equivalent to F-reductivity. As it turns out, there is indeed a generic notion of well-foundedness but this is strictly weaker than F-reductivity. The definition is given below, the facts just stated are left as exercises in the use of fixed point calculus.

Well-foundedness of relation R is equivalent to the equation $X :: X = X \circ R$ having a unique solution (which is obviously $\bot\!\!\bot$, the empty relation). This is easily generalised to the property that, for all relations S, the equation $X :: X = S \circ X \circ R$ has a unique solution. The generic notion of well-foundedness focusses on this unicity of the solution of equations.

Definition 94 (F-well-founded) Relation $R \in \mathsf{F}.I \leftarrow I$ is F-*well-founded* iff, for all relations $S \in J \leftarrow \mathsf{F}.J$ and $X \in J \leftarrow I$,

$$X = S \circ \mathsf{F}.X \circ R \quad \equiv \quad X = \mu\langle Y :: S \circ \mathsf{F}.Y \circ R\rangle \quad .$$

□

Exercise 95 *Verify the claim made immediately before definition 94. That is, show that* R *is well-founded equivales*

$$\forall\langle X, S:: \; X = S\circ X\circ R \;\equiv\; X = \mu\langle Y:: S\circ Y\circ R\rangle\rangle \quad .$$

In words, R is well-founded equivales R is Id-well-founded. (Hint: if R is well-founded then $\mu\langle Y:: S\circ Y\circ R\rangle = \bot\!\!\!\bot$.)

□

Exercise 96 *Prove that an* F*-reductive relation is* F*-well-founded.*

□

An example of a relation that is F-well-founded but not F-reductive can be constructed as follows. Define the relator F by $F.X = X\times X$. Suppose $R \in I\leftarrow I$ is a non-empty well-founded relation. Then the relation $R\triangle I$ of type $I\times I\leftarrow I$ (which relates a pair of values (x, y) each of type I to a single value z of type I iff x is related by R to z and $y = z$) is F-well-founded but not F-reductive. For a proof see [Doo96].

6.6 A calculus of F-reductivity

The introduction of a data structure —the relator F— as a parameter to the notion of reductivity is a significant advance because it admits the possibility of developing a calculus of reductivity and thus of program termination based on the structure of the parameter. A beginning has been made to the development of such a calculus [DB95, Doo96] sufficient to establish the termination of all the examples given in section 6.2 by a process akin to type checking.

Space only allows us to give a brief taste of the calculus here. The fundamental theorem is the following.

Theorem 97 The converse of an initial F-algebra is F-reductive.

Proof Let $\mathsf{in} \in I\leftarrow F.I$ be an initial F-algebra and A an arbitrary coreflexive of type $I\leftarrow I$. We must show that

$$I \subseteq A \quad\Leftarrow\quad \mathsf{in}^{\cup}\backslash_{\!\!\cdot}\, F.A \;\subseteq\; A \quad .$$

We start with the antecedent and derive the consequent:

$$\begin{array}{ll} & \mathsf{in}^{\cup}\backslash_{\!\!\cdot}\, F.A \subseteq A \\ \equiv & \quad \{ \quad \text{for function } f \text{ and coreflexive } B,\ f\backslash_{\!\!\cdot} B = f^{\cup}\circ B\circ f, \\ & \qquad \mathsf{in}^{\cup} \text{ is a function and } F.A \text{ is a coreflexive} \quad \} \end{array}$$

$$
\begin{array}{ll}
 & \mathsf{in} \circ \mathsf{F}.A \circ \mathsf{in}^{\cup} \subseteq A \\
\Rightarrow & \quad \{ \quad \text{Hylo theorem: (90) and (86) ,} \\
 & \qquad \mathsf{in} \text{ is an initial } \mathsf{F}\text{-algebra} \quad \} \\
 & (\!|\mathsf{in}|\!) \subseteq A \\
\equiv & \quad \{ \quad \text{identity rule: (88), } \mathsf{in} \in \mathsf{I} \leftarrow \mathsf{F}.\mathsf{I} \text{ is an initial } \mathsf{F}\text{-algebra} \quad \} \\
 & \mathsf{I} \subseteq A \quad .
\end{array}
$$

□

Theorem 97 has central importance because, if we examine all the programs in section 6.2 we see that the converse of a initial F-algebra is at the heart of the coalgebra in all the hylo equations. In the case, for example, of primitive recursion the generic equation has the form

$$X \;::\; X = \mathsf{R} \circ \mathsf{F}.(\mathsf{I} \times X) \circ \mathsf{F}.(\mathsf{I} \triangle \mathsf{I}) \circ \mathsf{in}^{\cup}$$

and the coalgebra is $\mathsf{F}.(\mathsf{I} \triangle \mathsf{I}) \circ \mathsf{in}^{\cup}$ where $\mathsf{in} \in \mathsf{I} \leftarrow \mathsf{F}.\mathsf{I}$ is an initial F-algebra. For the equation to define a terminating program (and consequently have a unique solution) we must show that the coalgebra is $(\mathsf{F}\bullet(\mathsf{I}\times))$-reductive. This is done by showing that $\mathsf{F}.(\mathsf{I} \triangle \mathsf{I})$ transforms any F-reductive relation into an $(\mathsf{F}\bullet(\mathsf{I}\times))$-reductive relation — which is a consequence of the fact that $\mathsf{F}.(\mathsf{I} \triangle \mathsf{I})$ is an instance of a natural transformation of the relator F to the relator $\mathsf{F}\bullet(\mathsf{I}\times)$.

