

The *inductive* and the *coinductive* types are two important forms of recursive type. Inductive types correspond to *least*, or *initial*, solutions of certain type equations, and coinductive types correspond to their *greatest*, or *final*, solutions. Intuitively, the elements of an inductive type are those that are given by a finite composition of its introduction forms. Consequently, if we specify the behavior of a function on each of the introduction forms of an inductive type, then its behavior is defined for all values of that type. Such a function is a *recursor*, or *catamorphism*. Dually, the elements of a coinductive type are those that behave properly in response to a finite composition of its elimination forms. Consequently, if we specify the behavior of an element on each elimination form, then we have fully specified a value of that type. Such an element is a *generator*, or *anamorphism*.

15.1 Motivating Examples

The most important example of an inductive type is the type of natural numbers as formalized in Chapter 9. The type nat is the *least* type containing \mathbf{z} and closed under $\mathbf{s}(-)$. The minimality condition is expressed by the existence of the iterator, $\text{iter } e \{z \hookrightarrow e_0 \mid \mathbf{s}(x) \hookrightarrow e_1\}$, which transforms a natural number into a value of type τ , given its value for zero, and a transformation from its value on a number to its value on the successor of that number. This operation is well-defined precisely because there are no other natural numbers.

With a view towards deriving the type nat as a special case of an inductive type, it is useful to combine zero and successor into a single introduction form, and to correspondingly combine the basis and inductive step of the iterator. The following rules specify the statics of this reformulation:

$$\frac{\Gamma \vdash e : \text{unit} + \text{nat}}{\Gamma \vdash \text{fold}_{\text{nat}}(e) : \text{nat}} \quad (15.1a)$$

$$\frac{\Gamma, x : \text{unit} + \tau \vdash e_1 : \tau \quad \Gamma \vdash e_2 : \text{nat}}{\Gamma \vdash \text{rec}_{\text{nat}}(x.e_1; e_2) : \tau} \quad (15.1b)$$

The expression $\text{fold}_{\text{nat}}(e)$ is the unique introduction form of the type nat . Using this, the expression \mathbf{z} is $\text{fold}_{\text{nat}}(\mathbf{1} \cdot \langle \rangle)$, and $\mathbf{s}(e)$ is $\text{fold}_{\text{nat}}(\mathbf{r} \cdot e)$. The recursor, $\text{rec}_{\text{nat}}(x.e_1; e_2)$, takes as argument the abstractor $x.e_1$ that combines the basis and inductive step into a single computation that, given a value of type $\text{unit} + \tau$, yields a value of type τ . Intuitively, if

x is replaced by the value $1 \cdot \langle \rangle$, then e_1 computes the base case of the recursion, and if x is replaced by the value $r \cdot e$, then e_1 computes the inductive step from the result e of the recursive call.

The dynamics of the combined representation of natural numbers is given by the following rules:

$$\frac{}{\text{fold}_{\text{nat}}(e) \text{ val}} \quad (15.2a)$$

$$\frac{e_2 \mapsto e'_2}{\text{rec}_{\text{nat}}(x.e_1; e_2) \mapsto \text{rec}_{\text{nat}}(x.e_1; e'_2)} \quad (15.2b)$$

$$\frac{}{\text{rec}_{\text{nat}}(x.e_1; \text{fold}_{\text{nat}}(e_2)) \mapsto [\text{map}\{t.\text{unit} + t\}(y.\text{rec}_{\text{nat}}(x.e_1; y))(e_2)/x]e_1} \quad (15.2c)$$

Rule (15.2c) uses (polynomial) generic extension (see Chapter 14) to apply the recursor to the predecessor, if any, of a natural number. If we expand the definition of the generic extension in place, we obtain this rule:

$$\frac{}{\text{rec}_{\text{nat}}(x.e_1; \text{fold}_{\text{nat}}(e_2)) \mapsto [\text{case } e_2 \{1 \cdot _ \hookrightarrow 1 \cdot \langle \rangle \mid r \cdot y \hookrightarrow r \cdot \text{rec}_{\text{nat}}(x.e_1; y)\}/x]e_1}$$

Exercise **15.2** asks for a derivation of the iterator, as defined in Chapter 9, from the recursor just given.

