Defining (Co)datatypes and Primitively (Co)recursive Functions in Isabelle/HOL

Julian Biendarra, Jasmin Blanchette,
Martin Desharnais, Lorenz Panny,
Andrei Popescu, and Dmitriy Traytel

15 April 2020

Abstract
This tutorial describes the definitional package for datatypes and codatatypes, and for primitively recursive and corecursive functions, in Isabelle/HOL. The following commands are provided: \texttt{datatype}, \texttt{datatype_compat}, \texttt{primrec}, \texttt{codatatype}, \texttt{primcorec}, \texttt{primcorecursive}, \texttt{bnf}, \texttt{lift_bnf}, \texttt{copy_bnf}, \texttt{bnf_axiomatization}, \texttt{print_bnsf}, and \texttt{free_constructors}.

Contents
1 Introduction .. 3
2 Defining Datatypes .. 5
  2.1 Introductory Examples .. 5
  2.1.1 Nonrecursive Types .. 5
  2.1.2 Simple Recursion .. 6
  2.1.3 Mutual Recursion .. 6
  2.1.4 Nested Recursion .. 7
  2.1.5 Auxiliary Constants .. 7
  2.2 Command Syntax .. 9
    2.2.1 \texttt{datatype} .. 9
    2.2.2 \texttt{datatype_compat} .. 12
  2.3 Generated Constants .. 13
  2.4 Generated Theorems .. 13
    2.4.1 Free Constructor Theorems .. 14
1 Introduction

The 2013 edition of Isabelle introduced a definitional package for freely generated datatypes and codatatypes. This package replaces the earlier implementation due to Berghofer and Wenzel [1]. Perhaps the main advantage of the new package is that it supports recursion through a large class of non-datatypes, such as finite sets:

\[
\text{datatype } 'a \text{ tree} = \text{Node} (\text{lbl}: 'a) (\text{sub} : 'a \text{ set} : "'a tree \ fset")
\]

Another strong point is the support for local definitions:
1 Introduction

context linorder
begin

datatype flag = Less | Eq | Greater

end

Furthermore, the package provides a lot of convenience, including automatically generated discriminators, selectors, and relators as well as a wealth of properties about them.

In addition to inductive datatypes, the package supports coinductive datatypes, or codatatypes, which allow infinite values. For example, the following command introduces the type of lazy lists, which comprises both finite and infinite values:

codatatype 'a llist = LNil | LCons 'a "'a llist"

Mixed inductive–coinductive recursion is possible via nesting. Compare the following four Rose tree examples:

datatype 'a treeff = Nodeff 'a "'a treeff list"
datatype 'a treefi = Nodefi 'a "'a treefi llist"
codatatype 'a treeiff = Nodeiff 'a "'a treeiff list"
codatatype 'a treeiif = Nodeiiff 'a "'a treeiiff llist"

The first two tree types allow only paths of finite length, whereas the last two allow infinite paths. Orthogonally, the nodes in the first and third types have finitely many direct subtrees, whereas those of the second and fourth may have infinite branching.

The package is part of Main. Additional functionality is provided by the theory --/src/HOL/Library/BNF_Axiomatization.thy.

The package, like its predecessor, fully adheres to the LCF philosophy [5]: The characteristic theorems associated with the specified (co)datatypes are derived rather than introduced axiomatically. The package is described in a number of scientific papers [2, 4, 9, 11]. The central notion is that of a bounded natural functor (BNF)—a well-behaved type constructor for which nested (co)recursion is supported.

This tutorial is organized as follows:

- Section 2, “Defining Datatypes,” describes how to specify datatypes using the datatype command.
- Section 3, “Defining Primitively Recursive Functions,” describes how to specify functions using primrec. (A separate tutorial [6] describes the more powerful fun and function commands.)

\[\text{However, some of the internal constructions and most of the internal proof obligations are omitted if the quick_and_dirty option is enabled.}\]
2 Defining Datatypes

Datatypes can be specified using the \texttt{datatype} command.

2.1 Introductory Examples

Datatypes are illustrated through concrete examples featuring different flavors of recursion. More examples can be found in the directory \texttt{~~/src/HOL/Datatype_Examples}.

2.1.1 Nonrecursive Types

Datatypes are introduced by specifying the desired names and argument types for their constructors. \textit{Enumeration} types are the simplest form of \texttt{datatype}. All their constructors are nullary:

\begin{verbatim}
  datatype bool = True | False | Perhaps
\end{verbatim}
True, False, and Perhaaps have the type trool.

Polymorphic types are possible, such as the following option type, modeled after its homologue from the HOL Option theory:

```plaintext
datatype 'a option = None | Some 'a
```

The constructors are None :: 'a option and Some :: 'a ⇒ 'a option.

The next example has three type parameters:

```plaintext
datatype ('a, 'b, 'c) triple = Triple 'a 'b 'c
```

The constructor is Triple :: 'a ⇒ 'b ⇒ 'c ⇒ ('a, 'b, 'c) triple. Unlike in Standard ML, curried constructors are supported. The uncurried variant is also possible:

```plaintext
datatype ('a, 'b, 'c) triple_u = Triple_u "'a * 'b * 'c"
```

Occurrences of nonatomic types on the right-hand side of the equal sign must be enclosed in double quotes, as is customary in Isabelle.

### 2.1.2 Simple Recursion

Natural numbers are the simplest example of a recursive type:

```plaintext
datatype nat = Zero | Succ nat
```

Lists were shown in the introduction. Terminated lists are a variant that stores a value of type 'b at the very end:

```plaintext
datatype ('a, 'b) tlist = TNil 'b | TCons 'a ('a, 'b) tlist
```

### 2.1.3 Mutual Recursion

Mutually recursive types are introduced simultaneously and may refer to each other. The example below introduces a pair of types for even and odd natural numbers:

```plaintext
datatype even_nat = Even_Zero | Even_Succ odd_nat
and odd_nat = Odd_Succ even_nat
```

Arithmetic expressions are defined via terms, terms via factors, and factors via expressions:

```plaintext
datatype ('a, 'b) exp =
  Term "('a, 'b) trm" | Sum "('a, 'b) trm" "('a, 'b) exp"
and ('a, 'b) trm =
  Factor "('a, 'b) fct" | Prod "('a, 'b) fct" "('a, 'b) trm"
and ('a, 'b) fct =
  Const 'a | Var 'b | Expr "('a, 'b) exp"
```
2.1.4 Nested Recursion

Nested recursion occurs when recursive occurrences of a type appear under a type constructor. The introduction showed some examples of trees with nesting through lists. A more complex example, that reuses our option type, follows:

```
datatype 'a btree =
    BNode 'a "a btree option" "a btree option"
```

Not all nestings are admissible. For example, this command will fail:

```
datatype 'a wrong = W1 | W2 "a wrong ⇒ 'a"
```

The issue is that the function arrow ⇒ allows recursion only through its right-hand side. This issue is inherited by polymorphic datatypes defined in terms of ⇒:

```
datatype ('a, 'b) fun_copy = Fun "a ⇒ 'b"
datatype 'a also_wrong = W1 | W2 "('a also_wrong, 'a) fun_copy"
```

The following definition of 'a-branching trees is legal:

```
datatype 'a ftree = FTLeaf 'a | FTNode "'a ⇒ 'a ftree"
```

And so is the definition of hereditarily finite sets:

```
datatype hfset = HFSet "hfset fset"
```

In general, type constructors ('a_1, ..., 'a_m) t allow recursion on a subset of their type arguments 'a_1, ..., 'a_m. These type arguments are called live; the remaining type arguments are called dead. In 'a ⇒ 'b and ('a, 'b) fun_copy, the type variable 'a is dead and 'b is live.

Type constructors must be registered as BNFs to have live arguments. This is done automatically for datatypes and codatatypes introduced by the `datatype` and `codatatype` commands. Section 6 explains how to register arbitrary type constructors as BNFs.

Here is another example that fails:

```
datatype 'a pow_list = PNil 'a | PCons "('a * 'a) pow_list"
```

This attempted definition features a different flavor of nesting, where the recursive call in the type specification occurs around (rather than inside) another type constructor.

2.1.5 Auxiliary Constants

The `datatype` command introduces various constants in addition to the constructors. With each datatype are associated set functions, a map function, a
predicator, a relator, discriminators, and selectors, all of which can be given custom names. In the example below, the familiar names null, hd, tl, set, map, and list_all2 override the default names is_Nil, un_Cons1, un_Cons2, set_list, map_list, and rel_list:

```datatype (set: 'a) list = 
null: Nil 
| Cons (hd: 'a) (tl: "a list")
for 
  map: map
  rel: list_all2
  pred: list_all
where 
  "tl Nil = Nil"
```

The types of the constants that appear in the specification are listed below.

Constructors:  
Nil :: 'a list
  Cons :: 'a ⇒ 'a list ⇒ 'a list

Discriminator:  
null :: 'a list ⇒ bool

Selectors:  
hd :: 'a list ⇒ 'a
tl :: 'a list ⇒ 'a list

Set function:  
set :: 'a list ⇒ 'a set

Map function:  
map :: ('a ⇒ 'b) ⇒ 'a list ⇒ 'b list

Relator:  
list_all2 :: ('a ⇒ 'b ⇒ bool) ⇒ 'a list ⇒ 'b list ⇒ bool

The discriminator null and the selectors hd and tl are characterized by the following conditional equations:

\[ null \, xs \implies xs = Nil \quad \neg null \, xs \implies Cons \, (hd \, xs) \, (tl \, xs) = xs \]

For two-constructor datatypes, a single discriminator constant is sufficient. The discriminator associated with Cons is simply \( \lambda xs. \neg null \, xs \).

The where clause at the end of the command specifies a default value for selectors applied to constructors on which they are not a priori specified. In the example, it is used to ensure that the tail of the empty list is itself (instead of being left unspecified).

Because Nil is nullary, it is also possible to use \( \lambda xs. xs = Nil \) as a discriminator. This is the default behavior if we omit the identifier null and the associated colon. Some users argue against this, because the mixture of constructors and selectors in the characteristic theorems can lead Isabelle’s automation to switch between the constructor and the destructor view in surprising ways.

The usual mixfix syntax annotations are available for both types and constructors. For example:
2 Defining Datatypes

\[
\text{datatype} \ (\texttt{\char127}a, \texttt{\char127}b) \ \text{prod} \ (\text{infixr} \ \texttt{\char34}^* \ 20) = \text{Pair} \ \texttt{\char127}a \ \texttt{\char127}b
\]

\[
\text{datatype} \ (\text{set:} \ a) \ \text{list} =
\ 
\text{null:} \ \text{Nil} \ (\texttt{"[]"})
\ 
| \text{Cons} \ (hd: \ a) \ (tl: \ \texttt{\char127}a \ \text{list}) \ (\text{infixr} \ \texttt{\char34}# \ 65)
\]

\begin{itemize}
  \item \text{for}
  \begin{itemize}
    \item map: \ map
    \item rel: \ \text{list\_all2}
    \item pred: \ \text{list\_all}
  \end{itemize}
\end{itemize}

Incidentally, this is how the traditional syntax can be set up:

\[
\text{syntax} \ \texttt{\_list} :: \ \texttt{\_args} \Rightarrow \textbf{\char127}a \ \texttt{list} \ (\texttt{"[\_]"})
\]

\[
\text{translations}
\ 
\texttt{\_\_list} \ \Rightarrow \ \texttt{\_\_args} \ \Rightarrow \ \textbf{\char127}a \ \texttt{list} \ (\texttt{"[\_]"})
\]

\[
\text{\_\_list} \ \Rightarrow \ \texttt{\_\_args} \ \Rightarrow \ \textbf{\char127}a \ \texttt{list} \ (\texttt{"[\_]"})
\]

2.2 Command Syntax

2.2.1 datatype

\[
\text{datatype} : \ \text{local\_theory} \rightarrow \text{local\_theory}
\]
The \texttt{datatype} command introduces a set of mutually recursive datatypes specified by their constructors. The syntactic entity \texttt{target} can be used to specify a local context (e.g., \texttt{(in linorder) [12]}), and \texttt{prop} denotes a HOL proposition.

The optional target is optionally followed by a combination of the following options:

- The \texttt{plugins} option indicates which plugins should be enabled (\texttt{only}) or disabled (\texttt{del}). By default, all plugins are enabled.
- The \texttt{discs\_sels} option indicates that discriminators and selectors should be generated. The option is implicitly enabled if names are specified for discriminators or selectors.

The optional \texttt{where} clause specifies default values for selectors. Each proposition must be an equation of the form \texttt{un\_D (C \ldots) = \ldots}, where \texttt{C} is a constructor and \texttt{un\_D} is a selector.
The left-hand sides of the datatype equations specify the name of the type to define, its type parameters, and additional information:

\[ \text{dt-name} \]

\[
\begin{align*}
\text{name} & \quad \text{tyargs} \\
\text{mixfix} & \quad \text{typefree}
\end{align*}
\]

**tyargs**

\[
\begin{align*}
\text{typefree} & \quad ( \\
\text{dead} & \quad \text{name} : \\
\text{typefree} & \quad )
\end{align*}
\]

The syntactic entity *name* denotes an identifier, *mixfix* denotes the usual parenthesized mixfix notation, and *typefree* denotes fixed type variable (\('a, 'b, \ldots\) \[12\]).

The optional names preceding the type variables allow to override the default names of the set functions (\(set_1_t, \ldots, set_m_t\)). Type arguments can be marked as dead by entering *dead* in front of the type variable (e.g., \((\text{dead } 'a)\)); otherwise, they are live or dead (and a set function is generated or not) depending on where they occur in the right-hand sides of the definition. Declaring a type argument as dead can speed up the type definition but will prevent any later (co)recursion through that type argument.

Inside a mutually recursive specification, all defined datatypes must mention exactly the same type variables in the same order.

\[ \text{dt-ctor} \]

\[
\begin{align*}
\text{name} & \quad : \\
\text{dt-ctor-arg} & \quad \text{mixfix}
\end{align*}
\]
The main constituents of a constructor specification are the name of the
constructor and the list of its argument types. An optional discriminator
name can be supplied at the front. If discriminators are enabled (cf. the
discs_sels option) but no name is supplied, the default is \( \lambda x. x = C_j \) for
nullary constructors and \( t.is_C_j \) otherwise.