Acknowledgements The material in this paper was developed whilst the author was heading the Mathematics of Program Construction group at Eindhoven University of Technology, particularly during the period 1990–1995. It would not have been possible to write the paper without the wonderfully stimulating and highly productive team effort that went into all we did at that time. The sections on well-foundedness and admitting induction are extracted from [DBvdW97] written jointly with Henk Doornbos and Jaap van der Woude, the sections on hylomorphisms and reductivity are extracted from Doornbos's thesis [Doo96] (see also [DB95, DB96]), and the hylomorphism theorem in the form presented here is joint work with Paul Hoogendijk [BH99].

7 Solutions to Exercises

3 The "at-most" relation is the least prefix point of the function f where, for any relation R,

$$f.R \;=\; \{n{:}\; n{\in}\mathbb{N}{:}\; (0, n)\} \cup \{m, n{:}\; (m, n){\in}R{:}\; (m{+}1\,,\; n{+}1)\} \quad .$$

□

13

$$\begin{array}{ll} & \mu f \sqsubseteq \nu f \\ \Leftarrow & \quad \{ \quad \text{induction} \quad \} \\ & f.\nu f \sqsubseteq \nu f \\ \equiv & \quad \{ \quad \text{computation: } f.\nu f = \nu f \quad \} \\ & \mathsf{true} \quad . \end{array}$$

□

15

$$\begin{array}{ll} & \leq \text{ is reflexive} \\ \equiv & \quad \{ \quad \text{definition of reflexive} \quad \} \\ & \forall\langle x{::}\, x \leq x\rangle \\ \equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\ & \forall\langle x{::}\, x{+}x = x\rangle \\ \equiv & \quad \{ \quad \text{definition of idempotent} \quad \} \\ & + \text{ is idempotent.} \end{array}$$

$$\begin{array}{ll} & \leq \text{ is transitive} \\ \equiv & \quad \{ \quad \text{definition of transitive} \quad \} \\ & \forall\langle x, y, z{::}\, x \leq y \wedge y \leq z \Rightarrow x \leq z\rangle \\ \equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\ & \forall\langle x, y, z{::}\, x{+}y = y \wedge y{+}z = z \Rightarrow x{+}z = z\rangle \end{array}$$

Now suppose $x{+}y = y$ and $y{+}z = z$. Then,

$$
\begin{array}{ll}
 & x+z \\
= & \quad \{ \quad y+z=z \quad \} \\
 & x+(y+z) \\
= & \quad \{ \quad + \text{ is associative} \quad \} \\
 & (x+y)+z \\
= & \quad \{ \quad x+y=y \quad \} \\
 & y+z \\
= & \quad \{ \quad y+z=z \quad \} \\
 & z \quad .
\end{array}
$$

Thus $\leq$ is transitive.
Now for the antisymmetry of $\leq$.

$$
\begin{array}{ll}
 & x \leq y \wedge y \leq x \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & x+y=y \wedge y+x=x \\
\equiv & \quad \{ \quad + \text{ is symmetric} \quad \} \\
 & x+y=y \wedge x+y=x \\
\Rightarrow & \quad \{ \quad \text{substitution} \quad \} \\
 & x=y \quad .
\end{array}
$$

We now show that $+$ is monotonic.

$$
\begin{array}{ll}
 & x+y \leq x+z \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & (x+y)+(x+z)=x+z \\
\equiv & \quad \{ \quad \text{arithmetic, } + \text{ is idempotent} \quad \} \\
 & x+y+z=x+z \\
\Leftarrow & \quad \{ \quad \text{Leibniz} \quad \} \\
 & y+z=z \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & y \leq z \quad .
\end{array}
$$

To show that $\cdot$ is monotonic we have to show that it is monotonic in both its arguments. We only show that it is monotonic in its right argument. The other case is symmetrical.

$$
\begin{array}{ll}
 & x \cdot y \leq x \cdot z \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & x \cdot y + x \cdot z = x \cdot z \\
\equiv & \quad \{ \quad \text{distributivity} \quad \} \\
 & x \cdot (y+z) = x \cdot z \\
\Leftarrow & \quad \{ \quad \text{Leibniz} \quad \} \\
 & y+z = z \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & y \leq z \quad .
\end{array}
$$

To show that $+$ is the binary supremum operator we have to show that

$$x+y \leq z \equiv x \leq z \wedge y \leq z \quad .$$

This we do as follows.

$$
\begin{array}{ll}
 & x+y \leq z \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & (x+y)+z = z \\
\equiv & \quad \{ \quad \text{substitution} \quad \} \\
 & (x+y)+z = z \wedge x+z = x+((x+y)+z) \\
\equiv & \quad \{ \quad + \text{ is associative and idempotent} \quad \} \\
 & (x+y)+z = z \wedge x+z = (x+y)+z \\
\Rightarrow & \quad \{ \quad \text{substitution} \quad \} \\
 & x+z = z \\
\equiv & \quad \{ \quad \text{definition of } \leq \quad \} \\
 & x \leq z \quad .
\end{array}
$$

Symmetrically, $x+y \leq z \Rightarrow y \leq z$. Finally, $x \leq z \wedge y \leq z \Rightarrow x+y \leq z$ by the monotonicity of $+$.
□

24 (a)

$$
\begin{array}{ll}
 & a \cdot b^* \leq c^* \cdot a \\
\Leftarrow & \quad \{ \quad (20) \quad \} \\
 & a + c^* \cdot a \cdot b \leq c^* \cdot a \\
\Leftarrow & \quad \{ \quad c^* \cdot a = a + c^* \cdot c \cdot a \text{ , monotonicity} \quad \} \\
 & a \cdot b \leq c \cdot a \quad .
\end{array}
$$