An illustrative example of a coinductive type is the type of *streams* of natural numbers. A stream is an infinite sequence of natural numbers such that an element of the stream can be computed only after computing all preceding elements in that stream. That is, the computations of successive elements of the stream are sequentially dependent in that the computation of one element influences the computation of the next. In this sense, the introduction form for streams is dual to the elimination form for natural numbers.

A stream is given by its behavior under the elimination forms for the stream type: $\text{hd}(e)$ returns the next, or head, element of the stream, and $\text{tl}(e)$ returns the tail of the stream, the stream resulting when the head element is removed. A stream is introduced by a *generator*, the dual of a recursor, that defines the head and the tail of the stream in terms of the current state of the stream, which is represented by a value of some type. The statics of streams is given by the following rules:

$$\frac{\Gamma \vdash e : \text{stream}}{\Gamma \vdash \text{hd}(e) : \text{nat}} \quad (15.3a)$$

$$\frac{\Gamma \vdash e : \text{stream}}{\Gamma \vdash \text{tl}(e) : \text{stream}} \quad (15.3b)$$

$$\frac{\Gamma \vdash e : \tau \quad \Gamma, x : \tau \vdash e_1 : \text{nat} \quad \Gamma, x : \tau \vdash e_2 : \tau}{\Gamma \vdash \text{strgen } x \text{ is } e \text{ in } \langle \text{hd} \hookrightarrow e_1, \text{tl} \hookrightarrow e_2 \rangle : \text{stream}} \quad (15.3c)$$

In rule (15.3c), the current state of the stream is given by the expression e of some type τ , and the head and tail of the stream are determined by the expressions e_1 and e_2 , respectively, as a function of the current state. (The notation for the generator is chosen to emphasize that every stream has both a head *and* a tail.)

The dynamics of streams is given by the following rules:

$$\frac{}{\text{strgen } x \text{ is } e \text{ in } \langle \text{hd} \hookrightarrow e_1, \text{tl} \hookrightarrow e_2 \rangle \text{ val}} \quad (15.4a)$$

$$\frac{e \mapsto e'}{\text{hd}(e) \mapsto \text{hd}(e')} \quad (15.4b)$$

$$\frac{}{\text{hd}(\text{strgen } x \text{ is } e \text{ in } \langle \text{hd} \hookrightarrow e_1, \text{tl} \hookrightarrow e_2 \rangle) \mapsto [e/x]e_1} \quad (15.4c)$$

$$\frac{e \mapsto e'}{\text{tl}(e) \mapsto \text{tl}(e')} \quad (15.4d)$$

$$\frac{}{\text{tl}(\text{strgen } x \text{ is } e \text{ in } \langle \text{hd} \hookrightarrow e_1, \text{tl} \hookrightarrow e_2 \rangle) \mapsto \text{strgen } x \text{ is } [e/x]e_2 \text{ in } \langle \text{hd} \hookrightarrow e_1, \text{tl} \hookrightarrow e_2 \rangle} \quad (15.4e)$$

Rules (15.4c) and (15.4e) express the dependency of the head and tail of the stream on its current state. Observe that the tail is obtained by applying the generator to the new state determined by e_2 from the current state.

To derive streams as a special case of a coinductive type, we combine the head and the tail into a single elimination form, and reorganize the generator correspondingly. Thus, we consider the following statics:

$$\frac{\Gamma \vdash e : \text{stream}}{\Gamma \vdash \text{unfold}_{\text{stream}}(e) : \text{nat} \times \text{stream}} \quad (15.5a)$$

$$\frac{\Gamma, x : \tau \vdash e_1 : \text{nat} \times \tau \quad \Gamma \vdash e_2 : \tau}{\Gamma \vdash \text{gen}_{\text{stream}}(x.e_1; e_2) : \text{stream}} \quad (15.5b)$$

Rule (15.5a) states that a stream may be unfolded into a pair consisting of its head, a natural number, and its tail, another stream. The head $\text{hd}(e)$ and tail $\text{tl}(e)$ of a stream e are the projections $\text{unfold}_{\text{stream}}(e) \cdot \text{l}$ and $\text{unfold}_{\text{stream}}(e) \cdot \text{r}$, respectively. Rule (15.5b) states that a stream is generated from the state element e_2 by an expression e_1 that yields the head element and the next state as a function of the current state.