2.2.2 datatype_compat

datatype_compat : local_theory \rightarrow local_theory

The datatype_compat command registers new-style datatypes as old-style
datatypes and invokes the old-style plugins. For example:

datatype_compat even_nat odd_nat

ML (Old_Datatype_Data.get_info theory type_name (even_nat))

The syntactic entity name denotes an identifier [12].
The command is sometimes useful when migrating from the old datatype
package to the new one.

A few remarks concern nested recursive datatypes:
2 Defining Datatypes

• The old-style, nested-as-mutual induction rule and recursor theorems are generated under their usual names but with “compat_” prefixed (e.g., compat_tree.induct, compat_tree.inducts, and compat_tree.rec). These theorems should be identical to the ones generated by the old datatype package, up to the order of the premises—meaning that the subgoals generated by the induct or induction method may be in a different order than before.

• All types through which recursion takes place must be new-style datatypes or the function type.

2.3 Generated Constants

Given a datatype \('a_1, \ldots, 'a_m) t\) with \(m\) live type variables and \(n\) constructors \(t.C_1, \ldots, t.C_n\), the following auxiliary constants are introduced:

- Case combinator: \(t.case_t\) (rendered using the familiar case–of syntax)
- Discriminators: \(t.is_{C_1}, \ldots, t.is_{C_n}\)
- Selectors: \(t.un_{C_1}1, \ldots, t.un_{C_1}k_1\)
  
  \[\vdots\]
  
  \(t.un_{C_n}1, \ldots, t.un_{C_n}k_n\)
- Set functions: \(t.set_1_t, \ldots, t.set_m_t\)
- Map function: \(t.map_t\)
- Relator: \(t.rel_t\)
- Recursor: \(t.rec_t\)

The discriminators and selectors are generated only if the discs_sels option is enabled or if names are specified for discriminators or selectors. The set functions, map function, predicator, and relator are generated only if \(m > 0\).

In addition, some of the plugins introduce their own constants (Section 8). The case combinator, discriminators, and selectors are collectively called destructors. The prefix “\(t\)” is an optional component of the names and is normally hidden.

2.4 Generated Theorems

The characteristic theorems generated by \texttt{datatype} are grouped in three broad categories:

• The free constructor theorems (Section 2.4.1) are properties of the constructors and destructors that can be derived for any freely generated type. Internally, the derivation is performed by \texttt{free_constructors}. 
- The functorial theorems (Section 2.4.2) are properties of datatypes related to their BNF nature.
- The inductive theorems (Section 2.4.3) are properties of datatypes related to their inductive nature.

The full list of named theorems can be obtained by issuing the command `print_theorems` immediately after the datatype definition. This list includes theorems produced by plugins (Section 8), but normally excludes low-level theorems that reveal internal constructions. To make these accessible, add the line

```
declare [[bnf_internals]]
```

### 2.4.1 Free Constructor Theorems

The free constructor theorems are partitioned in three subgroups. The first subgroup of properties is concerned with the constructors. They are listed below for 'a list:

- **`t.inject`** [iff, induct_simp]:
  \[ (x_{21} \# x_{22} = y_{21} \# y_{22}) = (x_{21} = y_{21} \wedge x_{22} = y_{22}) \]

- **`t.distinct`** [simp, induct_simp]:
  \[
  \emptyset \neq x_{21} \# x_{22} \\
  x_{21} \# x_{22} \neq \emptyset 
  \]

- **`t.exhaust`** [cases t, case_names C₁ ... Cₙ]:
  \[
  \begin{array}{c}
  y = \emptyset \implies P; \ \Lambda x_{21} x_{22}. \ y = x_{21} \# x_{22} \implies P \\
  \end{array} \implies P
  \]

- **`t.nchotomy`**:
  \[
  \forall \text{list}. \ \text{list} = \emptyset \lor (\exists x_{21} x_{22}. \ \text{list} = x_{21} \# x_{22})
  \]

In addition, these nameless theorems are registered as safe elimination rules:

- **`t.distinct`** [THEN notE, elim!]:
  \[
  \begin{array}{c}
  \emptyset = x_{21} \# x_{22} \implies R \\
  x_{21} \# x_{22} = \emptyset \implies R \\
  \end{array}
  \]

The next subgroup is concerned with the case combinator:

- **`t.case`** [simp, code]:
  \[
  \begin{array}{c}
  \text{(case } \emptyset \text{ of } \emptyset \Rightarrow f_{1} | x \# x_{a} \Rightarrow f_{2} x_{a}) = f_{1} \\
  \text{(case } x_{21} \# x_{22} \text{ of } \emptyset \Rightarrow f_{1} | x \# x_{a} \Rightarrow f_{2} x_{a}) = f_{2} x_{21} x_{22} \\
  \end{array}
  \]

The [code] attribute is set by the code plugin (Section 8.1).
The third subgroup revolves around discriminators and selectors:

\[ \texttt{t.case\_cong \ [fundef\_cong]}: \]
\[
\text{list} = \text{list'}, \text{list'} = \texttt{[]} \implies f1 = g1; \land \ x21 \ x22. \ \text{list'} = x21 \ # \ x22 \implies \\
\text{f2} \ x21 \ x22 = g2 \ x21 \ x22 \implies (\text{case list of } \texttt{[]} \implies f1 | \ x21 \ # \ x22 \implies \\
\text{f2} \ x21 \ x22) = (\text{case list' of } \texttt{[]} \implies g1 | \ x21 \ # \ x22 \implies g2 \ x21 \ x22) \]

\[ \texttt{t.case\_cong\_weak \ [cong]}: \]
\[
\text{list} = \text{list'} \implies (\text{case list of } \texttt{[]} \implies f1 | \ x \ # \ xa \implies f2 \ x \ xa) = (\text{case list' of } \texttt{[]} \implies f1 | \ x \ # \ xa \implies f2 \ x \ xa) \]

\[ \texttt{t.case\_distrib}: \]
\[
h \ (\text{case list of } \texttt{[]} \implies f1 | \ x \ # \ xa \implies f2 \ x \ xa) = (\text{case list of } \texttt{[]} \implies h \ (f1 \ x1 \ x2)) \]

\[ \texttt{t.split}: \]
\[
P \ (\text{case list of } \texttt{[]} \implies f1 | \ x \ # \ xa \implies f2 \ x \ xa) = ((\text{list} = \texttt{[]} \implies P \ f1) \land \\
(\forall \ x21 \ x22. \ \text{list} = x21 \ # \ x22 \implies P \ (f2 \ x21 \ x22))) \]

\[ \texttt{t.split\_asm}: \]
\[
P \ (\text{case list of } \texttt{[]} \implies f1 | \ x \ # \ xa \implies f2 \ x \ xa) = (\neg \ (\text{list} = \texttt{[]} \land \neg \ P \ f1 \lor \ (\exists \ x21 \ x22. \ \text{list} = x21 \ # \ x22 \land \neg \ P \ (f2 \ x21 \ x22))) \]

\[ \texttt{t.splits} = \text{split split\_asm} \]

The third subgroup revolves around discriminators and selectors:

\[ \texttt{t.disc \ [simp]}: \]
\[
\neg \ \text{null } \texttt{[]} \]
\[
\neg \ \text{null } (x21 \ # \ x22) \]

\[ \texttt{t.discI}: \]
\[
\text{list} = \texttt{[]} \implies \text{null list} \]
\[
\text{list} = x21 \ # \ x22 \implies \neg \ \text{null list} \]

\[ \texttt{t.sel \ [simp, \ code]}: \]
\[
\text{hd } (x21 \ # \ x22) = x21 \]
\[
\text{tl } (x21 \ # \ x22) = x22 \]

The [code] attribute is set by the code plugin (Section 8.1).

\[ \texttt{t.collapse \ [simp]}: \]
\[
\neg \ \text{null list} \implies \text{list} = \texttt{[]} \]
\[
\neg \ \text{null list} \implies \text{hd list} \ # \ \text{tl list} = \text{list} \]

The [simp] attribute is exceptionally omitted for datatypes equipped
with a single nullary constructor, because a property of the form \( x = C \) is not suitable as a simplification rule.

\[ \texttt{t.distinct\_disc \ [dest]}: \]
\[
\text{These properties are missing for } \texttt{'}a \text{ list} \text{ because there is only one} \]
proper discriminator. If the datatype had been introduced with a second discriminator called *nonnull*, they would have read as follows:

\[
null \text{ list } \Rightarrow \neg \text{nonnull list} \\
\text{nonnull list } \Rightarrow \neg \text{null list}
\]

\texttt{t.exhaust\_disc [case\_names } C_1 \ldots \ C_n]:

\[
[null \text{ list } \Rightarrow P; \neg \text{null list } \Rightarrow P] \Rightarrow P
\]

\texttt{t.exhaust\_sel [case\_names } C_1 \ldots \ C_n]:

\[
[\text{list } = [] \Rightarrow P; \text{list } = \text{hd list } \# \text{tl list } \Rightarrow P] \Rightarrow P
\]

\texttt{t.expand:}

\[
[\text{null list } = \text{null list'}; [\neg \text{null list}; \neg \text{null list’}] \Rightarrow \text{hd list } = \text{hd list’}
\wedge \text{tl list } = \text{tl list’}] \Rightarrow \text{list } = \text{list’}
\]

\texttt{t.split\_sel:}

\[
P (\text{case list of } []) \Rightarrow f_1 | \text{x } \# \text{xa } \Rightarrow f_2 \text{ x xa}) = ((\text{list } = [] \Rightarrow P f_1)
\wedge (\text{list } = \text{hd list } \# \text{tl list } \Rightarrow P (f_2 (\text{hd list} (\text{tl list}))))
\]

\texttt{t.split\_sel\_asm:}

\[
P (\text{case list of } []) \Rightarrow f_1 | \text{x } \# \text{xa } \Rightarrow f_2 \text{ x xa}) = (\neg (\text{list } = [] \wedge \neg P
f_1 \vee \text{list } = \text{hd list } \# \text{tl list } \wedge \neg P (f_2 (\text{hd list} (\text{tl list}))))
\]

\texttt{t.split\_sels} = split\_sel split\_sel\_asm

\texttt{t.case\_eq\_if:}

\[
(\text{case list of } []) \Rightarrow f_1 | \text{x } \# \text{xa } \Rightarrow f_2 \text{ x xa}) = (\text{if null list then } f_1 \text{ else } f_2 (\text{hd list} (\text{tl list}))
\]

\texttt{t.disc\_eq\_case:}

\[
\text{null list } = (\text{case list of } []) \Rightarrow \text{True} | \text{wu_ } \# \text{wua_ } \Rightarrow \text{False}
\neg \text{null list} = (\text{case list of } []) \Rightarrow \text{False} | \text{wu_ } \# \text{wua_ } \Rightarrow \text{True}
\]

In addition, equational versions of \texttt{t.disc} are registered with the \texttt{[code]} attribute. The \texttt{[code]} attribute is set by the \texttt{code} plugin (Section 8.1).

### 2.4.2 Functorial Theorems

The functorial theorems are generated for type constructors with at least one live type argument (e.g., ‘a list). They are partitioned in two subgroups. The first subgroup consists of properties involving the constructors or the destructors and either a set function, the map function, the predicator, or the relator:

\texttt{t.case\_transfer [transfer\_rule]:}

\[
\text{rel\_fun } S (\text{rel\_fun } (\text{rel\_fun } R (\text{rel\_fun } (\text{list\_all2 } R) S)) (\text{rel\_fun } (\text{list\_all2 } R) S)) \text{ case\_list case\_list}
\]

This property is generated by the \texttt{transfer} plugin (Section 8.3).
t.sel_transfer [transfer_rule]:
This property is missing for 'a list because there is no common selector to all constructors.
The [transfer_rule] attribute is set by the transfer plugin (Section 8.3).

t.ctr_transfer [transfer_rule]:
list_all2 R [] []
rel_fun R (rel_fun (list_all2 R) (list_all2 R)) (#) (#)
The [transfer_rule] attribute is set by the transfer plugin (Section 8.3).

t.disc_transfer [transfer_rule]:
rel_fun (list_all2 R) (=) null null
rel_fun (list_all2 R) (=) (λlist. ¬ null list) (λlist. ¬ null list)
The [transfer_rule] attribute is set by the transfer plugin (Section 8.3).

t.set [simp, code]:
set [] = {}
set (x21 # x22) = insert x21 (set x22)
The [code] attribute is set by the code plugin (Section 8.1).

t.set_cases [consumes 1, cases set: set_t]:
[e ∈ set a; ∀z2. a = e # z2 ⇒ thesis; ∀z1 z2. [a = z1 # z2; e ∈ set z2] ⇒ thesis] ⇒ thesis

t.set_intros:
x21 ∈ set (x21 # x22)
y ∈ set x22 ⇒ y ∈ set (x21 # x22)

t.set_sel:
¬ null a ⇒ hd a ∈ set a
[¬ null a; x ∈ set (tl a)] ⇒ x ∈ set a

t.map [simp, code]:
map f [] = []
map f (x21 # x22) = f x21 # map f x22
The [code] attribute is set by the code plugin (Section 8.1).

t.map_disc_iff [simp]:
null (map f a) = null a

t.map_sel:
¬ null a ⇒ hd (map f a) = f (hd a)
¬ null a ⇒ tl (map f a) = map f (tl a)

t.pred_inject [simp]:
list_all P []
list_all P (a # aa) = (P a ∧ list_all P aa)
t.rel_inject [simp]:
\[ \text{list_all} R [] \] 
\[ \text{list_all} R (x21 \# x22) (y21 \# y22) = (R x21 y21 \land \text{list_all} R x22 y22) \]

\[ t.\text{rel_distinct} [\text{simp}] : \]
\[- \text{list_all} R [] (y21 \# y22) \]
\[- \text{list_all} R (y21 \# y22) [] \]

t.rel_intros:
\[ \text{list_all} R [] [] \]
\[ [R x21 y21; \text{list_all} R x22 y22] \implies \text{list_all} R (x21 \# x22) (y21 \# y22) \]

t.rel_cases [consumes 1, case_names t_1 \ldots t_m, cases pred]:
\[ [\text{list_all} R a b; [a = []]; b = []] \implies \text{thesis}; \land x1 x2 y1 y2. [a = x1 \# x2; b = y1 \# y2; R x1 y1; \text{list_all} R x2 y2] \implies \text{thesis} \implies \text{thesis} \]

t.rel_sel:
\[ \text{list_all} R a b = (\text{null} a = \text{null} b \land (\neg \text{null} a \rightarrow \neg \text{null} b \rightarrow R (\text{hd} a) (\text{hd} b) \land \text{list_all} R (\text{tl} a) (\text{tl} b))) \]

In addition, equational versions of \text{t.rel_inject} and \text{rel_distinct} are registered with the \text{[code]} attribute. The \text{[code]} attribute is set by the code plugin (Section 8.1).