(b)

$$
\begin{array}{ll}
 & c^* \cdot a \leq a \cdot b^* \\
\Leftarrow & \quad \{ \quad (17) \quad \} \\
 & a + c \cdot a \cdot b^* \leq a \cdot b^* \\
\Leftarrow & \quad \{ \quad a \cdot b^* = a + a \cdot b \cdot b^* \text{ , monotonicity} \quad \} \\
 & c \cdot a \leq a \cdot b \quad .
\end{array}
$$

(c) By combining (a) and (b) using the anti-symmetry of the ordering relation we obtain

$$a \cdot b^* = c^* \cdot a \;\Leftarrow\; a \cdot b = c \cdot a \quad .$$

Thus, with $b,c \;:=\; b \cdot a \,,\, a \cdot b$, we have:

$$a \cdot (b \cdot a)^* = (a \cdot b)^* \cdot a \;\Leftarrow\; a \cdot (b \cdot a) = (a \cdot b) \cdot a \quad .$$

The property now follows from the associativity of product.
(d) First we use induction:

$$
\begin{array}{ll}
 & (a+b)^* \leq b^* \cdot (a \cdot b^*)^* \\
\Leftarrow & \quad \{ \quad (17) \quad \} \\
 & 1 + (a+b) \cdot b^* \cdot (a \cdot b^*)^* \leq b^* \cdot (a \cdot b^*)^*
\end{array}
$$

Now we show that $1 + (a+b) \cdot b^* \cdot (a \cdot b^*)^* \;=\; b^* \cdot (a \cdot b^*)^*$.

$$
\begin{array}{ll}
 & 1 + (a+b) \cdot b^* \cdot (a \cdot b^*)^* \\
= & \quad \{ \quad \text{arithmetic} \quad \} \\
 & 1 + a \cdot b^* \cdot (a \cdot b^*)^* + b \cdot b^* \cdot (a \cdot b^*)^* \\
= & \quad \{ \quad 1 + x \cdot x^* = x^* \;\text{ with }\; x \;:=\; a \cdot b^* \quad \} \\
 & (a \cdot b^*)^* + b \cdot b^* \cdot (a \cdot b^*)^*
\end{array}
$$

$$
\begin{array}{ll}
= & \{ \quad \text{arithmetic} \quad \} \\
& (1 + b \cdot b^*) \cdot (a \cdot b^*)^* \\
= & \{ \quad 1 + x \cdot x^* = x^* \ \text{ with } \ x := b \quad \} \\
& b^* \cdot (a \cdot b^*)^* \quad .
\end{array}
$$

The opposite inequality is proved as follows:

$$
\begin{array}{ll}
& b^* \cdot (a \cdot b^*)^* \\
\leq & \{ \quad \text{monotonicity} \quad \} \\
& (a+b)^* \cdot ((a+b) \cdot (a+b)^*)^* \\
\leq & \{ \quad x \cdot x^* \leq x^* \ \text{ with } \ x := a+b \,, \ \text{monotonicity} \quad \} \\
& (a+b)^* \cdot ((a+b)^*)^* \\
= & \{ \quad x^* = x^* \cdot x^* = (x^*)^* \ \text{ with } \ x := a+b \quad \} \\
& (a+b)^* \quad .
\end{array}
$$

Finally, $b^* \cdot (a \cdot b^*)^* = (b^* \cdot a)^* \cdot b^*$ is an instance of (c).
(e)

$$
\begin{array}{ll}
& (a^*)^* = a^* \\
\equiv & \{ \quad \text{anti-symmetry} \quad \} \\
& (a^*)^* \leq a^* \wedge a^* \leq (a^*)^* \\
\Leftarrow & \{ \quad {}^* \text{ is a closure operator and } {}^* \text{ is monotonic} \quad \} \\
& a^* \leq a^* \wedge a \leq a^* \\
\equiv & \{ \quad \text{reflexivity and } \ a^* = 1 + a \cdot a^* = 1 + a + a \cdot a \cdot a^* \quad \} \\
& \mathsf{true} \quad .
\end{array}
$$

□

25 (a) $a^+ = \{\text{definition}\}\ a \cdot a^* \geq \{a^* \geq 1\}\ a \cdot 1 = a$.
(b) This is an application of the mirror rule, 24(c), with b instantiated to 1 .
(c)

$$
\begin{array}{ll}
& a^+ \cdot a^+ \\
= & \{ \quad \text{definition, (b)} \quad \} \\
& a \cdot a^* \cdot a^* \cdot a \\
= & \{ \quad a^* = a^* \cdot a^* \quad \}
\end{array}
$$

$$
\begin{array}{ll}
 & a \cdot a^* \cdot a \\
\leq & \quad \{ \quad a^* = 1 + a^* \cdot a \quad \} \\
 & a \cdot a^* \\
= & \quad \{ \quad \text{definition} \quad \} \\
 & a^+ \ .
\end{array}
$$

(d)

$$
\begin{array}{ll}
 & a + x \cdot x \leq x \\
\Rightarrow & \quad \{ \quad \text{induction (20), with } a,b,x := x,a,x \quad \} \\
 & a \cdot x^* \leq x \\
\equiv & \quad \{ \quad x^* = 1 + x \cdot x^* \text{, arithmetic} \quad \} \\
 & a + a \cdot x \cdot x^* \leq x \\
\Rightarrow & \quad \{ \quad 1 \leq x^* \text{, addition and multiplication are monotonic} \quad \} \\
 & a + a \cdot x \leq x \\
\Rightarrow & \quad \{ \quad \text{induction (17), with } a,b,x := a,a,x \quad \} \\
 & a^* \cdot a \leq x \\
\equiv & \quad \{ \quad \text{(a)} \quad \} \\
 & a^+ \leq x \ .
\end{array}
$$