The dynamics of streams is given by the following rules:

$$\frac{}{\text{gen}_{\text{stream}}(x.e_1; e_2) \text{ val}} \quad (15.6a)$$

$$\frac{e \mapsto e'}{\text{unfold}_{\text{stream}}(e) \mapsto \text{unfold}_{\text{stream}}(e')} \tag{15.6b}$$

$$\frac{\text{unfold}_{\text{stream}}(\text{gen}_{\text{stream}}(x.e_1; e_2))}{\mapsto} \tag{15.6c}$$

$$\text{map}\{t.\text{nat} \times t\}(y.\text{gen}_{\text{stream}}(x.e_1; y))([e_2/x]e_1)$$

Rule (15.6c) uses generic extension to generate a new stream whose state is the second component of $[e_2/x]e_1$. Expanding the generic extension we obtain the following reformulation of this rule:

$$\frac{\text{unfold}_{\text{stream}}(\text{gen}_{\text{stream}}(x.e_1; e_2))}{\mapsto} \langle ([e_2/x]e_1) \cdot 1, \text{gen}_{\text{stream}}(x.e_1; ([e_2/x]e_1) \cdot r) \rangle$$

Exercise 15.3 asks for a derivation of $\text{strgen } x \text{ is } e \text{ in } \langle \text{hd} \leftrightarrow e_1, \text{tl} \leftrightarrow e_2 \rangle$ from the coinductive generation form.

15.2 Statics

We may now give a general account of inductive and coinductive types, which are defined in terms of positive type operators. We will consider a variant of **T**, which we will call **M**, with natural numbers replaced by functions, products, sums, and a rich class of inductive and coinductive types.

15.2.1 Types

The syntax of inductive and coinductive types involves *type variables*, which are, of course, variables ranging over types. The abstract syntax of inductive and coinductive types is given by the following grammar:

$$\begin{array}{lll} \text{Typ } \tau ::= & t & \text{self-reference} \\ & \text{ind}(t.\tau) & \mu(t.\tau) \text{ inductive} \\ & \text{coi}(t.\tau) & \nu(t.\tau) \text{ coinductive} \end{array}$$

Type formation judgments have the form

$$t_1 \text{ type}, \dots, t_n \text{ type} \vdash \tau \text{ type},$$

where t_1, \dots, t_n are type names. We let Δ range over finite sets of hypotheses of the form t type, where t is a type name. The type formation judgment is inductively defined by the

following rules:

$$\frac{}{\Delta, t \text{ type} \vdash t \text{ type}} \quad (15.7a)$$

$$\frac{}{\Delta \vdash \text{unit type}} \quad (15.7b)$$

$$\frac{\Delta \vdash \tau_1 \text{ type} \quad \Delta \vdash \tau_2 \text{ type}}{\Delta \vdash \text{prod}(\tau_1; \tau_2) \text{ type}} \quad (15.7c)$$

$$\frac{}{\Delta \vdash \text{void type}} \quad (15.7d)$$

$$\frac{\Delta \vdash \tau_1 \text{ type} \quad \Delta \vdash \tau_2 \text{ type}}{\Delta \vdash \text{sum}(\tau_1; \tau_2) \text{ type}} \quad (15.7e)$$

$$\frac{\Delta \vdash \tau_1 \text{ type} \quad \Delta \vdash \tau_2 \text{ type}}{\Delta \vdash \text{arr}(\tau_1; \tau_2) \text{ type}} \quad (15.7f)$$

$$\frac{\Delta, t \text{ type} \vdash \tau \text{ type} \quad \Delta \vdash t.\tau \text{ pos}}{\Delta \vdash \text{ind}(t.\tau) \text{ type}} \quad (15.7g)$$

$$\frac{\Delta, t \text{ type} \vdash \tau \text{ type} \quad \Delta \vdash t.\tau \text{ pos}}{\Delta \vdash \text{coi}(t.\tau) \text{ type}} \quad (15.7h)$$