The second subgroup consists of more abstract properties of the set functions, the map function, the predicator, and the relator:

t.inj_map:
\[ \text{inj} f \implies \text{inj} (\text{map} f) \]

t.inj_map_strong:
\[ [\land z. \text{z} \in \text{set} x; \text{za} \in \text{set xa}; f \text{z} = f \text{za}] \implies \text{z} = \text{za}; \text{map} f x = \text{map} f a \text{xa} \implies x = xa \]

t.map_comp:
\[ \text{map} g (\text{map} f v) = \text{map} (g \circ f) v \]

t.map_cong0:
\[ (\land z. \text{z} \in \text{set} x \implies f \text{z} = g \text{z}) \implies \text{map} f x = \text{map} g x \]

t.map_cong [fundef_cong]:
\[ [x = ya; \land z. \text{z} \in \text{set} ya \implies f \text{z} = g \text{z}] \implies \text{map} f x = \text{map} g ya \]

t.map_cong_pred:
\[ [x = ya; \text{list_all} (\land z. f \text{z} = g \text{z}) ya] \implies \text{map} f x = \text{map} g ya \]
t.map_cong_simp:
\[ x = ya; \forall z. z \in \text{set} ya \implies f z = g z \] \Rightarrow \text{map} f x = \text{map} g ya

t.map_id0:
map id = id

t.map_id:
map id t = t

t.map_ident:
map (\lambda x. x) t = t

t.map_transfer [transfer_rule]:
rel_fun (rel_fun Rb Sd) (rel_fun (list_all2 Rb) (list_all2 Sd)) map map
The [transfer_rule] attribute is set by the transfer plugin (Section 8.3) for type constructors with no dead type arguments.

t.pred_cong [fundef_cong]:
\[ x = ya; \forall z. z \in \text{set} ya \implies P z = Pa z \] \Rightarrow \text{list_all} P x = \text{list_all} Pa ya

t.pred_cong_simp:
\[ x = ya; \forall z. z \in \text{set} ya \implies P z = Pa z \] \Rightarrow \text{list_all} P x = \text{list_all} Pa ya

t.pred_map:
\text{list_all} Q (map f x) = \text{list_all} (Q \circ f) x

t.pred_mono_strong:
\[ \text{list_all} P x; \forall z. [z \in \text{set} x; P z] \implies Pa z \] \Rightarrow \text{list_all} Pa x

t.pred_rel:
\text{list_all} P x = \text{list_all2} (\text{eq_onp} P) x x

t.pred_set:
\text{list_all} P = (\lambda x. \text{Ball} (\text{set} x) P)

t.pred_transfer [transfer_rule]:
rel_fun (rel_fun R (\equiv)) (rel_fun (list_all2 R) (\equiv)) list_all list_all
The [transfer_rule] attribute is set by the transfer plugin (Section 8.3) for type constructors with no dead type arguments.

t.pred_True:
\text{list_all} (\lambda_. \text{True}) = (\lambda_. \text{True})

t.set_map:
\text{set} (\text{map} f v) = f ^ {\text{+}} \text{set} v
t.set_transfer [transfer_rule]:
rel_fun (list_all2 R) (rel_set R) set set
The [transfer_rule] attribute is set by the transfer plugin (Section 8.3)
for type constructors with no dead type arguments.

t.rel_compp [relator_distr]:
list_all2 (R OO S) = list_all2 R OO list_all2 S
The [relator_distr] attribute is set by the lifting plugin (Section 8.4).

t.rel_conversep:
list_all2 R−− = (list_all2 R)−−

t.rel_eq:
list_all2 (=) = (=)

t.rel_eq_onp:
list_all2 (eq_onp P) = eq_onp (list_all P)

t.rel_flip:
list_all2 R−− a b = list_all2 R b a

t.rel_map:
list_all2 Sb (map i x) y = list_all2 (λx. Sb (i x)) x y
list_all2 Sa x (map g y) = list_all2 (λx y. Sa x (g y)) x y

t.rel_mono [mono, relator_mono]:
R ≤ Ra ⇒ list_all2 R ≤ list_all2 Ra
The [relator_mono] attribute is set by the lifting plugin (Section 8.4).

t.rel_mono_strong:
[[list_all2 R x y; \ z yb. [z ∈ set x; yb ∈ set y; R z yb] ⇒ Ra z yb] ⇒ list_all2 Ra x y

t.rel_cong [fundef_cong]:
[[x = ya; y = xa; \ z yb. [z ∈ set ya; yb ∈ set xa] ⇒ R z yb = Ra z yb] ⇒ list_all2 R x y = list_all2 Ra ya xa

t.rel_cong_simp:
[[x = ya; y = xa; \ z yb. z ∈ set ya =simp=> yb ∈ set xa =simp=> R z yb = Ra z yb] ⇒ list_all2 R x y = list_all2 Ra ya xa

t.rel_refl:
(\ x. Ra x x) ⇒ list_all2 Ra x x

t.rel_refl_strong:
(\ z. z ∈ set x ⇒ Ra z z) ⇒ list_all2 Ra x x

t.rel_reflp:
reflp R ⇒ reflp (list_all2 R)
2. Defining Datatypes

\[\text{rel}_\text{symp}:\]
\[\text{symp} R \implies \text{symp} (\text{list}_\text{all}2 R)\]

\[\text{rel}_\text{transp}:\]
\[\text{transp} R \implies \text{transp} (\text{list}_\text{all}2 R)\]

\[\text{rel}_\text{transfer} \text{[transfer_rule]}:\]
\[\text{rel}_\text{fun} (\text{rel}_\text{fun} S_\text{a} (\text{rel}_\text{fun} S_\text{c} (=))) (\text{rel}_\text{fun} (\text{list}_\text{all}2 S_\text{a}) (\text{rel}_\text{fun} (\text{list}_\text{all}2 S_\text{c}) (=))) \text{ list}_\text{all}2 \text{ list}_\text{all}2\]

The [transfer_rule] attribute is set by the transfer plugin (Section 8.3) for type constructors with no dead type arguments.

2.4.3 Inductive Theorems

The inductive theorems are as follows:

\[\text{induct} \text{[case_names C}_1 \ldots \text{ C}_n, \text{ induct t]}:\]
\[\left[P []; \land x_1 x_2. P x_2 \implies P (x_1 \neq x_2)\right] \implies P \text{ list}\]

\[\text{rel}_\text{induct} \text{[case_names C}_1 \ldots \text{ C}_n, \text{ induct pred]}:\]
\[\left[\text{list}_\text{all}2 R x y; Q [] []; \land a_{21} a_{22} b_{21} b_{22}. \left[R a_{21} b_{21}; Q a_{22} b_{22}\right] \implies Q (a_{21} \neq a_{22}) (b_{21} \neq b_{22})\right] \implies Q x y\]

\[t_{1, \ldots , t_m}.\text{induct} \text{[case_names C}_1 \ldots \text{ C}_n]::\]

\[t_{1, \ldots , t_m}.\text{rel}_\text{induct} \text{[case_names C}_1 \ldots \text{ C}_n]::\]

Given \(m > 1\) mutually recursive datatypes, this induction rule can be used to prove \(m\) properties simultaneously.

\[\text{rec} \text{[simp, code]}:\]
\[\text{rec}_\text{list} f_1 f_2 [] = f_1\]
\[\text{rec}_\text{list} f_1 f_2 (x_{21} \neq x_{22}) = f_2 x_{21} x_{22} (\text{rec}_\text{list} f_1 f_2 x_{22})\]

The [code] attribute is set by the code plugin (Section 8.1).

\[\text{rec}_\text{o_map}::\]
\[\text{rec}_\text{list} g g a \circ \text{map} f = \text{rec}_\text{list} g (\lambda x x a. g a (f x)) (\text{map} f x a)\]

\[\text{rec}_\text{transfer} \text{[transfer_rule]}::\]
\[\text{rel}_\text{fun} S (\text{rel}_\text{fun} (\text{rel}_\text{fun} R (\text{rel}_\text{fun} (\text{list}_\text{all}2 R) (\text{rel}_\text{fun} S S))) (\text{rel}_\text{fun} (\text{list}_\text{all}2 R) S)) \text{ rec}_\text{list} \text{ rec}_\text{list}\]

The [transfer_rule] attribute is set by the transfer plugin (Section 8.3) for type constructors with no dead type arguments.

For convenience, datatype also provides the following collection:

\[t.\text{simps} = t.\text{inject} t.\text{distinct} t.\text{case} t.\text{rec} t.\text{map} t.\text{rel}_\text{inject} t.\text{rel}_\text{distinct} t.\text{set}\]
2.5 Proof Method

2.5.1 countable_datatype

The theory 
```
--/src/HOL/Library/Countable.thy
```
provides a proof method called *countable_datatype* that can be used to prove the countability of many datatypes, building on the countability of the types appearing in their definitions and of any type arguments. For example:

```
instance list :: (countable) countable
  by countable_datatype
```

2.6 Antiquotation

2.6.1 datatype

The *datatype* antiquotation, written `\datatype t` or @{datatype *t*}, where *t* is a type name, expands to LaTeX code for the definition of the datatype, with each constructor listed with its argument types. For example, if *t* is `option`:

```
datatype 'a option = None | Some 'a
```

2.7 Compatibility Issues

The command *datatype* has been designed to be highly compatible with the old, pre-Isabelle2015 command, to ease migration. There are nonetheless a few incompatibilities that may arise when porting:

- *The Standard ML interfaces are different.* Tools and extensions written to call the old ML interfaces will need to be adapted to the new interfaces. The *BNF_LFP_Compat* structure provides convenience functions that simulate the old interfaces in terms of the new ones.

- *The recursor rec_t has a different signature for nested recursive datatypes.* In the old package, nested recursion through non-functions was internally reduced to mutual recursion. This reduction was visible in the type of the recursor, used by *primrec*. Recursion through functions was handled specially. In the new package, nested recursion (for functions and non-functions) is handled in a more modular fashion. The old-style recursor can be generated on demand using *primrec* if the recursion is via new-style datatypes, as explained in Section 3.1.5, or using *datatype_compat*. 
Accordingly, the induction rule is different for nested recursive datatypes. Again, the old-style induction rule can be generated on demand using `primrec` if the recursion is via new-style datatypes, as explained in Section 3.1.5, or using `datatype_compat`. For recursion through functions, the old-style induction rule can be obtained by applying the `[unfolded all_mem_range]` attribute on `t.induct`.

The size function has a slightly different definition. The new function returns 1 instead of 0 for some nonrecursive constructors. This departure from the old behavior made it possible to implement `size` in terms of the generic function `t.size_t`. Moreover, the new function considers nested occurrences of a value, in the nested recursive case. The old behavior can be obtained by disabling the `size` plugin (Section 8) and instantiating the `size` type class manually.

The internal constructions are completely different. Proof texts that unfold the definition of constants introduced by the old command will be difficult to port.

Some constants and theorems have different names. For non-mutually recursive datatypes, the alias `t.inducts` for `t.induct` is no longer generated. For $m > 1$ mutually recursive datatypes, `rec_t1_. . . _t_m_. i` has been renamed `rec_t_i` for each $i \in \{1, \ldots, m\}$; `t1_. . . _t_m_.inducts(i)` has been renamed `t_i.induct` for each $i \in \{1, \ldots, m\}$, and the collection `t1_. . . _t_m_.size` (generated by the `size` plugin, Section 8.2) has been divided into `t1_.size`, `\ldots`, `t_m_.size`.

The `t.simps` collection has been extended. Previously available theorems are available at the same index as before.

Variables in generated properties have different names. This is rarely an issue, except in proof texts that refer to variable names in the `[where ...]` attribute. The solution is to use the more robust `[of ...]` syntax.

## 3 Defining Primitively Recursive Functions

Recursive functions over datatypes can be specified using the `primrec` command, which supports primitive recursion, or using the `fun`, `function`, and `partial_function` commands. In this tutorial, the focus is on `primrec`; `fun` and `function` are described in a separate tutorial [6].

Because it is restricted to primitive recursion, `primrec` is less powerful than `fun` and `function`. However, there are primitively recursive specifications (e.g., based on infinitely branching or mutually recursive datatypes)
for which fun’s termination check fails. It is also good style to use the simpler primrec mechanism when it works, both as an optimization and as documentation.

3.1 Introductory Examples

Primitive recursion is illustrated through concrete examples based on the datatypes defined in Section 2.1. More examples can be found in the directory ~/src/HOL/Datatype_Examples.

3.1.1 Nonrecursive Types

Primitive recursion removes one layer of constructors on the left-hand side in each equation. For example:

primrec (nonexhaustive) bool_of_trool :: “trool ⇒ bool” where
“bool_of_trool False” ←→ False
| “bool_of_trool True” ←→ True

primrec the_list :: “a option ⇒ ’a list” where
“the_list None” = []
| “the_list (Some a)” = [a]”

primrec the_default :: “’a ⇒ ’a option ⇒ ’a” where
“the_default d None” = d
| “the_default _ (Some a)” = a”

primrec mirror :: “(’a, ’b, ’c) triple ⇒ (’c, ’b, ’a) triple” where
“mirror (Triple a b c)” = Triple c b a”

The equations can be specified in any order, and it is acceptable to leave out some cases, which are then unspecified. Pattern matching on the left-hand side is restricted to a single datatype, which must correspond to the same argument in all equations.