(e) First, we have that a^* is a prefix point of the function $\langle x :: 1 + a + x \cdot x \rangle$ since $1 \leq a^*$, $a \leq a^*$ and $a^* \cdot a^* \leq a^*$. We now show that it is the least prefix point:

$$
\begin{array}{ll}
 & 1 + a + x \cdot x \leq x \\
\equiv & \quad \{ \quad \text{definition of suprema} \quad \} \\
 & 1 \leq x \wedge a + x \cdot x \leq x \\
\Rightarrow & \quad \{ \quad \text{(d)} \quad \} \\
 & 1 \leq x \wedge a^+ \leq x \\
\equiv & \quad \{ \quad \text{definition of suprema} \quad \} \\
 & 1 + a^+ \leq x \\
\equiv & \quad \{ \quad a^* = 1 + a \cdot a^* = 1 + a^+ \quad \} \\
 & a^* \leq x \ .
\end{array}
$$

□

30 Rolling rule: First,

$$
\begin{array}{ll}
 & \mu(g \bullet f) \preceq g.\mu(f \bullet g) \\
\Leftarrow & \quad \{ \quad \text{induction} \quad \} \\
 & (g \bullet f).(g.\mu(f \bullet g)) \preceq g.\mu(f \bullet g) \\
\equiv & \quad \{ \quad \text{definition of } (g \bullet f) \quad \} \\
 & g.(f.(g.\mu(f \bullet g))) \preceq g.\mu(f \bullet g) \\
\Leftarrow & \quad \{ \quad g \text{ is monotonic} \quad \} \\
 & f.(g.\mu(f \bullet g)) \sqsubseteq \mu(f \bullet g) \\
\equiv & \quad \{ \quad \text{definition of } (f \bullet g) \quad \} \\
 & (f \bullet g).\mu(f \bullet g) \sqsubseteq \mu(f \bullet g) \\
\equiv & \quad \{ \quad \text{computation rule} \quad \} \\
 & \text{true} \quad .
\end{array}
$$

Second, we suppose that y is an arbitrary prefix point of $g \bullet f$ and calculate as follows:

$$
\begin{array}{ll}
 & g.\mu(f \bullet g) \preceq y \\
\Leftarrow & \quad \{ \quad \bullet \quad (g \bullet f).y \preceq y \quad \text{, transitivity} \quad \} \\
 & g.\mu(f \bullet g) \preceq (g \bullet f).y \\
\Leftarrow & \quad \{ \quad \text{definition of } g \bullet f \text{, monotonicity of } g \quad \} \\
 & \mu(f \bullet g) \sqsubseteq f.y \\
\Leftarrow & \quad \{ \quad \text{induction} \quad \} \\
 & (f \bullet g).(f.y) \sqsubseteq f.y \\
\Leftarrow & \quad \{ \quad (f \bullet g).(f.y) = f.((g \bullet f).y) \text{, monotonicity of } f \quad \} \\
 & (g \bullet f).y \preceq y \\
\equiv & \quad \{ \quad y \text{ is a prefix point of } g \bullet f \quad \} \\
 & \text{true} \quad .
\end{array}
$$

Square rule:

$$
\begin{array}{ll}
 & \mu f = \mu(f^2) \\
\equiv & \quad \{ \quad \mathbf{T} \quad \}
\end{array}
$$

$$
\begin{array}{ll}
 & \mu f \sqsubseteq \mu(f^2) \wedge \mu(f^2) \sqsubseteq \mu f \\
\Leftarrow & \quad \{ \quad \text{induction} \quad \} \\
 & f.\mu(f^2) \sqsubseteq \mu(f^2) \wedge f^2.\mu f \sqsubseteq \mu f \\
\equiv & \quad \{ \quad \text{rolling rule and computation rule (applied twice)} \quad \} \\
 & \mathsf{true} \quad .
\end{array}
$$

Diagonal rule: First suppose that $\mu\langle y :: x \oplus y\rangle$ exists, for all x, as does $\mu\langle x :: \mu\langle y :: x \oplus y\rangle\rangle$. We prove that the latter satisfies the definition of the least prefix point of the function $\langle x :: x \oplus x\rangle$.

For brevity let us denote $\mu\langle y :: x \oplus y\rangle$ by $m.x$ and $\mu\langle x :: \mu\langle y :: x \oplus y\rangle\rangle$ by M.
First M is a fixed point of the function:

$$
\begin{array}{ll}
 & M = M \oplus M \\
\equiv & \quad \{ \quad \text{fixed point: } M = m.M = M \oplus m.M \quad \} \\
 & M \oplus m.M = M \oplus M \\
\equiv & \quad \{ \quad \text{fixed point: } M = m.M \quad \} \\
 & \mathsf{true} \quad .
\end{array}
$$

Second, M is at most every prefix point of the function:

$$
\begin{array}{ll}
 & M \sqsubseteq x \\
\Leftarrow & \quad \{ \quad \text{induction: } M = \mu m \quad \} \\
 & m.x \sqsubseteq x \\
\Leftarrow & \quad \{ \quad \text{induction: } m.x = \mu\langle y :: x \oplus y\rangle \quad \} \\
 & x \oplus x \sqsubseteq x \quad .
\end{array}
$$

Now, with the same abbreviations as above, we assume that $m.x$ exists for all x. We also introduce the abbreviation N for $\mu\langle x :: x \oplus x\rangle$ which we assume to exist. We have to show that N is the least prefix point of the function m.
First it is a prefix point of m.