15.2.2 Expressions

The abstract syntax of **M** is given by the following grammar:

$$\begin{array}{llll} \text{Exp } e ::= & \text{fold}\{t.\tau\}(e) & \text{fold}_{t,\tau}(e) & \text{constructor} \\ & \text{rec}\{t.\tau\}(x.e_1; e_2) & \text{rec}(x.e_1; e_2) & \text{recursor} \\ & \text{unfold}\{t.\tau\}(e) & \text{unfold}_{t,\tau}(e) & \text{destructor} \\ & \text{gen}\{t.\tau\}(x.e_1; e_2) & \text{gen}(x.e_1; e_2) & \text{generator} \end{array}$$

The subscripts on the concrete syntax forms are often omitted when they are clear from context.

The statics for **M** is given by the following typing rules:

$$\frac{\Gamma \vdash e : [\text{ind}(t.\tau)/t]\tau}{\Gamma \vdash \text{fold}\{t.\tau\}(e) : \text{ind}(t.\tau)} \quad (15.8a)$$

$$\frac{\Gamma, x : [\tau'/t]\tau \vdash e_1 : \tau' \quad \Gamma \vdash e_2 : \text{ind}(t.\tau)}{\Gamma \vdash \text{rec}\{t.\tau\}(x.e_1; e_2) : \tau'} \quad (15.8b)$$

$$\frac{\Gamma \vdash e : \text{coi}(t.\tau)}{\Gamma \vdash \text{unfold}\{t.\tau\}(e) : [\text{coi}(t.\tau)/t]\tau} \quad (15.8c)$$

$$\frac{\Gamma \vdash e_2 : \tau_2 \quad \Gamma, x : \tau_2 \vdash e_1 : [\tau_2/t]\tau}{\Gamma \vdash \text{gen}\{t.\tau\}(x.e_1; e_2) : \text{coi}(t.\tau)} \quad (15.8d)$$

15.3 Dynamics

The dynamics of \mathbf{M} is given in terms of the positive generic extension operation described in Chapter 14. The following rules specify a lazy dynamics for \mathbf{M} :

$$\frac{}{\text{fold}\{t.\tau\}(e) \text{ val}} \quad (15.9a)$$

$$\frac{e_2 \mapsto e'_2}{\text{rec}\{t.\tau\}(x.e_1; e_2) \mapsto \text{rec}\{t.\tau\}(x.e_1; e'_2)} \quad (15.9b)$$

$$\frac{}{\text{rec}\{t.\tau\}(x.e_1; \text{fold}\{t.\tau\}(e_2)) \mapsto [\text{map}^+\{t.\tau\}(y.\text{rec}\{t.\tau\}(x.e_1; y))(e_2)/x]e_1} \quad (15.9c)$$

$$\frac{}{\text{gen}\{t.\tau\}(x.e_1; e_2) \text{ val}} \quad (15.9d)$$

$$\frac{e \mapsto e'}{\text{unfold}\{t.\tau\}(e) \mapsto \text{unfold}\{t.\tau\}(e')} \quad (15.9e)$$

$$\frac{}{\text{unfold}\{t.\tau\}(\text{gen}\{t.\tau\}(x.e_1; e_2)) \mapsto \text{map}^+\{t.\tau\}(y.\text{gen}\{t.\tau\}(x.e_1; y))([e_2/x]e_1)} \quad (15.9f)$$

Rule (15.9c) states that to evaluate the recursor on a value of recursive type, we inductively apply the recursor as guided by the type operator to the value, and then apply the inductive step to the result. Rule (15.9f) is simply the dual of this rule for coinductive types.