3.1.2 Simple Recursion

For simple recursive types, recursive calls on a constructor argument are allowed on the right-hand side:

primrec replicate :: “nat ⇒ ’a ⇒ ’a list” where
“replicate Zero _” = []
| “replicate (Succ n) x” = x # replicate n x”

primrec (nonexhaustive) at :: “’a list ⇒ nat ⇒ ’a” where
Defining Primitively Recursive Functions

"\( \text{at} (x \# xs) \ j = \)
\( \text{(case } j \text{ of}
\quad \text{Zero} \Rightarrow x
\quad \text{| Succ } j' \Rightarrow \text{at } xs \ j') \)"

\[ \text{primrec } \text{tfold} :: \left( \begin{array}{l}
\left( 'a \Rightarrow 'b \Rightarrow 'b \right) \Rightarrow \left( 'a, 'b \right) \Rightarrow 'b \\
\end{array} \right) \text{ while}
\quad \text{tfold } _{\text{TNil } y} = y
\quad \text{| tfold } f \ (TCons \ x \ xs) = f \ x \ (\text{tfold } f \ xs) \]

Pattern matching is only available for the argument on which the recursion takes place. Fortunately, it is easy to generate pattern-maching equations using the \textsf{simps\_of\_case} command provided by the theory ~/src/HOL/Library/Simps\_Case\_Conv.thy.

\[ \text{simps\_of\_case } \text{at\_simps\_alt: at.simps} \]

This generates the lemma collection \text{at\_simps\_alt}:

\[ \text{at} (x \# xs) \ \text{Zero} = x \quad \text{at} (xa \# xs) \ (\text{Succ } x) = at \ xs \ x \]

The next example is defined using \textsf{fun} to escape the syntactic restrictions imposed on primitively recursive functions:

\[ \text{fun } \text{at\_least\_two} :: "nat } \Rightarrow \text{ bool" where}
\quad \text{at\_least\_two } (\text{Succ (Succ _)}) \leftrightarrow \text{ True"}
\quad \text{| at\_least\_two } _{\leftrightarrow } \text{ False"}

3.1.3 Mutual Recursion

The syntax for mutually recursive functions over mutually recursive data-types is straightforward:

\[ \text{primrec}
\quad \text{nat\_of\_even\_nat } :: "even\_nat } \Rightarrow \text{ nat" and}
\quad \text{nat\_of\_odd\_nat } :: "odd\_nat } \Rightarrow \text{ nat" where}
\quad \text{nat\_of\_even\_nat } \text{Even\_Zero } = \text{ Zero"}
\quad \text{| nat\_of\_even\_nat } (\text{Even\_Succ } n) = \text{ Succ } (\text{nat\_of\_odd\_nat } n)"
\quad \text{| nat\_of\_odd\_nat } (\text{Odd\_Succ } n) = \text{ Succ } (\text{nat\_of\_even\_nat } n)"

\[ \text{primrec}
\quad \text{eval}_{e} :: "(a } \Rightarrow \text{ int) } \Rightarrow (b } \Rightarrow \text{ int) } \Rightarrow (a, b) \Rightarrow \text{ int" and}
\quad \text{eval}_{t} :: "(a } \Rightarrow \text{ int) } \Rightarrow (b } \Rightarrow \text{ int) } \Rightarrow (a, b) \Rightarrow \text{ int" and}
\quad \text{eval}_{f} :: "(a } \Rightarrow \text{ int) } \Rightarrow (b } \Rightarrow \text{ int) } \Rightarrow (a, b) \Rightarrow \text{ int" where}
\quad \text{eval}_{e} \gamma \xi \ (\text{Term } t) = \text{eval}_{t} \gamma \xi \ t"
\quad \text{| eval}_{e} \gamma \xi \ (\text{Sum } t \ c) = \text{eval}_{t} \gamma \xi \ t + \text{eval}_{e} \gamma \xi \ c"
3 Defining Primitively Recursive Functions

| “eval\_\gamma\xi (Factor f) = eval\_\gamma\xi f” |
| “eval\_\gamma\xi (Prod f t) = eval\_\gamma\xi f + eval\_\gamma\xi t” |
| “eval\_\gamma\_ (Const a) = \gamma a” |
| “eval\_\gamma\xi (Var b) = \xi b” |
| “eval\_\gamma\xi (Expr e) = eval\_e \gamma\xi e” |

Mutual recursion is possible within a single type, using fun:

```
fun
  even :: “nat ⇒ bool” and
  odd :: “nat ⇒ bool”
where
  “even Zero = True” |
  “even (Suc n) = odd n” |
  “odd Zero = False” |
  “odd (Suc n) = even n”
```

3.1.4 Nested Recursion

In a departure from the old datatype package, nested recursion is normally handled via the map functions of the nesting type constructors. For example, recursive calls are lifted to lists using map:

```
primrec at\_ff :: “’a tree\_ff ⇒ nat list ⇒ ’a” where
  “at\_ff (Node\_ff a ts) js =
    (case js of
      [] ⇒ a
    | j # js’ ⇒ at\_ff (map\_λ t js’) ts) j”
```

The next example features recursion through the option type. Although option is not a new-style datatype, it is registered as a BNF with the map function map\_option:

```
primrec sum\_btree :: “’(a::{zero, plus}) btree ⇒ ’a” where
  “sum\_btree (BNode a lt rt) =
    a + the\_default 0 (map\_option sum\_btree lt) +
    the\_default 0 (map\_option sum\_btree rt)”
```

The same principle applies for arbitrary type constructors through which recursion is possible. Notably, the map function for the function type (⇒) is simply composition ((○)):

```
primrec relabel\_ft :: “’(a ⇒ ’a) ⇒ ’a ftree ⇒ ’a ftree” where
  “relabel\_ft f (FTLeaf x) = FTLeaf (f x)” |
  “relabel\_ft f (FTNode g) = FTNode (relabel\_ft f ○ g)”
```

For convenience, recursion through functions can also be expressed using λ-abstractions and function application rather than through composition. For
example:

```
primrec relabel_ft :: "'a ⇒ 'a ftree ⇒ 'a ftree" where
    "relabel_ft f (FTLeaf x) = FTLeaf (f x)"
 | "relabel_ft f (FTNode g) = FTNode (λx. relabel_ft f (g x))"
```

```
primrec (nonexhaustive) subtree_ft :: "'a ⇒ 'a ftree ⇒ 'a ftree" where
    "subtree_ft x (FTNode g) = g x"
```

For recursion through curried $n$-ary functions, $n$ applications of $(•)$ are necessary. The examples below illustrate the case where $n = 2$:

```
datatype 'a ftree2 = FTLeaf2 'a | FTNode2 "'a ⇒ 'a ⇒ 'a ftree2"
```

```
primrec relabel_ft2 :: "('a ⇒ 'a) ⇒ 'a ftree2 ⇒ 'a ftree2" where
    "relabel_ft2 f (FTLeaf2 x) = FTLeaf2 (f x)"
 | "relabel_ft2 f (FTNode2 g) = FTNode2 (λx y. relabel_ft2 f) g"
```

```
primrec relabel_ft2 :: "('a ⇒ 'a) ⇒ 'a ftree2 ⇒ 'a ftree2" where
    "relabel_ft2 f (FTLeaf2 x) = FTLeaf2 (f x)"
 | "relabel_ft2 f (FTNode2 g) = FTNode2 (λx y. relabel_ft2 f (g x y))"
```

```
primrec (nonexhaustive) subtree_ft2 :: "'a ⇒ 'a ⇒ 'a ftree2" where
    "subtree_ft2 x y (FTNode2 g) = g x y"
```

For any datatype featuring nesting, the predicator can be used instead of the map function, typically when defining predicates. For example:

```
primrec increasing_tree :: "int ⇒ int tree ⇒ bool" where
    "increasing_tree m (Node ff f f n ts) ←→
      n ≥ m ∧ list_all (increasing_tree (n + 1)) ts"
```

### 3.1.5 Nested-as-Mutual Recursion

For compatibility with the old package, but also because it is sometimes convenient in its own right, it is possible to treat nested recursive datatypes as mutually recursive ones if the recursion takes place though new-style datatypes. For example:

```
primrec (nonexhaustive)
  atff :: "'a treeff ⇒ nat list ⇒ 'a" and
  atsff :: "'a treeff list ⇒ nat ⇒ nat list ⇒ 'a"
where
  "atff (Nodeff a ts) js =
    (case js of
      [] ⇒ a
    | j # js' ⇒ atsff ts j js')"
 | "atsff (t # ts) j =
```
(case j of
    Zero ⇒ at\_ff t
  | Succ j' ⇒ ats\_ff ts j')"

Appropriate induction rules are generated as \( at\_ff.\_induct \), \( ats\_ff.\_induct \), and \( ats\_ff.\_ats\_ff.\_induct \). The induction rules and the underlying recursors are generated dynamically and are kept in a cache to speed up subsequent definitions.

Here is a second example:

\begin{verbatim}
primrec
  sum\_btree :: "('a::{zero,plus}) btree ⇒ 'a" and
  sum\_btree\_option :: "'a btree option ⇒ 'a"
where
  "sum\_btree (BNode a lt rt) =
    a + sum\_btree\_option lt + sum\_btree\_option rt"
  "sum\_btree\_option None = 0"
  "sum\_btree\_option (Some t) = sum\_btree t"
\end{verbatim}

### 3.2 Command Syntax

#### 3.2.1 primrec

\texttt{primrec} : \textit{local\_theory} \(\rightarrow\) \textit{local\_theory}
The \texttt{primrec} command introduces a set of mutually recursive functions over datatypes.

The syntactic entity \texttt{target} can be used to specify a local context, \texttt{fixes} denotes a list of names with optional type signatures, \texttt{thmdecl} denotes an optional name for the formula that follows, and \texttt{prop} denotes a HOL proposition [12].

The optional target is optionally followed by a combination of the following options:

- The \texttt{plugins} option indicates which plugins should be enabled (\textit{only}) or disabled (\textit{del}). By default, all plugins are enabled.
- The \texttt{nonexhaustive} option indicates that the functions are not necessarily specified for all constructors. It can be used to suppress the warning that is normally emitted when some constructors are missing.
- The \texttt{transfer} option indicates that an unconditional transfer rule should be generated and proved by \texttt{transfer\_prover}. The \texttt{[transfer\_rule]} attribute is set on the generated theorem.

### 3.3 Generated Theorems

The \texttt{primrec} command generates the following properties (listed for \texttt{tfold}):
3 Defining Primitively Recursive Functions

\[ f.\text{simps} \ [\text{simp}, \text{code}]: \]
\[ \text{tfold } uu \ (TNil \ y) = y \]
\[ \text{tfold } f \ (TCons \ x \ xs) = f \ x \ (\text{tfold } f \ xs) \]
The [code] attribute is set by the code plugin (Section 8.1).

\[ f.\text{transfer} \ [\text{transfer\_rule}]: \]
\[ \text{rel\_fun } (\text{rel\_fun } R2 \ (\text{rel\_fun } R1 \ R1)) \ (\text{rel\_fun } (\text{rel\_tlist } R2 \ R1) \ R1) \ \text{tfold } \text{tfold} \]
This theorem is generated by the transfer plugin (Section 8.3) for functions declared with the transfer option enabled.

\[ f.\text{induct} \ [\text{case\_names } C_1 \ldots C_n]: \]
This induction rule is generated for nested-as-mutual recursive functions (Section 3.1.5).

\[ f_1\ldots f_m.\text{induct} \ [\text{case\_names } C_1 \ldots C_n]: \]
This induction rule is generated for nested-as-mutual recursive functions (Section 3.1.5). Given \( m > 1 \) mutually recursive functions, this rule can be used to prove \( m \) properties simultaneously.

3.4 Recursive Default Values for Selectors

A datatype selector \( \text{un}_D \) can have a default value for each constructor on which it is not otherwise specified. Occasionally, it is useful to have the default value be defined recursively. This leads to a chicken-and-egg situation, because the datatype is not introduced yet at the moment when the selectors are introduced. Of course, we can always define the selectors manually afterward, but we then have to state and prove all the characteristic theorems ourselves instead of letting the package do it.

Fortunately, there is a workaround that relies on overloading to relieve us from the tedium of manual derivations:

1. Introduce a fully unspecified constant \( \text{un}_D_0 :: 'a \) using \text{consts}.
2. Define the datatype, specifying \( \text{un}_D_0 \) as the selector’s default value.
3. Define the behavior of \( \text{un}_D_0 \) on values of the newly introduced data-type using the \text{overloading} command.
4. Derive the desired equation on \( \text{un}_D \) from the characteristic equations for \( \text{un}_D_0 \).

The following example illustrates this procedure:

\text{consts} \ \text{termi}_0 :: 'a
4 Defining Codatatypes

Codatatypes can be specified using the \texttt{codatatype} command. The command is first illustrated through concrete examples featuring different flavors of corecursion. More examples can be found in the directory \texttt{~/src/HOL/Datatype_Examples}. The Archive of Formal Proofs also includes some useful codatatypes, notably for lazy lists [7].

4.1 Introductory Examples

4.1.1 Simple Corecursion

Non-corecursive codatatypes coincide with the corresponding datatypes, so they are rarely used in practice. \textit{Corecursive codatatypes} have the same
syntax as recursive datatypes, except for the command name. For example, here is the definition of lazy lists:

```ocaml
codatatype ('a) llist = 
  lnull: LNil 
| LCons (lhd: 'a) (llt: "'a llist")
for 
  map: lmap 
  rel: llist_all2 
  pred: llist_all
where 
  "llt LNil = LNil"
```

Lazy lists can be infinite, such as LCons 0 (LCons 0 (...)) and LCons 0 (LCons 1 (LCons 2 (...))). Here is a related type, that of infinite streams:

```ocaml
codatatype ('a) stream = 
  SCons (shd: 'a) (stl: "'a stream")
for 
  map: smap 
  rel: stream_all2
```

Another interesting type that can be defined as a codatatype is that of the extended natural numbers:

```ocaml
codatatype enat = EZero | ESucc enat
```

This type has exactly one infinite element, ESucc (ESucc (ESucc (...))), that represents ∞. In addition, it has finite values of the form ESucc (... (ESucc EZero)...).

Here is an example with many constructors:

```ocaml
codatatype 'a process = 
  Fail 
| Skip (cont: "'a process") 
| Action (prefix: 'a) (cont: "'a process") 
| Choice (left: "'a process") (right: "'a process")
```

Notice that the cont selector is associated with both Skip and Action.