$$
\begin{array}{ll}
 & m.N \sqsubseteq N \\
\Leftarrow & \quad \{ \quad \text{induction: } m.N = \mu\langle y :: N \oplus y\rangle \quad \} \\
 & N \oplus N \sqsubseteq N \\
\equiv & \quad \{ \quad \text{prefix point: } N = \mu\langle x :: x \oplus x\rangle \quad \} \\
 & \mathsf{true} \quad .
\end{array}
$$

Second, N is at most every prefix point of m:

$$
\begin{array}{ll}
 & N \sqsubseteq x \\
\Leftarrow & \quad \{ \quad \text{induction: } N = \mu\langle x :: x \oplus x \rangle \quad \} \\
 & x \oplus x \sqsubseteq x \\
\Leftarrow & \quad \{ \quad \text{transitivity} \quad \} \\
 & x \oplus x \sqsubseteq x \oplus m.x \sqsubseteq m.x \sqsubseteq x \\
\equiv & \quad \{ \quad \text{fixed point: } m.x = x \oplus m.x \quad \} \\
 & x \oplus x \sqsubseteq x \oplus m.x \wedge m.x \sqsubseteq x \\
\equiv & \quad \{ \quad \oplus \text{ is monotonic} \quad \} \\
 & m.x \sqsubseteq x \quad .
\end{array}
$$

□

31

$$
\begin{array}{ll}
 & \mu\langle X :: 1 + a \cdot X \cdot b \cdot X + b \cdot X \cdot a \cdot X \rangle \\
= & \quad \{ \quad \text{diagonal rule (29)} \quad \} \\
 & \mu\langle X :: \mu\langle Y :: 1 + a \cdot X \cdot b \cdot Y + b \cdot X \cdot a \cdot Y \rangle \rangle \\
= & \quad \{ \quad \text{definition of } {}^* \quad \} \\
 & \mu\langle X :: (a \cdot X \cdot b + b \cdot X \cdot a)^* \rangle \\
= & \quad \{ \quad \text{exercise 25(e)} \quad \} \\
 & \mu\langle X :: \mu\langle Y :: 1 + a \cdot X \cdot b + a \cdot X \cdot b + Y \cdot Y \rangle \rangle \\
= & \quad \{ \quad \text{diagonal rule (29)} \quad \} \\
 & \mu\langle X :: 1 + a \cdot X \cdot b + a \cdot X \cdot b + X \cdot X \rangle \quad .
\end{array}
$$

□

34 (a) We have:

$$
\begin{array}{ll}
 & \mu(g \bullet f) \sqsubseteq \mu(h \bullet f) \\
\Leftarrow & \quad \{ \quad \text{induction} \quad \} \\
 & (g \bullet f) \,.\, \mu(h \bullet f) \sqsubseteq \mu(h \bullet f) \\
\equiv & \quad \{ \quad \text{composition and rolling rule} \quad \} \\
 & g \,.\, \mu(f \bullet h) \sqsubseteq \mu(h \bullet f)
\end{array}
$$

$\Leftarrow$ { μ-fusion, g is a lower adjoint }

$g \bullet f \bullet h \mathrel{\dot{\sqsubseteq}} h \bullet f \bullet g$.

Further,

$\mu(f \bullet g) \sqsubseteq \mu(f \bullet h)$

$\equiv$ { rolling rule }

$f \,.\, \mu(g \bullet f) \sqsubseteq f \,.\, \mu(h \bullet f)$

$\Leftarrow$ { monotonicity of f }

$\mu(g \bullet f) \sqsubseteq \mu(h \bullet f)$

$\Leftarrow$ { above }

$g \bullet f \bullet h \mathrel{\dot{\sqsubseteq}} h \bullet f \bullet g$.

(b) This part is a straightforward corollary of (a) and the antisymmetry of the ordering relations.

As remarked above, the exchange rule says that g and h can be exchanged within a μ term in which they are composed after or before an arbitrary function f. The generalisation we seek is that they can be exchanged anywhere in a μ term. That is, we seek a condition such that

$$\mu(f_0 \bullet g \bullet f_1) = \mu(f_0 \bullet h \bullet f_1) \quad .$$

We derive such a condition as follows:

$\mu(f_0 \bullet g \bullet f_1) = \mu(f_0 \bullet h \bullet f_1)$

$\equiv$ { rolling rule }

$f_0 \,.\, \mu(g \bullet f_1 \bullet f_0) = f_0 \,.\, \mu(h \bullet f_1 \bullet f_0)$

$\Leftarrow$ { exchange rule }

$g \bullet f_1 \bullet f_0 \bullet h = h \bullet f_1 \bullet f_0 \bullet g$.

We don't advise you to remember this rule. It is overly complicated and is so easily recovered from the basic rule.

□

37 In the case of reachability, graph b is acyclic. In the case of shortest paths, the sum of the edge lengths on any cycle of edges is greater than zero. In the case of bottlenecks, there is no cycle of edges each of weight ∞. That is, there is no subset of nodes connected by a cycle of edges with each edge offering no bottleneck.