Lemma 15.1. *If $e : \tau$ and $e \mapsto e'$, then $e' : \tau$.*

Proof By rule induction on rules (15.9). □

Lemma 15.2. *If $e : \tau$, then either $e \text{ val}$ or there exists e' such that $e \mapsto e'$.*

Proof By rule induction on rules (15.8). □

Although a proof of this fact lies beyond our current reach, all programs in \mathbf{M} terminate.

Theorem 15.3 (Termination for \mathbf{M}). *If $e : \tau$, then there exists $e' \text{ val}$ such that $e \mapsto^* e'$.*

It may, at first, seem surprising that a language with infinite data structures, such as streams, can enjoy such a termination property. But bear in mind that infinite data structures, such as streams, are represented as in a continuing state of creation, and not as a completed whole.

15.4 Solving Type Equations

For a positive type operator $t.\tau$, we may say that the inductive type $\mu(t.\tau)$ and the coinductive type $\nu(t.\tau)$ are both *solutions* (up to isomorphism) of the type equation $t \cong \tau$:

$$\begin{aligned}\mu(t.\tau) &\cong [\mu(t.\tau)/t]\tau \\ \nu(t.\tau) &\cong [\nu(t.\tau)/t]\tau.\end{aligned}$$

Intuitively speaking, this means that every value of an inductive type is the folding of a value of the unfolding of the inductive type, and that, similarly, every value of the unfolding of a coinductive type is the unfolding of a value of the coinductive type itself. It is a good exercise to define functions back and forth between the isomorphic types and to convince yourself informally that they are mutually inverse to one another.

Whereas both are solutions to the same type equation, they are not isomorphic to each other. To see why, consider the inductive type $\text{nat} \triangleq \mu(t.\text{unit} + t)$ and the coinductive type $\text{conat} \triangleq \nu(t.\text{unit} + t)$. Informally, nat is the smallest (most restrictive) type containing zero, given by $\text{fold}(1 \cdot \langle \rangle)$, and closed under formation of the successor of any other e of type nat , given by $\text{fold}(r \cdot e)$. Dually, conat is the largest (most permissive) type of expressions e for which the unfolding, $\text{unfold}(e)$, is either zero, given by $1 \cdot \langle \rangle$, or to the successor of some other e' of type conat , given by $r \cdot e'$.

Because nat is defined by the composition of its introduction forms and sum injections, it is clear that only finite natural numbers can be constructed in finite time. Because conat is defined by the composition of its elimination forms (unfoldings plus case analyses), it is clear that a co-natural number can only be explored to finite depth in finite time—essentially we can only examine some finite number of predecessors of a given co-natural number in a terminating program. Consequently,

1. there is a function $i : \text{nat} \rightarrow \text{conat}$ embedding every finite natural number into the type of possibly infinite natural numbers; and
2. there is an “actually infinite” co-natural number ω that is essentially an infinite composition of successors.

Defining the embedding of nat into conat is the subject of Exercise **15.1**. The infinite co-natural number ω is defined as follows:

$$\omega \triangleq \text{gen}(x.r \cdot x; \langle \rangle).$$

One may check that $\text{unfold}(\omega) \mapsto^* r \cdot \omega$, which means that ω is its own predecessor. The co-natural number ω is larger than any finite natural number in that any finite number of predecessors of ω is non-zero.

Summing up, the mere fact of being a solution to a type equation does not uniquely characterize a type: there can be many different solutions to the same type equation, the natural and the co-natural numbers being good examples of the discrepancy. However, we will show in Part VIII that type equations have unique solutions (up to isomorphism) and that the restriction to polynomial type operators is no longer required. The price we pay for the additional expressive power is that programs are no longer guaranteed to terminate.

15.5 Notes

The language **M** is named after Mendler, on whose work the present treatment is based (Mendler, 1987). Mendler's work is grounded in category theory, specifically the concept of an algebra for a functor (MacLane, 1998; Taylor, 1999). The functorial action of a type constructor (described in Chapter 14) plays a central role. Inductive types are initial algebras and coinductive types are final coalgebras for the functor given by a (polynomial or positive) type operator.