### 4.1.2 Mutual Corecursion

The example below introduces a pair of mutually corecursive types:

```ocaml
codatatype even_enat = Even_EZero | Even_ESucc odd_enat
and odd_enat = Odd_ESucc even_enat
```
4.1.3 Nested Corecursion

The next examples feature nested corecursion:

\[
\text{codatatype } 'a \text{ tree}_i = \text{Node}_i (\text{lbl}_i : 'a) \ (\text{sub}_i : "'a tree}_i \ \text{llist}"
\]
\[
\text{codatatype } 'a \text{ tree}_i s = \text{Node}_i s (\text{lbl}_i s : 'a) \ (\text{sub}_i s : "'a tree}_i s \ \text{fset}"
\]
\[
\text{codatatype } 'a \text{ sm} = \text{SM} (\text{accept} : \text{bool}) \ (\text{trans} : "'a } \Rightarrow 'a \text{ sm})
\]

4.2 Command Syntax

4.2.1 codatatype

\[
\text{codatatype} \quad \text{codatatype} : \text{local\_theory} \rightarrow \text{local\_theory}
\]

Definitions of codatatypes have almost exactly the same syntax as for data-
types (Section 2.2). The \text{discs\_sels} option is superfluous because discrimi-
nators and selectors are always generated for codatatypes.

4.3 Generated Constants

Given a codatatype \(('a_1, \ldots, 'a_m) t\) with \(m > 0\) live type variables and \(n\) constructosrs \(t.\text{C}_1, \ldots, t.\text{C}_n\), the same auxiliary constants are generated as for datatypes (Section 2.3), except that the recursor is replaced by a dual concept:

Corecursor: \(t.\text{corec}_t\)

4.4 Generated Theorems

The characteristic theorems generated by \text{codatatype} are grouped in three broad categories:

- The \textit{free constructor theorems} (Section 2.4.1) are properties of the con-
structors and destructors that can be derived for any freely generated type.
The functorial theorems (Section 2.4.2) are properties of datatypes related to their BNF nature.

The coinductive theorems (Section 4.4.1) are properties of datatypes related to their coinductive nature.

The first two categories are exactly as for datatypes.

### 4.4.1 Coinductive Theorems

The coinductive theorems are listed below for 'a list:

- **t.coinduct** [consumes m, case_names t₁ ... tₘ, case_conclusion D₁ ... Dₙ, coinduct t₁]:
  
  \[
  \begin{align*}
  & [R \text{llist'}; \land \text{llist'}, R \text{llist'} \Rightarrow \text{lnull llist} = \text{lnull llist'} \land \\
  & (\neg \text{lnull llist} \rightarrow \neg \text{lnull llist'} \rightarrow \text{lhd llist} = \text{lhd llist'} \land R (\text{ltl llist}) \\
  & (\text{ltl llist'})] \implies \text{llist} = \text{llist'}
  \end{align*}
  \]

- **t.coinduct_strong** [consumes m, case_names t₁ ... tₘ, case_conclusion D₁ ... Dₙ]:
  
  \[
  \begin{align*}
  & [R \text{llist'}; \land \text{llist'}, R \text{llist'} \Rightarrow \text{lnull llist} = \text{lnull llist'} \land \\
  & (\neg \text{lnull llist} \rightarrow \neg \text{lnull llist'} \rightarrow \text{lhd llist} = \text{lhd llist'} \land (R (\text{ltl llist'}) \lor \text{ltl llist} = \text{ltl llist'}))] \implies \text{llist} = \text{llist'}
  \end{align*}
  \]

- **t.rel_coinduct** [consumes m, case_names t₁ ... tₘ, case_conclusion D₁ ... Dₙ, coinduct pred]:
  
  \[
  \begin{align*}
  & [P \text{x y}; \land \text{llist'}, P \text{llist'} \Rightarrow \text{lnull llist} = \text{lnull llist'} \land (\neg \text{lnull llist} \rightarrow \neg \text{lnull llist'} \rightarrow (\text{lhd llist} (\text{ltl llist'}) \land P (\text{lhd llist'})) \land \text{ltl llist'}) \implies \text{llist} = \text{llist'}]
  \end{align*}
  \]

- **t₁...ₘ.coinduct** [case_names t₁ ... tₘ, case_conclusion D₁ ... Dₙ]:
  
  \[
  \begin{align*}
  & [x \in \text{lset a}; \land z₁ z₂, P z₁ (\text{LCons z₁ z₂}); \land z₁ z₂ x a. \{x a \in \text{lset z₂; P x a z₂}] \Rightarrow P x a (\text{LCons z₁ z₂}) \Rightarrow P x a
  \end{align*}
  \]

Given m > 1 mutually corecursive codatatypes, these coinduction rules can be used to prove m properties simultaneously.

- **t₁...ₘ.set_induct** [case_names C₁ ... Cₙ, induct set: set₁...ₜ₁, ..., induct set: setₗ...ₜₘ]:
  
  \[
  \begin{align*}
  & [x \in \text{lset a; \land z₁ z₂, P z₁ (LCons z₁ z₂); \land z₁ z₂ x a. \{x a \in \text{lset z₂; P x a z₂}] \Rightarrow P x a (\text{LCons z₁ z₂}) \Rightarrow P x a
  \end{align*}
  \]

If m = 1, the attribute [consumes 1] is generated as well.
t.corec:
\[ p \ a \Rightarrow \text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a = \text{LNil} \]
\[ \neg p \ a \Rightarrow \text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a = \text{LCons} \ (q^{21} \ a) \ (\text{if} \ q^{22} \ a \ \text{then} \ g^{221} \ a \ \text{else} \ \text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ (g^{222} \ a)) \]

t.corec\_code [code]:
\[ \text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a = (\text{if} \ p \ a \ \text{then} \ \text{LNil} \ \text{else} \ \text{LCons} \ (q^{21} \ a) \ (\text{if} \ q^{22} \ a \ \text{then} \ g^{221} \ a \ \text{else} \ \text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ (g^{222} \ a))) \]
The [code] attribute is set by the code plugin (Section 8.1).

t.corec\_disc:
\[ p \ a \Rightarrow \text{lnull} \ (\text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a) \]
\[ \neg p \ a \Rightarrow \neg \text{lnull} \ (\text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a) \]

t.corec\_disc\_iff [simp]:
\[ \text{lnull} \ (\text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a) = p \ a \]
\[ (\neg p \ a) = \neg \text{lnull} \ (\text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a) \]

t.corec\_sel [simp]:
\[ \neg p \ a \Rightarrow \text{lhd} \ (\text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a) = g^{21} \ a \]
\[ \neg p \ a \Rightarrow \text{ltl} \ (\text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ a) = (\text{if} \ q^{22} \ a \ \text{then} \ g^{221} \ a \ \text{else} \ \text{corec\_llist} \ p \ g^{21} \ q^{22} \ g^{221} \ g^{222} \ (g^{222} \ a)) \]

t.map\_o\_corec:
\[ \text{lmap} \ f \circ \text{corec\_llist} \ g \ ga \ gb \ gc \ gd = \text{corec\_llist} \ g \ (f \circ ga) \ gb \ (\text{lmap} \ f \circ gc) \ gd \]

t.corec\_transfer [transfer\_rule]:
\[ \text{rel\_fun} \ (\text{rel\_fun} \ S \ (=)) \ (\text{rel\_fun} \ (\text{rel\_fun} \ S \ R) \ (\text{rel\_fun} \ (\text{rel\_fun} \ S \ (\text{llist\_all2} \ R)) \ (\text{rel\_fun} \ (\text{rel\_fun} \ S \ (\text{llist\_all2} \ R))))) \ (\text{corec\_llist} \ \text{corec\_llist}) \]
The [transfer\_rule] attribute is set by the transfer plugin (Section 8.3) for type constructors with no dead type arguments.

For convenience, codatatype also provides the following collection:

\[ t.simps = t.inject \ t.distinct \ t.case \ t.corec\_disc\_iff \ t.corec\_sel \ t.map \ t.rel\_inject \ t.rel\_distinct \ t.set \]

4.5 Antiquotation

4.5.1 codatatype

The codatatype antiquotation, written \texttt{\textless codatatype\texttt{\textgreater} \(t\)} or \texttt{@\{codatatype} \texttt{\textgreater} \texttt{t\}}, where \textit{t} is a type name, expands to \LaTeX{} code for the definition of
the codatatype, with each constructor listed with its argument types. For example, if \( t \) is \( \text{llist} \):

\[
\text{codatatype } a \text{llist} = \text{LNil} \mid \text{LCons } a \ (a \text{llist})
\]

5 Defining Primitively Corecursive Functions

Corecursive functions can be specified using the \texttt{primcorec} and \texttt{primcorecursive} commands, which support primitive corecursion. Other approaches include the more general \texttt{partial_function} command, the \texttt{corec} and \texttt{corecursive} commands, and techniques based on domains and topologies [8]. In this tutorial, the focus is on \texttt{primcorec} and \texttt{primcorecursive}; \texttt{corec} and \texttt{corecursive} are described in a separate tutorial [3]. More examples can be found in the directories \~/src/HOL/Datatype_Examples and \~/src/HOL/Corec_Examples.

Whereas recursive functions consume datatypes one constructor at a time, corecursive functions construct codatatypes one constructor at a time. Partly reflecting a lack of agreement among proponents of coalgebraic methods, Isabelle supports three competing syntaxes for specifying a function \( f \):

- The \textit{destructor view} specifies \( f \) by implications of the form
  \[
  \ldots \implies \text{is}_{C_j}(f \ x_1 \ldots \ x_n)
  \]
  and equations of the form
  \[
  \text{un}_{C_j}i\ (f \ x_1 \ldots \ x_n) = \ldots
  \]
  This style is popular in the coalgebraic literature.

- The \textit{constructor view} specifies \( f \) by equations of the form
  \[
  \ldots \implies f \ x_1 \ldots \ x_n = C_j \ldots
  \]
  This style is often more concise than the previous one.

- The \textit{code view} specifies \( f \) by a single equation of the form
  \[
  f \ x_1 \ldots \ x_n = \ldots
  \]
  with restrictions on the format of the right-hand side. Lazy functional programming languages such as Haskell support a generalized version of this style.

All three styles are available as input syntax. Whichever syntax is chosen, characteristic theorems for all three styles are generated.
5.1 Introductory Examples

Primitive corecursion is illustrated through concrete examples based on the codatatypes defined in Section 4.1. More examples can be found in the directory ~/src/HOL/Datatype_Examples. The code view is favored in the examples below. Sections 5.1.5 and 5.1.6 present the same examples expressed using the constructor and destructor views.

5.1.1 Simple Corecursion

Following the code view, corecursive calls are allowed on the right-hand side as long as they occur under a constructor, which itself appears either directly to the right of the equal sign or in a conditional expression:

```
primcorec literate :: "('a ⇒ 'a) ⇒ 'a ⇒ 'a llist" where
  "literate g x = LCons x (literate g (g x))"
```

```
primcorec siterate ::="('a ⇒ 'a) ⇒ 'a ⇒ 'a stream" where
  "siterate g x = SCons x (siterate g (g x))"
```

The constructor ensures that progress is made---i.e., the function is productive. The above functions compute the infinite lazy list or stream [x, g x, g (g x), ...]. Productivity guarantees that prefixes [x, g x, g (g x), ..., (g ^^ k) x] of arbitrary finite length k can be computed by unfolding the code equation a finite number of times.

Corecursive functions construct codatatype values, but nothing prevents them from also consuming such values. The following function drops every second element in a stream:

```
primcorec every_snd :: "'a stream ⇒ 'a stream" where
  "every_snd s = SCons (shd s) (stl (stl s))"
```

Constructs such as let–in, if–then–else, and case–of may appear around constructors that guard corecursive calls:

```
primcorec lapp :: "'a llist ⇒ 'a llist ⇒ 'a llist" where
  "lapp xs ys =
    (case xs of
      LNil ⇒ ys
    | LCons x xs' ⇒ LCons x (lapp xs' ys))"
```

For technical reasons, case–of is only supported for case distinctions on (co)datatypes that provide discriminators and selectors.

Pattern matching is not supported by primcorec. Fortunately, it is easy to generate pattern-matching equations using the simps_of_case command provided by the theory ~/src/HOL/Library/Simps_Case_Conv.thy.
5 Defining Primitively Corecursive Functions

simps_of_case lapp_simps: lapp.code

This generates the lemma collection lapp_simps:

\[
\begin{align*}
lapp \, \text{LNil} \, \text{ys} &= \text{ys} \\
lapp \, (\text{LCons} \, \text{xa} \, x) \, \text{ys} &= \text{LCons} \, \text{xa} \, (\text{lapp} \, x \, \text{ys})
\end{align*}
\]

Corecursion is useful to specify not only functions but also infinite objects:

primcorec infty :: enat where
  “infty = ESucc infty”

The example below constructs a pseudorandom process value. It takes a stream of actions \(s\), a pseudorandom function generator \(f\), and a pseudorandom seed \(n\):

primcorec random_process :: “a stream ⇒ (int ⇒ int) ⇒ int ⇒ 'a process”
where
  “random_process \ s \ f \ n =
  (if \ n \ mod \ 4 = 0 \ then
   Fail
  else if \ n \ mod \ 4 = 1 \ then
   Skip \ (random_process \ s \ f \ (f \ n))
  else if \ n \ mod \ 4 = 2 \ then
   Action \ (shd \ s) \ (random_process \ (stl \ s) \ f \ (f \ n))
  else
   Choice \ (random_process \ (every_snd \ s) \ (f \circ f) \ (f \ n))
   (random_process \ (every_snd \ (stl \ s)) \ (f \circ f) \ (f \ (f \ n))))”

The main disadvantage of the code view is that the conditions are tested sequentially. This is visible in the generated theorems. The constructor and destructor views offer nonsequential alternatives.