□

39

$$
\begin{array}{ll}
 & \mu\langle g::\ \langle a::\ a \oplus g.a\rangle\rangle \\
= & \quad \{ \quad \text{definition of } \oplus \quad \} \\
 & \mu\langle g::\ \langle a::\ f.(g.a)\rangle\rangle \\
= & \quad \{ \quad \text{definition of } (f\bullet) \quad \} \\
 & \mu(f\bullet) \\
= & \quad \{ \quad \text{abstraction rule} \quad \} \\
 & \langle a::\ \mu\langle x::\ a \oplus x\rangle\rangle \\
= & \quad \{ \quad \text{definition of } \oplus \quad \} \\
 & \langle a::\ \mu\langle x::\ f.x\rangle\rangle \\
= & \quad \{ \quad \eta\text{-reduction} \quad \} \\
 & \langle a::\ \mu f\rangle
\end{array}
$$

Thus the rule is: $\mu(f\bullet) = K.\mu f$ where K is the constant combinator.
□

44 X is the relation on pairs (w,k)

$$w \in L \ \equiv \ w' = \varepsilon \wedge k' = 0$$

where w is a word, k is a natural number and L is the language given in the statement of the problem. The invariant is the relation

$$w \in L \ \equiv \ w' \in L^{k'} \quad ,$$

the initialisation is the relation

$$w = w' \wedge k' = 1 \quad ,$$

the loop body is the relation

$$k \neq 0 \ \wedge \ (\exists\langle v::\ w = av \wedge w' = v \wedge k' = k-1\rangle \vee \exists\langle v::\ w = bv \wedge w' = v \wedge k' = k+1\rangle) \quad ,$$

and the termination is the relation

$$w = w' = \varepsilon \ \wedge \ k' = k = 0 \quad .$$

□

55 Let p denote $\nu\langle X :: X \circ R\rangle$. Then p is characterised by the two properties

(a) $\quad p = p \circ R$, and

(b) $\quad \forall\langle q:\ q = \top\top \circ q:\ q \subseteq p \Leftarrow q = q \circ R\rangle$.

Interpreting p as a set, (a) is the property that $x \in p$ equivales $\exists\langle y: y \in p: y [\![R]\!] x\rangle$ and (b) is the property that p is the largest such set. In words, p is the largest set of elements such that each element begins an infinite chain of R-related elements.
□

56 By theorem 51 we have to show that the greatest fixed point of $\langle X :: X \circ f\cup \circ R \circ f\rangle$ equals $\bot\!\!\!\bot$. We do this by first rewriting $\nu\langle X :: X \circ f\cup \circ R \circ f\rangle$ in terms of $\nu\langle X :: X \circ R\rangle$. We have:

$$
\begin{array}{ll}
 & \nu\langle X :: X \circ f\cup \circ R \circ f\rangle \\
= & \quad \{ \quad \text{rolling rule} \quad \} \\
 & \nu\langle X :: X \circ f \circ f\cup \circ R\rangle \circ f \\
\subseteq & \quad \{ \quad f \circ f\cup \subseteq I \text{, by definition of functional; } \nu \text{ is monotonic} \quad \} \\
 & \nu\langle X :: X \circ R\rangle \circ f \quad .
\end{array}
$$

The more general statement is thus

$$\nu\langle X :: X \circ f\cup \circ R \circ f\rangle \ \subseteq\ \nu\langle X :: X \circ R\rangle \circ f \quad .$$

Making use of it, we have:

$$
\begin{array}{ll}
 & \nu\langle X :: X \circ f\cup \circ R \circ f\rangle \subseteq \bot\!\!\!\bot \\
\Leftarrow & \quad \{ \quad \text{transitivity} \quad \} \\
 & \nu\langle X :: X \circ R\rangle \circ f \subseteq \bot\!\!\!\bot \\
\Leftarrow & \quad \{ \quad \bot\!\!\!\bot \text{ is zero of composition and composition is monotonic} \quad \} \\
 & \nu\langle X :: X \circ R\rangle \subseteq \bot\!\!\!\bot \quad .
\end{array}
$$

□

57 It suffices to show that if R is well-founded then $R \cap I \subseteq \bot\!\!\!\bot$ because R is well-founded equivales R^+ is well-founded.

$$
\begin{array}{ll}
 & R \cap I \subseteq \bot\!\!\!\bot \\
\Leftarrow & \quad \{ \quad \text{assumption, } R \text{ is well-founded} \quad \} \\
 & R \cap I \subseteq (R \cap I) \circ R
\end{array}
$$

$\Leftarrow$ { $R \cap I \subseteq R$, monotonicity and transitivity }

$R \cap I \subseteq (R \cap I) \circ (R \cap I)$

$\equiv$ { for all coreflexives A, $A = A \circ A$.

$R \cap I$ is a coreflexive }

true .

Well-foundedness of R is equivalent to $R^+ \cap I \subseteq \bot\!\!\!\bot$ when R is a relation on a finite set. (This is proved by induction on the size of the set.) When R is a relation on an infinite set the two conditions are not equivalent. For example, the less-than ordering on integers is not well-founded but its intersection with the identity relation is empty. □

62 We use 60 as definition of admitting induction: with A denoting $\mu((R \circ \mu(R \backslash\!\!\cdot)) \backslash\!\!\cdot)$ we must show $I = A$, or equivalently (because A is a coreflexive) $I \subseteq A$. The proof has two phases. In the first phase we reduce the goal to the formally weaker $\mu(R \backslash\!\!\cdot) \subseteq A$.

$I \subseteq A$

$\equiv$ { definition of A, $\mu F = F.\mu F$ }

$I \subseteq (R \circ \mu(R \backslash\!\!\cdot)) \backslash\!\!\cdot A$

$\equiv$ { monotype factors }

$(R \circ \mu(R \backslash\!\!\cdot))^< \subseteq A$

$\equiv$ { $\mu(R \backslash\!\!\cdot)$ is a monotype }

$(R \circ \mu(R \backslash\!\!\cdot) \circ \mu(R \backslash\!\!\cdot))^< \subseteq A$

$\equiv$ { monotype factors; definition of A; $\mu F = F.\mu F$ }

$\mu(R \backslash\!\!\cdot) \subseteq A$

Now, in the second phase, we have the opportunity to apply Knaster-Tarski.