Exercises

- 15.1.** Define a function $i : \text{nat} \rightarrow \text{conat}$ that sends every natural number to “itself” in the sense that every finite natural number is sent to its correlate as a co-natural number.
- (a) $\text{unfold}(i(z)) \mapsto^* 1 \cdot \langle \rangle$.
- (b) $\text{unfold}(i(s(\bar{n}))) \mapsto^* r \cdot i(\bar{n})$.
- 15.2.** Derive the iterator, $\text{iter } e \{z \hookrightarrow e_0 \mid s(x) \hookrightarrow e_1\}$, described in Chapter 9 from the recursor for the inductive type of natural numbers given in Section 15.1.
- 15.3.** Derive the stream generator, $\text{strgen } x \text{ is } e \text{ in } \langle \text{hd} \hookrightarrow e_1, \text{tl} \hookrightarrow e_2 \rangle$ from the generator for the coinductive stream type given in Section 15.1.
- 15.4.** Consider the type $\text{seq} \triangleq \text{nat} \rightarrow \text{nat}$ of infinite sequences of natural numbers. Every stream can be turned into a sequence by the following function:

$$\lambda(\text{stream} : s) \lambda(n : \text{nat}) \text{hd}(\text{iter } n \{z \hookrightarrow s \mid s(x) \hookrightarrow \text{tl}(x)\}).$$

Show that every sequence can be turned into a stream whose n th element is the n th element of the given sequence.

- 15.5.** The type of lists of natural numbers is defined by the following introduction and elimination forms:

$$\frac{}{\Gamma \vdash \text{nil} : \text{natlist}} \tag{15.10a}$$

$$\frac{\Gamma \vdash e_1 : \text{nat} \quad \Gamma \vdash e_2 : \text{natlist}}{\Gamma \vdash \text{cons}(e_1; e_2) : \text{natlist}} \quad (15.10b)$$

$$\frac{\Gamma \vdash e : \text{natlist} \quad \Gamma \vdash e_0 : \tau \quad \Gamma x : \text{nat} y : \tau \vdash e_1 : \tau}{\Gamma \vdash \text{rec } e \{ \text{nil} \hookrightarrow e_0 \mid \text{cons}(x; y) \hookrightarrow e_1 \} : \tau} \quad (15.10c)$$

The associated dynamics, whether eager or lazy, can be derived from that of the recursor for the type `nat` given in Chapter 9. Give a definition of `natlist` as an inductive type, including the definitions of its associated introduction and elimination forms. Check that they validate the expected dynamics.

- 15.6.** Consider the type `itree` of possibly infinite binary trees with the following introduction and elimination forms:

$$\frac{\Gamma \vdash e : \text{itree}}{\Gamma \vdash \text{view}(e) : (\text{itree} \times \text{itree}) \text{opt}} \quad (15.11a)$$

$$\frac{\Gamma \vdash e : \tau \quad \Gamma x : \tau \vdash e' : (\tau \times \tau) \text{opt}}{\Gamma \vdash \text{itgen } x \text{ is } e \text{ in } e' : \text{itree}} \quad (15.11b)$$

Because a possibly infinite tree must be in a state of continual generation, viewing a tree exposes only its top-level structure, an optional pair of possibly infinite trees.¹ If the `view` is `null`, the tree is empty, and if it is `just(e1)e2`, then it is non-empty, with children given by `e1` and `e2`. To generate an infinite tree, choose a type `τ` of its state of generation, and provide its current state `e` and a state transformation `e'` that, when applied to the current state, announces whether or not generation is complete, and, if not, provides the state for each of the children.

- (a) Give a precise dynamics for the `itree` operations as just described informally. *Hint:* use generic programming!
- (b) Reformulate the type `itree` as a coinductive type, and derive the statics and dynamics of its introduction and elimination forms.
- 15.7.** Exercise **11.5** asked you to define an RS latch as a signal transducer, in which signals are expressed explicitly as functions of time. Here you are asked again to define an RS latch as a signal transducer, but this time with signals expressed as streams of booleans. Under such a representation, time is implicitly represented by the successive elements of the stream. Define an RS latch as a transducer of signals consisting of pairs of booleans.

Note

¹ See Chapter 11 for the definition of option types.