5.1.2 Mutual Corecursion

The syntax for mutually corecursive functions over mutually corecursive datatypes is unsurprising:

primcorec
  even_infty :: even_enat and
  odd_infty :: odd_enat
where
  “even_infty = Even_ESucc odd_infty”
  | “odd_infty = Odd_ESucc even_infty”
5.1.3 Nested Corecursion

The next pair of examples generalize the `literate` and `siterate` functions (Section 5.1.3) to possibly infinite trees in which subnodes are organized either as a lazy list (`tree_i llist`) or as a finite set (`tree_i fset`). They rely on the map functions of the nesting type constructors to lift the corecursive calls:

```
primcorec iterate_i_l :: "'(a ⇒ 'a llist) ⇒ 'a ⇒ 'a tree_i llist" where
  "iterate_i_l g x = Node_i_l x (lmap (iterate_i_l) (g x))"
```

```
primcorec iterate_i_s :: "'(a ⇒ 'a fset) ⇒ 'a ⇒ 'a tree_i fset" where
  "iterate_i_s g x = Node_i_s x (fimage (iterate_i_s) (g x))"
```

Both examples follow the usual format for constructor arguments associated with nested recursive occurrences of the datatype. Consider `iterate_i_l`. The term `g x` constructs an `'a llist` value, which is turned into an `'a tree_i llist` value using `lmap`. This format may sometimes feel artificial. The following function constructs a tree with a single, infinite branch from a stream:

```
primcorec tree_i_l_of_stream :: "'a stream ⇒ 'a tree_i llist" where
  "tree_i_l_of_stream s = Node_i_l (shd s) (lmap tree_i_l_of_stream (LCons (stl s) LNil))"
```

A more natural syntax, also supported by Isabelle, is to move corecursive calls under constructors:

```
primcorec tree_i_l_of_stream :: "'a stream ⇒ 'a tree_i llist" where
  "tree_i_l_of_stream s = Node_i_l (shd s) (LCons (tree_i_l_of_stream (stl s)) LNil)"
```

The next example illustrates corecursion through functions, which is a bit special. Deterministic finite automata (DFAs) are traditionally defined as 5-tuples `(Q, Σ, δ, q₀, F)`, where `Q` is a finite set of states, `Σ` is a finite alphabet, `δ` is a transition function, `q₀` is an initial state, and `F` is a set of final states. The following function translates a DFA into a state machine:

```
primcorec sm_of_dfa :: "'(q ⇒ 'a ⇒ 'q) ⇒ 'q set ⇒ 'q ⇒ 'a sm" where
  "sm_of_dfa δ F q = SM (q ∈ F) (sm_of_dfa δ F o δ q)"
```

The map function for the function type `(⇒)` is composition `( ○ )`. For convenience, corecursion through functions can also be expressed using `λ`-abstractions and function application rather than through composition. For example:

```
primcorec sm_of_dfa :: "'(q ⇒ 'a ⇒ 'q) ⇒ 'q set ⇒ 'q ⇒ 'a sm" where
  "sm_of_dfa δ F q = SM (q ∈ F) (λa. sm_of_dfa δ F (δ q a))"
```

```
primcorec empty_sm :: "'a sm" where
```
"empty_sm = SM False (∧. empty_sm)"

\textbf{primcorec} \texttt{not_sm} :: "a sm ⇒ 'a sm" where
"not_sm M = SM (∼ accept M) (∧. not_sm (trans M a))"

\textbf{primcorec} \texttt{or_sm} :: "a sm ⇒ 'a sm ⇒ 'a sm" where
"or_sm M N = SM (accept M ∨ accept N) (∧. or_sm (trans M a) (trans N a))"

For recursion through curried \( n \)-ary functions, \( n \) applications of \((\circ)\) are necessary. The examples below illustrate the case where \( n = 2 \):

\textbf{codatatype} ('a, 'b) \texttt{sm}2 =
SM2 (accept2 : bool) (trans2: "'a ⇒ 'b ⇒ ('a, 'b) sm2")

\textbf{primcorec}
\texttt{sm}2\_of_dfa :: "('q ⇒ 'a ⇒ 'b ⇒ 'q) ⇒ 'q set ⇒ 'q ⇒ ('a, 'b) sm2"
where
"\texttt{sm}2\_of_dfa δ F q = SM2 (q ∈ F) ((\circ) ((\circ) (\texttt{sm}2\_of_dfa δ F)) (δ q))"

\textbf{primcorec}
\texttt{sm}2\_of_dfa :: "('q ⇒ 'a ⇒ 'b ⇒ 'q) ⇒ 'q set ⇒ 'q ⇒ ('a, 'b) sm2"
where
"\texttt{sm}2\_of_dfa δ F q = SM2 (q ∈ F) (∧. b. \texttt{sm}2\_of_dfa δ F (δ q a b))"

\section*{5.1.4 Nested-as-Mutual Corecursion}

Just as it is possible to recurse over nested recursive datatypes as if they were mutually recursive (Section 3.1.5), it is possible to pretend that nested codatatypes are mutually corecursive. For example:

\textbf{primcorec}
\texttt{iterate}ii : "('a ⇒ 'a llist) ⇒ 'a ⇒ 'a treeii" and
\texttt{iterates}ii :: "('a ⇒ 'a llist) ⇒ 'a llist ⇒ 'a treeii llist"
where
"\texttt{iterate}ii g x = Node\_ii x (\texttt{iterates}ii g (g x))"
| "\texttt{iterates}ii g xs =
| (case xs of
| LNil ⇒ LNil
| LCons x xs′ ⇒ LCons (\texttt{iterate}ii g x) (\texttt{iterates}ii g xs′))"

Coinduction rules are generated as \texttt{iterate}ii\_coinduct, \texttt{iterates}ii\_coinduct, and \texttt{iterate}ii\_iteratesii\_coinduct and analogously for \texttt{coinduct\_strong}. These rules and the underlying corecurors are generated dynamically and are kept in a cache to speed up subsequent definitions.
5.1.5 Constructor View

The constructor view is similar to the code view, but there is one separate conditional equation per constructor rather than a single unconditional equation. Examples that rely on a single constructor, such as `literate` and `siterate`, are identical in both styles.

Here is an example where there is a difference:

```
primcorec lapp :: "a list ⇒ 'a list ⇒ 'a list" where
  "null xs ⇒ null ys ⇒ lapp xs ys = LNil"
| "_ ⇒ lapp xs ys = LCons (hd (if null xs then ys else xs))
  (if xs = LNil then ltl ys else lapp (ltl xs) ys)"
```

With the constructor view, we must distinguish between the `LNil` and the `LCons` case. The condition for `LCons` is left implicit, as the negation of that for `LNil`.

For this example, the constructor view is slightly more involved than the code equation. Recall the code view version presented in Section 5.1.1. The constructor view requires us to analyze the second argument (`ys`). The code equation generated from the constructor view also suffers from this.

In contrast, the next example is arguably more naturally expressed in the constructor view:

```
primcorec
  random_process :: "a stream ⇒ (int ⇒ int) ⇒ int ⇒ 'a process"
where
  "n mod 4 = 0 ⇒ random_process s f n = Fail"
| "n mod 4 = 1 ⇒
    random_process s f n = Skip (random_process s f (f n))"
| "n mod 4 = 2 ⇒
    random_process s f n = Action (shd s) (random_process (stl s) f (f n))"
| "n mod 4 = 3 ⇒
    random_process s f n = Choice (random_process (every_snd s) f (f n))
    (random_process (every_snd (stl s)) f (f n))"
```

Since there is no sequentiality, we can apply the equation for `Choice` without having first to discharge `n mod 4 ≠ 0`, `n mod 4 ≠ 1`, and `n mod 4 ≠ 2`. The price to pay for this elegance is that we must discharge exclusiveness proof obligations, one for each pair of conditions (`n mod 4 = i, n mod 4 = j`) with `i < j`. If we prefer not to discharge any obligations, we can enable the `sequential` option. This pushes the problem to the users of the generated properties.
5.1.6 Destructor View

The destructor view is in many respects dual to the constructor view. Conditions determine which constructor to choose, and these conditions are interpreted sequentially or not depending on the *sequential* option. Consider the following examples:

```haskell
primcorec literate :: "'a ⇒ 'a ⇒ 'a llist" where
  "⇒ lnull (literate _ x)"
| "lhd (literate _ x) = x"
| "ltl (literate g x) = literate g (g x)"

primcorec siterate :: "'a ⇒ 'a ⇒ 'a stream" where
  "shd (siterate _ x) = x"
| "stl (siterate g x) = siterate g (g x)"

primcorec every_snd :: "a stream ⇒ 'a stream" where
  "shd (every_snd s) = shd s"
| "stl (every_snd s) = stl (stl s)"
```

The first formula in the *local.literate* specification indicates which constructor to choose. For *local.siterate* and *local.every_snd*, no such formula is necessary, since the type has only one constructor. The last two formulas are equations specifying the value of the result for the relevant selectors. Corecursive calls appear directly to the right of the equal sign. Their arguments are unrestricted.

The next example shows how to specify functions that rely on more than one constructor:

```haskell
primcorec lapp :: "'a llist ⇒ 'a llist ⇒ 'a llist" where
  "⇒ lnull xs ⇒ lnull ys ⇒ lnull (lapp xs ys)"
| "lhd (lapp xs ys) = lhd (if lnull xs then ys else xs)"
| "ltl (lapp xs ys) = (if xs = LNil then ltl ys else lapp (ltl xs) ys)"
```

For a codatatype with *n* constructors, it is sufficient to specify *n − 1* disriminator formulas. The command will then assume that the remaining constructor should be taken otherwise. This can be made explicit by adding

"⇒ ¬ lnull (lapp xs ys)"

to the specification. The generated selector theorems are conditional.

The next example illustrates how to cope with selectors defined for several constructors:

```haskell
primcorec
  random_process :: "a stream ⇒ (int ⇒ int) ⇒ int ⇒ 'a process" where
  "n mod 4 = 0 ⇒ random_process s f n = Fail"
```

“n mod 4 = 1 ⇒ is_Skip \left( \text{random\_process } s f n \right)”
“n mod 4 = 2 ⇒ is_Action \left( \text{random\_process } s f n \right)”
“n mod 4 = 3 ⇒ is_Choice \left( \text{random\_process } s f n \right)”
“cont \left( \text{random\_process } s f n \right) = \text{random\_process } s f \left( f n \right)” of Skip
“prefix \left( \text{random\_process } s f n \right) = \text{shd } s”
“cont \left( \text{random\_process } s f n \right) = \text{random\_process } \left( \text{every\_snd } s \right) f \left( f n \right)”
“right \left( \text{random\_process } s f n \right) = \text{random\_process } \left( \text{every\_snd } \left( \text{stl } s \right) \right) f \left( f n \right)”

Using the of keyword, different equations are specified for cont depending on which constructor is selected.

Here are more examples to conclude:

\begin{verbatim}
primcorec
even_infty :: even_enat and
odd_infty :: odd_enat
where
“even_infty \neq \text{Even\_EZero}”
| “un\_Even\_ESucc even_infty = odd_infty”
| “un\_Odd\_ESucc odd_infty = even_infty”
primcorec iteratei :: (\text{'a } \Rightarrow \text{'a list}) \Rightarrow \text{'a } \Rightarrow \text{'a treei}” where
“\text{lbli} \left( \text{iteratei } g x \right) = x”
| “\text{subi} \left( \text{iteratei } g x \right) = \text{lmap} \left( \text{iteratei } g \right) \left( g x \right)”
\end{verbatim}

5.2 Command Syntax

5.2.1 primcorec and primcorecursive

\begin{verbatim}
primcorec :: local\_theory \rightarrow local\_theory
primcorecursive :: local\_theory \rightarrow \text{proof(prove)}
\end{verbatim}
The `primcorec` and `primcorecursive` commands introduce a set of mutually corecursive functions over codatatypes.

The syntactic entity `target` can be used to specify a local context, `fixes` denotes a list of names with optional type signatures, `thmdecl` denotes an optional name for the formula that follows, and `prop` denotes a HOL proposition [12].

The optional target is optionally followed by a combination of the following options:

- The `plugins` option indicates which plugins should be enabled (only) or disabled (del). By default, all plugins are enabled.
- The `sequential` option indicates that the conditions in specifications expressed using the constructor or destructor view are to be interpreted sequentially.
- The `exhaustive` option indicates that the conditions in specifications expressed using the constructor or destructor view cover all possible cases. This generally gives rise to an additional proof obligation.
- The `transfer` option indicates that an unconditional transfer rule should be generated and proved by `transfer_prover`. The `[transfer_rule]` attribute is set on the generated theorem.
The \texttt{primcorec} command is an abbreviation for \texttt{primcorecursive} with \texttt{by auto?} to discharge any emerging proof obligations.

### 5.3 Generated Theorems

The \texttt{primcorec} and \texttt{primcorecursive} commands generate the following properties (listed for \texttt{literate}):

- \texttt{f.code [code]}:
  \[
  \text{literate } g \ x = \text{LCons } x \ (\text{literate } g \ (g \ x))
  \]
  The [\texttt{code}] attribute is set by the \texttt{code} plugin (Section 8.1).

- \texttt{f.ctr}:
  \[
  \text{literate } g \ x = \text{LCons } x \ (\text{literate } g \ (g \ x))
  \]

- \texttt{f.disc [simp, code]}:
  \[
  \neg \text{lnull } (\text{literate } g \ x)
  \]
  The [\texttt{code}] attribute is set by the \texttt{code} plugin (Section 8.1). The [\texttt{simp}] attribute is set only for functions for which \texttt{f.disc_iff} is not available.

- \texttt{f.disc_iff [simp]}:
  \[
  \neg \text{lnull } (\text{literate } g \ x)
  \]
  This property is generated only for functions declared with the \texttt{exhaustive} option or whose conditions are trivially exhaustive.

- \texttt{f.sel [simp, code]}:
  \[
  \neg \text{lnull } (\text{literate } g \ x)
  \]
  The [\texttt{code}] attribute is set by the \texttt{code} plugin (Section 8.1).

- \texttt{f.exclude}:
  These properties are missing for \texttt{literate} because no exclusiveness proof obligations arose. In general, the properties correspond to the discharged proof obligations.

- \texttt{f.exhaust}:
  This property is missing for \texttt{literate} because no exhaustiveness proof obligation arose. In general, the property correspond to the discharged proof obligation.

- \texttt{f.coinduct [consumes m, case_names t_1 \ldots \ t_m, case_conclusion D_1 \ldots D_n]}:
  This coinduction rule is generated for nested-as-mutual corecursive functions (Section 5.1.4).
This coinduction rule is generated for nested-as-mutual corecursive functions (Section 5.1.4).

Given \( m > 1 \) mutually corecursive functions, this rule can be used to prove \( m \) properties simultaneously.