$\mu(R \backslash\!\!\cdot) \subseteq A$

$\Leftarrow$ { Knaster-Tarski }

$R \backslash\!\!\cdot A \subseteq A$

$\equiv$ { definition of A; $\mu F = F.\mu F$ }

$R \backslash\!\!\cdot A \subseteq (R \circ \mu(R \backslash\!\!\cdot)) \backslash\!\!\cdot A$

$\Leftarrow$ { monotype factor is anti-monotone in left argument }

$R \circ \mu(R \backslash\!\!\cdot) \subseteq R$

$\equiv$ { $\mu(R\backslash) \subseteq I$; monotonicity of composition }

true

□

65 We first show that $R\backslash S = \neg(R^{\cup} \circ \neg S)$. We use indirect equality.

$X \subseteq \neg(R^{\cup} \circ \neg S)$

$\equiv$ { complements }

$\neg X \supseteq R^{\cup} \circ \neg S$

$\equiv$ { middle exchange }

$S \supseteq R \circ X$

$\equiv$ { factors }

$X \subseteq R\backslash S$.

Now we construct function f as suggested:

$\nu(\circ R) = f.\mu(R\backslash)$

$\Leftarrow$ { fusion, assuming f is antimonotonic }

$\forall\langle X :: f.X \circ R = f.(R\backslash X)\rangle$

$\equiv$ { $R\backslash S = \neg(R^{\cup} \circ \neg S)$ }

$\forall\langle X :: f.X \circ R = f.\neg(R^{\cup} \circ \neg X)\rangle$

$\Leftarrow$ { try $f.X = g.\neg X$ }

$\forall\langle Y :: g.Y \circ R = g.(R^{\cup} \circ Y)\rangle$

$\Leftarrow$ { converse }

$\forall\langle Y :: g.Y = Y^{\cup}\rangle$.

Thus, $\nu(\circ R) = \neg\mu(R\backslash)^{\cup}$. For the final step, we have

$\nu(\circ R) \subseteq \bot\!\!\!\bot$

$\equiv$ { above }

$\neg\mu(R\backslash)^{\cup} \subseteq \bot\!\!\!\bot$

$\equiv$ { complement, converse }

$\top\!\!\!\top^{\cup} \subseteq \mu(R\backslash)$

$\equiv$ { converse }

$\top\!\!\!\top \subseteq \mu(R\backslash)$.

The interpretation of $\mu(R\backslash)$ is the set of all points such that all chains starting in such a point are guaranteed to be of finite length.
□

95 By instantiating S to the identity relation, it is clear that the follows-from property is true. We only need to prove the implication. So assume that R is well-founded. Then, for all X and S, we have:

$$
\begin{array}{ll}
 & X = S \circ X \circ R \\
\Rightarrow & \quad \{ \quad \text{Leibniz} \quad \} \\
 & \top\!\!\top \circ X = \top\!\!\top \circ S \circ X \circ R \\
\Rightarrow & \quad \{ \quad \top\!\!\top \circ S \subseteq \top\!\!\top \quad \} \\
 & \top\!\!\top \circ X \subseteq \top\!\!\top \circ X \circ R \\
\Rightarrow & \quad \{ \quad \text{assumption: } R \text{ is well-founded. So } \nu\langle X :: X \circ R \rangle = \bot\!\!\bot \quad \} \\
 & \top\!\!\top \circ X \subseteq \bot\!\!\bot \\
\Rightarrow & \quad \{ \quad \bot\!\!\bot \subseteq X \subseteq \top\!\!\top \circ X \quad \} \\
 & X = \bot\!\!\bot \\
\equiv & \quad \{ \quad \bot\!\!\bot = S \circ \bot\!\!\bot \circ R \text{ and } \bot\!\!\bot \subseteq Y \text{ for all } Y \quad \} \\
 & X = \mu\langle Y :: S \circ Y \circ R \rangle \\
\Rightarrow & \quad \{ \quad \text{computation rule} \quad \} \\
 & X = S \circ X \circ R \quad .
\end{array}
$$

□

96 Suppose $R \in F.I \leftarrow I$ is F-reductive. Suppose $S \in J \leftarrow F.J$ and that $X \in J \leftarrow I$ satisfies

$$X = S \circ F.X \circ R \quad .$$

We have to show that $X = \mu\langle Y :: S \circ F.Y \circ R \rangle$.

$$
\begin{array}{ll}
 & X = \mu\langle Y :: S \circ F.Y \circ R \rangle \\
\equiv & \quad \{ \quad X = S \circ F.X \circ R \text{, definition of } \mu \quad \} \\
 & X \subseteq \mu\langle Y :: S \circ F.Y \circ R \rangle \\
\equiv & \quad \{ \quad R \in F.I \leftarrow I \text{ is } F\text{-reductive. So } \mu((R\backslash_{\bullet})\bullet F) = \mathrm{id}_I \quad \} \\
 & X \circ \mu((R\backslash_{\bullet})\bullet F) \subseteq \mu\langle Y :: S \circ F.Y \circ R \rangle \\
\Leftarrow & \quad \{ \quad \text{fusion} \quad \}
\end{array}
$$

$$
\begin{array}{ll}
 & \forall\langle A:\ A \subseteq I:\ X \circ R \backslash_{\bullet} F.A \subseteq S \circ F.(X \circ A) \circ R\rangle \\
\equiv & \quad \{ \quad \text{assumption: } X = S \circ F.X \circ R\,, \\
 & \qquad\quad F \text{ distributes through composition} \quad \} \\
 & \forall\langle A:\ A \subseteq I:\ S \circ F.X \circ R \circ R \backslash_{\bullet} F.A \subseteq S \circ F.X \circ F.A \circ R\rangle \\
\Leftarrow & \quad \{ \quad \text{monotonicity} \quad \} \\
 & \forall\langle A:\ A \subseteq I:\ R \circ R \backslash_{\bullet} F.A \subseteq F.A \circ R\rangle \\
\equiv & \quad \{ \quad \text{factors: (47)} \quad \} \\
 & \text{true} \quad .
\end{array}
$$

□

References

[Bac86] R.C. Backhouse. *Program Construction and Verification.* Prentice-Hall International, 1986.