For convenience, \texttt{primcorec} and \texttt{primcorecursive} also provide the following collection:

\[ f \cdot \texttt{simps} = f \cdot \texttt{disc iff} \text{ (or } f \cdot \texttt{disc)} \text{ t.sel} \]

6 Registering Bounded Natural Functors

The (co)datatype package can be set up to allow nested recursion through arbitrary type constructors, as long as they adhere to the BNF requirements and are registered as BNFs. It is also possible to declare a BNF abstractly without specifying its internal structure.

6.1 Bounded Natural Functors

Bounded natural functors (BNFs) are a semantic criterion for where (co)recursion may appear on the right-hand side of an equation [4,11].

An \( n \)-ary BNF is a type constructor equipped with a map function (functorial action), \( n \) set functions (natural transformations), and an infinite cardinal bound that satisfy certain properties. For example, \( 'a \) llist is a unary BNF. Its predicator \texttt{llist\_all} :: \( ('a \Rightarrow \textit{bool}) \Rightarrow 'a \texttt{llist} \Rightarrow \textit{bool} \) extends unary predicates over elements to unary predicates over lazy lists. Similarly, its relator \texttt{llist\_all2} :: \( ('a \Rightarrow 'b \Rightarrow \textit{bool}) \Rightarrow 'a \texttt{llist} \Rightarrow 'b \texttt{llist} \Rightarrow \textit{bool} \) extends binary predicates over elements to binary predicates over parallel lazy lists. The
cardinal bound limits the number of elements returned by the set function; it may not depend on the cardinality of \( 'a \).

The type constructors introduced by `datatype` and `codatatype` are automatically registered as BNFs. In addition, a number of old-style datatypes and non-free types are preregistered.

Given an \( n \)-ary BNF, the \( n \) type variables associated with set functions, and on which the map function acts, are `live`; any other variables are `dead`. Nested (co)recursion can only take place through live variables.

### 6.2 Introductory Examples

The example below shows how to register a type as a BNF using the `bnf` command. Some of the proof obligations are best viewed with the bundle `cardinal_syntax` included.

The type is simply a copy of the function space \( 'd \Rightarrow 'a \), where \( 'a \) is live and \( 'd \) is dead. We introduce it together with its map function, set function, predicator, and relator.

```plaintext
typedef \( ('d, 'a) \text{fn} = \text{"UNIV :: ('d ⇒ 'a) set"} \) by simp

setup_lifting type_definition_fn

lift_definition map_fn :: \( ('a ⇒ 'b) ⇒ ('d, 'a) \text{fn} ⇒ ('d, 'b) \text{fn} \) is \( \text{(o)} \) .

lift_definition set_fn :: \( ('d, 'a) \text{fn} ⇒ 'a \text{ set} \) is range .

lift_definition pred_fn :: \( ('a ⇒ \text{bool}) ⇒ ('d, 'a) \text{fn} ⇒ \text{bool} \) is
\( \text{"pred_fun (λ_. True)"} \) .

lift_definition rel_fn :: \( ('a ⇒ 'b ⇒ \text{bool}) ⇒ ('d, 'a) \text{fn} ⇒ ('d, 'b) \text{fn} ⇒ \text{bool} \) is
\( \text{"rel_fun (=)"} \) .

bnf \( ('d, 'a) \text{fn} \)

map: map_fn
sets: set_fn
bd: \( \text{"natLeq +c | UNIV :: 'd set|"} \)
rel: rel_fn
pred: pred_fn

proof –

show \( \text{"map_fn id = id"} \)
```
by transfer auto

next
fix f :: "'a ⇒ 'b" and g :: "'b ⇒ 'c"
show "map_fn (g ∘ f) = map_fn g ∘ map_fn f"
by transfer (auto simp add: comp_def)

next
fix F :: "'(d, 'a) fn" and f g :: "'a ⇒ 'b"
assume "∀x. x ∈ set_fn F ==⇒ f x = g x"
then show "map_fn f F = map_fn g F"
by transfer auto

next
fix F :: "'(d, 'a ⇒ 'b) fn"
show "card_order (natLeq +c ′d set|)"
apply (rule card_order_csum)
apply (rule natLeq_card_order)
by (rule card_of_card_order_on)

next
fix R :: "'a ⇒ 'b ⇒ bool" and S :: "'b ⇒ 'c ⇒ bool"
show "rel_fn R OO rel_fn S ≤ rel_fn (R OO S)"
by (rule, transfer) (auto simp add: rel_fun_def)

next
fix R :: "'a ⇒ 'b ⇒ bool"
show "rel_fn R = (λx y. ∃z. set_fn z ⊆ {(x, y). R x y} ∧ map_fn fst z = x ∧ map_fn snd z = y)"
unfolding fun_eq_iff relcompp.simps conversep.simps
by transfer (force simp: rel_fun_def subset_iff)

next
fix P :: "'a ⇒ bool"


show "pred_fn P = (\x. Ball (set_fn x) P)"

unfolding fun_eq_iff by transfer simp

qed

print_theorems
print_bnfs

Using print_theorems and print_bnfs, we can contemplate and show the world what we have achieved.

This particular example does not need any nonemptiness witness, because the one generated by default is good enough, but in general this would be necessary. See --/src/HOL/Basic_BNFs.thy, --/src/HOL/Library/Countable_Set_Type.thy, --/src/HOL/Library/FSet.thy, and --/src/HOL/Library/Multiset.thy for further examples of BNF registration, some of which feature custom witnesses.

For many typedefs and quotient types, lifting the BNF structure from the raw type to the abstract type can be done uniformly. This is the task of the lift_bnf command. Using lift_bnf, the above registration of \( (\alpha, \alpha) \) fn as a BNF becomes much shorter:

\[
\text{lift_bnf } (\alpha, \alpha) \text{ fn by force+}
\]

For type copies (typedefs with UNIV as the representing set), the proof obligations are so simple that they can be discharged automatically, yielding another command, copy_bnf, which does not emit any proof obligations:

\[
\text{copy_bnf } (\alpha, \alpha) \text{ fn}
\]

Since record schemas are type copies, copy_bnf can be used to register them as BNFs:

\[
\text{record } \alpha \text{ point } = \\
xval :: \alpha \\
yval :: \alpha
\]

\[
\text{copy_bnf } (\alpha, \beta) \text{ point_ext}
\]

In the general case, the proof obligations generated by lift_bnf are simpler than the actual BNF properties. In particular, no cardinality reasoning is required. Consider the following type of nonempty lists:

\[
\text{typedef } \alpha \text{ nonempty_list } = \{xs :: \alpha \text{ list}. \; xs \neq []\} \text{ by auto}
\]

The lift_bnf command requires us to prove that the set of nonempty lists is closed under the map function and the zip function. The latter only occurs implicitly in the goal, in form of the variable zs.

\[
\text{lift_bnf } \alpha \text{ nonempty_list}
\]
proof –
  fix \( f \) and \( xs :: "'a list" \)
  assume "\( xs \in \{ xs. \; xs \neq [] \} \)"
  then show "\( \text{map } f \; xs \in \{ xs. \; xs \neq [] \} \)"
    by (cases \( xs \)) auto
next
  fix \( zs :: (\'a \times \'b) \text{ list} \)
  assume "\( \text{map } \text{fst} \; zs \in \{ z\times. \; \text{fst } z\times \neq [] \} \)"
  "\( \text{map } \text{snd} \; zs \in \{ z\times. \; \text{snd } z\times \neq [] \} \)"
  then show \( \exists \; zs\times\in\{ z\times. \; \text{fst } z\times \neq [] \} \).
    \( \text{set } zs\times' \subseteq \text{set } zs \land \)
    \( \text{map } \text{fst} \; zs\times' = \text{map } \text{fst} \; zs \land \)
    \( \text{map } \text{snd} \; zs\times' = \text{map } \text{snd} \; zs \)"
    by (cases \( zs \)) (auto intro: exI[of _ zs])
qed

The \texttt{lift\_bnf} command also supports quotient types. Here is an example
that defines the option type as a quotient of the sum type. The proof obligations generated by \texttt{lift\_bnf} for quotients are different from the ones for typedefs. You can find additional examples of usages of \texttt{lift\_bnf} for both
quotients and subtypes in the session \texttt{HOL-Datatype\_Examples}.

\begin{verbatim}
inductive ignore\_Inl :: "'a + 'a \Rightarrow bool" where
  "ignore\_Inl (Inl x) (Inl y)"
| "ignore\_Inl (Inr x) (Inr x)"

lemma ignore\_Inl\_equivp:
  "ignore\_Inl x x"
  "ignore\_Inl x y \Longrightarrow ignore\_Inl y x"
  "ignore\_Inl x y \Longrightarrow ignore\_Inl y z \Longrightarrow ignore\_Inl x z"
  by (cases x; cases y; cases z; auto)+

quotient\_type 'a myoption = "'a + 'a" / ignore\_Inl
unfolding equiv\_reflp\_symp\_transp reflp\_def symp\_def transp\_def
by (blast intro: ignore\_Inl\_equivp)

lift\_bnf 'a myoption
proof –
  fix \( P :: "'a \Rightarrow 'b \Rightarrow bool" \) and \( Q :: "'b \Rightarrow 'c \Rightarrow bool" \)
  assume "\( P \; OO \; Q \neq \bot \)"
  then show "\( \text{rel\_sum } P \; P \; OO \; \text{ignore\_Inl } OO \; \text{rel\_sum } Q \; Q \leq \) \( \text{ignore\_Inl } OO \; \text{rel\_sum } (P \; OO \; Q) \; (P \; OO \; Q) \; OO \; \text{ignore\_Inl} \)"
    by (fastforce)
next
\end{verbatim}
fix \( S :: "a set set" \)
let \(?eq = \{(x, x'). ~\text{ignore}_\text{Inl} \ x \ x'\}\)"
let \(?in = "\forall x. \text{Basic\_BNFs} \text{.setl} \ x \ \cup \text{Basic\_BNFs} \text{.setr} \ x \ \subseteq A"\)"
assume "\(S \neq \{\}\)" "\(\bigcap S \neq \{\}\)"
show "\(\bigcap A \in S. \ ?eq \ " ?in \ A \ \subseteq \ ?eq \ " ?in \ (\bigcap S)"\)"
proof (intro subsetI)
  fix \(x\)
  assume "\(x \in (\bigcap A \in S. \ ?eq \ " ?in \ A)"\)
  with \(\bigcap S \neq \{\}\) show "\(x \in ?eq \ " ?in \ (\bigcap S)"\)"
  by (cases \(x\)) (fastforce)+
qed
qed

The next example declares a BNF axiomatically. This can be convenient for reasoning abstractly about an arbitrary BNF. The \texttt{bnf\_axiomatization} command below introduces a type \(\langle 'a, 'b, 'c \rangle \ F\), three set constants, a map function, a predicator, a relator, and a nonemptiness witness that depends only on \('a\). The type \( 'a \Rightarrow \langle 'a, 'b, 'c \rangle \ F\) of the witness can be read as an implication: Given a witness for \('a\), we can construct a witness for \(\langle 'a, 'b, 'c \rangle \ F\). The BNF properties are postulated as axioms.

\texttt{bnf\_axiomatization \(setA: 'a, setB: 'b, setC: 'c \rangle \ F\)}
\[ wits: "'a \Rightarrow \langle 'a, 'b, 'c \rangle \ F"\]

\texttt{print\_theorems}
\texttt{print\_bnfs}

\section*{6.3 Command Syntax}

\subsection*{6.3.1 bnf}

\texttt{bnf : local\_theory \rightarrow proof(prove)}
The \texttt{bnf} command registers an existing type as a bounded natural functor (BNF). The type must be equipped with an appropriate map function (functorial action). In addition, custom set functions, predicators, relators, and nonemptiness witnesses can be specified; otherwise, default versions are used.

The syntactic entity \texttt{target} can be used to specify a local context, \texttt{type} denotes a HOL type, and \texttt{term} denotes a HOL term \cite{12}.

The \texttt{plugins} option indicates which plugins should be enabled (\texttt{only}) or disabled (\texttt{del}). By default, all plugins are enabled.

6.3.2 \texttt{lift\_bnf}

\[
\text{lift\_bnf} : \text{local\_theory} \rightarrow \text{proof(prove)}
\]
The `lift_bnf` command registers as a BNF an existing type (the *abstract type*) that was defined as a subtype of a BNF (the *raw type*) using the `typedef` command or as a quotient type of a BNF (also, the *raw type*) using the `quotient_type`. To achieve this, it lifts the BNF structure on the raw type to the abstract type following a `type_definition` or a `Quotient` theorem. The theorem is usually inferred from the type, but can also be explicitly supplied by means of the optional `via` clause. In case of quotients, it is sometimes also necessary to supply a second theorem of the form `refl eq`, that expresses
the reflexivity (and thus totality) of the equivalence relation. In addition, custom names for the set functions, the map function, the predicator, and the relator, as well as nonemptiness witnesses can be specified.

Nonemptiness witnesses are not lifted from the raw type’s BNF, as this would be incomplete. They must be given as terms (on the raw type) and proved to be witnesses. The command warns about witness types that are present in the raw type’s BNF but not supplied by the user. The warning can be disabled by specifying the \textit{no\_warn\_wits} option.

\section{copy\_bnf}

\texttt{copy\_bnf} : \texttt{local\_theory} $\rightarrow$ \texttt{local\_theory}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{copy_bnf_diagram.png}
\caption{Diagram of the \texttt{copy\_bnf} command.}
\end{figure}

\texttt{cb-options}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cb_options_diagram.png}
\caption{Diagram of the \texttt{cb-options} annotation.}
\end{figure}

The \texttt{copy\_bnf} command performs the same lifting as \texttt{lift\_bnf} for type copies (\texttt{typedef}s with \textit{UNIV} as the representing set), without requiring the user to discharge any proof obligations or provide nonemptiness witnesses.

\section{bnf\_axiomatization}

\texttt{bnf\_axiomatization} : \texttt{local\_theory} $\rightarrow$ \texttt{local\_theory}
The \texttt{bnf_axiomatization} command declares a new type and associated constants (map, set, predicator, relator, and cardinal bound) and asserts the BNF properties for these constants as axioms.

The syntactic entity \texttt{target} can be used to specify a local context, \texttt{name} denotes an identifier, \texttt{typefree} denotes fixed type variable ('a, 'b, ...), \texttt{mixfix} denotes the usual parenthesized mixfix notation, and \texttt{types} denotes a space-separated list of types [12].

The \texttt{plugins} option indicates which plugins should be enabled (\texttt{only}) or disabled (\texttt{del}). By default, all plugins are enabled.

Type arguments are live by default; they can be marked as dead by entering \texttt{dead} in front of the type variable (e.g., (\texttt{dead 'a})) instead of an identifier for the corresponding set function. Witnesses can be specified by their types. Otherwise, the syntax of \texttt{bnf_axiomatization} is identical to the left-hand side of a \texttt{datatype} or \texttt{codatatype} definition.

The command is useful to reason abstractly about BNFs. The axioms are safe because there exist BNFs of arbitrary large arities. Applications must import the \texttt{-/src/HOL/Library/BNF_Axiomatization.thy} theory to use this functionality.
6.3.5 print_bnfs

\[ \text{print_bnfs} : \text{local\_theory} \rightarrow \]

---

7 Deriving Destructors and Theorems for Free Constructors

The derivation of convenience theorems for types equipped with free constructors, as performed internally by \texttt{datatype} and \texttt{codatatype}, is available as a stand-alone command called \texttt{free\_constructors}.

7.1 Command Syntax

7.1.1 free_constructors

\[ \text{free\_constructors} : \text{local\_theory} \rightarrow \text{proof(prove)} \]

---

\textbf{free\_constructors} \\
\textbf{dt-options} \\
\textbf{target} \\
\textbf{name} \\
\textbf{for} \\
\textbf{fc-ctor} \\
\textbf{where} \\
\textbf{prop}
The **free_constructors** command generates destructor constants for freely constructed types as well as properties about constructors and destructors. It also registers the constants and theorems in a data structure that is queried by various tools (e.g., **function**).

The syntactic entity *target* can be used to specify a local context, *name* denotes an identifier, *prop* denotes a HOL proposition, and *term* denotes a HOL term [12].

The syntax resembles that of **datatype** and **codatatype** definitions (Sections 2.2 and 4.2). A constructor is specified by an optional name for the discriminator, the constructor itself (as a term), and a list of optional names for the selectors.

Section 2.4 lists the generated theorems. For bootstrapping reasons, the generally useful **fundef_cong** attribute is not set on the generated **case_cong** theorem. It can be added manually using **declare**.

### 7.1.2 simps_of_case

```
simps_of_case : local_theory → local_theory
```

The diagram for **simps_of_case** illustrates the input and output types, as well as the relationship between the *target*, *name*, *thm*, and *splits* parameters.
The \texttt{simps\_of\_case} command provided by theory \texttt{~/src/HOL/Library/Simps\_Case\_Conv.thy} converts a single equation with a complex case expression on the right-hand side into a set of pattern-matching equations. For example,

\begin{verbatim}
  simps_of_case lapp_simps: lapp.code
\end{verbatim}

translates \texttt{lapp \textit{xs} \textit{ys} = (case \textit{xs} of \texttt{LNil} \Rightarrow \textit{ys} | \texttt{LCons \textit{x} \textit{x\textquotesingle} \Rightarrow \texttt{LCons \textit{x} (lapp \textit{x\textquotesingle} \textit{ys})})} into

\begin{align*}
  \texttt{lapp LNil \textit{ys} = \textit{ys}} \\
  \texttt{lapp (LCons \textit{xa} \textit{x}) \textit{ys} = LCons \textit{xa} (lapp \textit{x} \textit{ys})}
\end{align*}

### 7.1.3 case\_of\_simps

\begin{verbatim}
  case_of_simps : local\_theory \rightarrow local\_theory
\end{verbatim}

The \texttt{case\_of\_simps} command provided by theory \texttt{~/src/HOL/Library/Simps\_Case\_Conv.thy} converts a set of pattern-matching equations into single equation with a complex case expression on the right-hand side (cf. \texttt{simps\_of\_case}). For example,

\begin{verbatim}
  case_of_simps lapp_case: lapp_simps
\end{verbatim}

translates

\begin{align*}
  \texttt{lapp LNil \textit{ys} = \textit{ys}} \\
  \texttt{lapp (LCons \textit{xa} \textit{x}) \textit{ys} = LCons \textit{xa} (lapp \textit{x} \textit{ys})}
\end{align*}

into \texttt{lapp \textit{xb} \textit{x3a} = (case \textit{xb} of \texttt{LNil} \Rightarrow \textit{x3a} | \texttt{LCons \textit{x2} \textit{x1} \Rightarrow LCons \textit{x2} (lapp \textit{x1} \textit{x3a})}).
8 Selecting Plugins

Plugins extend the (co)datatype package to interoperate with other Isabelle packages and tools, such as the code generator, Transfer, Lifting, and Quickcheck. They can be enabled or disabled individually using the `plugins` option to the commands `datatype`, `primrec`, `codatatype`, `primcorec`, `primcorecursive`, `bnf`, `bnf_axiomatization`, and `free_constructors`. For example:

```
datatype (plugins del: code “quickcheck”) color = Red | Black
```

Beyond the standard plugins, the Archive of Formal Proofs includes a `derive` command that derives class instances of datatypes [10].

8.1 Code Generator

The `code` plugin registers freely generated types, including (co)datatypes, and (co)recursive functions for code generation. No distinction is made between datatypes and codatatypes. This means that for target languages with a strict evaluation strategy (e.g., Standard ML), programs that attempt to produce infinite codatatype values will not terminate.

For types, the plugin derives the following properties:

```
t.eq.refl [code nbe]:
  equal_class.equal x x ≡ True
```

```
t.eq.simps [code]:
  equal_class.equal [] (x21 ≠ x22) ≡ False
  equal_class.equal (x21 ≠ x22) [] ≡ False
  equal_class.equal (x21 ≠ x22) [] ≡ False
  equal_class.equal [] (x21 ≠ x22) ≡ False
  equal_class.equal (x21 ≠ x22) (y21 ≠ y22) ≡ x21 = y21 ∧ x22 = y22
  equal_class.equal [] [] ≡ True
```

In addition, the plugin sets the `[code]` attribute on a number of properties of freely generated types and of (co)recursive functions, as documented in Sections 2.4, 3.3, 4.4, and 5.3.

8.2 Size

For each datatype `t`, the `size` plugin generates a generic size function `t.size_t` as well as a specific instance `size :: t ⇒ nat` belonging to the `size` type class.
The **fun** command relies on **size** to prove termination of recursive functions on datatypes.

The plugin derives the following properties:

\[
\begin{align*}
t.size \ [\text{simp, code}]: & \quad \text{size}\_\text{list} \ x \ [] = 0 \\
& \quad \text{size}\_\text{list} \ x \ (x21 \# x22) = x \times x21 + \text{size}\_\text{list} \ x \ x22 + \text{Suc} \ 0 \\
& \quad \text{size} \ [] = 0 \\
& \quad \text{size} \ (x21 \# x22) = \text{size} \ x22 + \text{Suc} \ 0 \\
t.size\_\text{gen}: & \quad \text{size}\_\text{list} \ x \ [] = 0 \\
& \quad \text{size}\_\text{list} \ x \ (x21 \# x22) = x \times x21 + \text{size}\_\text{list} \ x \ x22 + \text{Suc} \ 0 \\
t.size\_\text{gen}\_\text{o}\_\text{map}: & \quad \text{size}\_\text{list} \ f \circ \text{map} \ g = \text{size}\_\text{list} \ (f \circ g) \\
t.size\_\text{neq}: & \quad \text{This property is missing for 'a list. If the size function always evaluates to a non-zero value, this theorem has the form size} \ x \neq 0.
\end{align*}
\]

The **t.size** and **t.size\_t** functions generated for datatypes defined by nested recursion through a datatype \(u\) depend on \(u.size_u\).

If the recursion is through a non-datatype \(u\) with type arguments 'a_1, \ldots, 'a_m, by default \(u\) values are given a size of 0. This can be improved upon by registering a custom size function of type ('a_1 \Rightarrow \text{nat}) \Rightarrow \ldots \Rightarrow ('a_m \Rightarrow \text{nat}) \Rightarrow u \Rightarrow \text{nat} using the ML function BNF\_LFP\_Size.register_size or BNF\_LFP\_Size.register_size\_global. See theory ~/src/HOL/Library/Multiset.thy for an example.

### 8.3 Transfer

For each (co)datatype with live type arguments and each manually registered BNF, the **transfer** plugin generates a predicator **t.pred\_t** and properties that guide the Transfer tool.

For types with at least one live type argument and no dead type arguments, the plugin derives the following properties:

\[
\begin{align*}
t.Domainp\_\text{rel} \ [\text{relator}\_\text{domain}]: & \quad \text{Domainp} \ (\text{list}\_\text{all}\_2 \ R) = \text{list}\_\text{all} \ (\text{Domainp} \ R) \\
t.left\_\text{total}\_\text{rel} \ [\text{transfer}\_\text{rule}]: & \quad \text{left}\_\text{total} \ R \implies \text{left}\_\text{total} \ (\text{list}\_\text{all}\_2 \ R)
\end{align*}
\]
\texttt{t.left\_unique\_rel} \texttt{[transfer\_rule]}:
\begin{equation}
\text{left\_unique } R \implies \text{left\_unique } (\text{list\_all2 } R)
\end{equation}

\texttt{t.right\_total\_rel} \texttt{[transfer\_rule]}:
\begin{equation}
\text{right\_total } R \implies \text{right\_total } (\text{list\_all2 } R)
\end{equation}

\texttt{t.right\_unique\_rel} \texttt{[transfer\_rule]}:
\begin{equation}
\text{right\_unique } R \implies \text{right\_unique } (\text{list\_all2 } R)
\end{equation}

\texttt{t.bi\_total\_rel} \texttt{[transfer\_rule]}:
\begin{equation}
\text{bi\_total } R \implies \text{bi\_total } (\text{list\_all2 } R)
\end{equation}

\texttt{t.bi\_unique\_rel} \texttt{[transfer\_rule]}:
\begin{equation}
\text{bi\_unique } R \implies \text{bi\_unique } (\text{list\_all2 } R)
\end{equation}

For (co)datatypes with at least one live type argument, the plugin sets the \texttt{[transfer\_rule]} attribute on the following (co)datatypes properties: \texttt{t.case\_transfer}, \texttt{t.sel\_transfer}, \texttt{t.ctr\_transfer}, \texttt{t.disc\_transfer}, \texttt{t.rec\_transfer}, and \texttt{t.corec\_transfer}. For (co)datatypes that further have no dead type arguments, the plugin sets \texttt{[transfer\_rule]} on \texttt{t.set\_transfer}, \texttt{t.map\_transfer}, and \texttt{t.rel\_transfer}.

For \texttt{primrec}, \texttt{primcorec}, and \texttt{primcorecursive}, the plugin implements the generation of the \texttt{f.transfer} property, conditioned by the \texttt{transfer} option, and sets the \texttt{[transfer\_rule]} attribute on these.

### 8.4 Lifting

For each (co)datatype and each manually registered BNF with at least one live type argument and no dead type arguments, the \texttt{lifting} plugin generates properties and attributes that guide the Lifting tool.

The plugin derives the following property:

\texttt{t.Quotient} \texttt{[quot\_map]}:
\begin{equation}
\text{Quotient } R \text{ Abs } \text{ Rep } T \implies \text{Quotient } (\text{list\_all2 } R) (\text{map } \text{Abs}) (\text{map } \text{Rep}) (\text{list\_all2 } T)
\end{equation}

In addition, the plugin sets the \texttt{[relator\_eq]} attribute on a variant of the \texttt{t.rel\_eq\_onp} property, the \texttt{[relator\_mono]} attribute on \texttt{t.rel\_mono}, and the \texttt{[relator\_distr]} attribute on \texttt{t.rel\_compp}.

### 8.5 Quickcheck

The integration of datatypes with Quickcheck is accomplished by the \texttt{quickcheck} plugin. It combines a number of subplugins that instantiate specific
type classes. The subplugins can be enabled or disabled individually. They are listed below:

```
quickcheck_random
quickcheck_exhaustive
quickcheck_bounded_forall
quickcheck_full_exhaustive
quickcheck_narrowing
```

### 8.6 Program Extraction

The *extraction* plugin provides realizers for induction and case analysis, to enable program extraction from proofs involving datatypes. This functionality is only available with full proof objects, i.e., with the *HOL-Proofs* session.

### 9 Known Bugs and Limitations

This section lists the known bugs and limitations of the (co)datatype package at the time of this writing.

1. *Defining mutually (co)recursive (co)datatypes can be slow.* Fortunately, it is always possible to recast mutual specifications to nested ones, which are processed more efficiently.

2. *Locally fixed types and terms cannot be used in type specifications.* The limitation on types can be circumvented by adding type arguments to the local (co)datatypes to abstract over the locally fixed types.

3. *The *primcorec* command does not allow user-specified names and attributes next to the entered formulas.* The less convenient syntax, using the *lemmas* command, is available as an alternative.

4. *The *primcorec* command does not allow corecursion under case–of for datatypes that are defined without discriminators and selectors.*

5. *There is no way to use an overloaded constant from a syntactic type class, such as 0, as a constructor.*

6. *There is no way to register the same type as both a datatype and a codatatype.* This affects types such as the extended natural numbers, for which both views would make sense (for a different set of constructors).
7. The names of variables are often suboptimal in the properties generated by the package.

8. The compatibility layer sometimes produces induction principles with a slightly different ordering of the premises than the old package.

Acknowledgment

Tobias Nipkow and Makarius Wenzel encouraged us to implement the new (co)datatype package. Andreas Lochbihler provided lots of comments on earlier versions of the package, especially on the coinductive part. Brian Huffman suggested major simplifications to the internal constructions. Ondřej Kunčar implemented the transfer and lifting plugins. Christian Sternagel and René Thiemann ported the derive command from the Archive of Formal Proofs to the new datatypes. Gerwin Klein and Lars Noschinski implemented the simps_of_case and case_of.simps commands. Florian Haftmann, Christian Urban, and Makarius Wenzel provided general advice on Isabelle and package writing. Stefan Milius and Lutz Schröder found an elegant proof that eliminated one of the BNF proof obligations. Mamoun Filali-Amine, Gerwin Klein, Andreas Lochbihler, Tobias Nipkow, and Christian Sternagel suggested many textual improvements to this tutorial.

References


REFERENCES