[BBH+92] R.C. Backhouse, P. de Bruin, P. Hoogendijk, G. Malcolm, T.S. Voermans, and J. van der Woude. Polynomial relators. In M. Nivat, C.S. Rattray, T. Rus, and G. Scollo, editors, *Proceedings of the 2nd Conference on Algebraic Methodology and Software Technology, AMAST'91*, pages 303–326. Springer-Verlag, Workshops in Computing, 1992.

[BBM+91] R.C. Backhouse, P. de Bruin, G. Malcolm, T.S. Voermans, and J. van der Woude. Relational catamorphisms. In Möller B., editor, *Proceedings of the IFIP TC2/WG2.1 Working Conference on Constructing Programs from Specifications*, pages 287–318. Elsevier Science Publishers B.V., 1991.

[BC75] R.C. Backhouse and B.A. Carré. Regular algebra applied to path-finding problems. *Journal of the Institute of Mathematics and its Applications*, 15:161–186, 1975.

[BH99] Roland Backhouse and Paul Hoogendijk. Final dialgebras: From categories to allegories. Accepted for publication in Theoretical Informatics and Applications. Extended abstract presented at Workshop on Fixed Points in Computer Science, Brno, August 1998, 1999.

[BJJM99] Roland Backhouse, Patrik Jansson, Johan Jeuring, and Lambert Meertens. Generic programming. An introduction. In S.D. Swierstra, editor, *3rd International Summer School on Advanced Functional Programming, Braga, Portugal, 12th-19th September, 1998*, volume LNCS 1608, pages 28–115. Springer Verlag, 1999.

[BW93] R.C. Backhouse and J. van der Woude. Demonic operators and monotype factors. *Mathematical Structures in Computer Science*, 3(4):417–433, December 1993.

[DB95] Henk Doornbos and Roland Backhouse. Induction and recursion on datatypes. In B. Möller, editor, *Mathematics of Program Construction, 3rd International Conference*, volume 947 of *LNCS*, pages 242–256. Springer-Verlag, July 1995.

[DB96] Henk Doornbos and Roland Backhouse. Reductivity. *Science of Computer Programming*, 26(1–3):217–236, 1996.

[DBvdW97] H. Doornbos, R.C. Backhouse, and J. van der Woude. A calculation approach to mathematical induction. *Theoretical Computer Science*, (179):103–135, 1997.

[Doo96] H. Doornbos. *Reductivity arguments and program construction.* PhD thesis, Eindhoven University of Technology, Department of Mathematics and Computing Science, June 1996.

[DS90] E.W. Dijkstra and C.S. Scholten. *Predicate Calculus and Program Semantics.* Springer-Verlag, Berlin, 1990.

[Fok92] Maarten M. Fokkinga. *Law and Order in Algorithmics.* PhD thesis, Universiteit Twente, The Netherlands, 1992.

[Fv90] P.J. Freyd and A. Ščedrov. *Categories, Allegories.* North-Holland, 1990.

[Gri81] D. Gries. *The Science of Programming.* Springer-Verlag, New York, 1981.

[HdM00] Paul Hoogendijk and Oege de Moor. What is a container type? *Journal of Functional Programming*, 2000. to appear.

[HH86] C.A.R. Hoare and Jifeng He. The weakest prespecification. *Fundamenta Informaticae*, 9:51–84, 217–252, 1986.

[Hoo97] Paul Hoogendijk. *A Generic Theory of Datatypes.* PhD thesis, Department of Mathematics and Computing Science, Eindhoven University of Technology, 1997.

[Kle56] S.C. Kleene. Representation of events in nerve nets and finite automata. In Shannon and McCarthy, editors, *Automata Studies*, pages 3–41. Princeton Univ. Press, 1956.

[Koz94] Dexter Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. *Information and Computation*, 110(2):366–390, 1994.

[Lam68] J. Lambek. A fixpoint theorem for complete categories. *Mathematische Zeitschrift*, 103:151–161, 1968.

[Mal90a] G. Malcolm. *Algebraic data types and program transformation.* PhD thesis, Groningen University, 1990.

[Mal90b] G. Malcolm. Data structures and program transformation. *Science of Computer Programming*, 14(2–3):255–280, October 1990.

[Mee92] L. Meertens. Paramorphisms. *Formal Aspects of Computing*, 4(5):413–424, 1992.

[Nau63] P. (Ed.) Naur. Revised report on the algorithmic language ALGOL 60. *Comm. ACM*, 6:1–20, Also in *The Computer Journal*, 5: 349–67 (1963); *Numerische Mathematik*, 4: 420–52 (1963) 1963.

[SdM93] Doaitse Swierstra and Oege de Moor. Virtual data structures. In Helmut Partsch, Bernhard Möller, and Steve Schuman, editors, *Formal Program Development*, volume 755 of *LNCS*, pages 355–371. Springer-Verlag, 1993.