Theorem Proving in Lean

Jeremy Avigad
Leonardo de Moura
Soonho Kong

Version f9a1d8e, updated at 2017-03-15 18:05:12 -0400

Contents

1

Introduction

1.1 Computers and Theorem Proving

Formal verification involves the use of logical and computational methods to establish claims that are expressed in precise mathematical terms. These can include ordinary mathematical theorems, as well as claims that pieces of hardware or software, network protocols, and mechanical and hybrid systems meet their specifications. In practice, there is not a sharp distinction between verifying a piece of mathematics and verifying the correctness of a system: formal verification requires describing hardware and software systems in mathematical terms, at which point establishing claims as to their correctness becomes a form of theorem proving. Conversely, the proof of a mathematical theorem may require a lengthy computation, in which case verifying the truth of the theorem requires verifying that the computation does what it is supposed to do.

The gold standard for supporting a mathematical claim is to provide a proof, and twentieth-century developments in logic show most if not all conventional proof methods can be reduced to a small set of axioms and rules in any of a number of foundational systems. With this reduction, there are two ways that a computer can help establish a claim: it can help find a proof in the first place, and it can help verify that a purported proof is correct.

Automated theorem proving focuses on the "finding" aspect. Resolution theorem provers, tableau theorem provers, fast satisfiability solvers, and so on provide means of establishing the validity of formulas in propositional and first-order logic. Other systems provide search procedures and decision procedures for specific languages and domains, such as linear or nonlinear expressions over the integers or the real numbers. Architectures like SMT ("satisfiability modulo theories") combine domain-general search methods with domain-specific procedures. Computer algebra systems and specialized mathematical software packages provide means of carrying out mathematical computations, establishing mathematical bounds, or finding mathematical objects. A calculation can be viewed as a proof as well, and these systems, too, help establish mathematical claims.

Automated reasoning systems strive for power and efficiency, often at the expense of guaranteed soundness. Such systems can have bugs, and it can be difficult to ensure that the results they deliver are correct. In contrast, *interactive theorem proving* focuses on the "verification" aspect of theorem proving, requiring that every claim is supported by a proof in a suitable axiomatic foundation. This sets a very high standard: every rule of inference and every step of a calculation has to be justified by appealing to prior definitions and theorems, all the way down to basic axioms and rules. In fact, most such systems provide fully elaborated "proof objects" that can be communicated to other systems and checked independently. Constructing such proofs typically requires much more input and interaction from users, but it allows us to obtain deeper and more complex proofs.

The *Lean Theorem Prover* aims to bridge the gap between interactive and automated theorem proving, by situating automated tools and methods in a framework that supports user interaction and the construction of fully specified axiomatic proofs. The goal is to support both mathematical reasoning and reasoning about complex systems, and to verify claims in both domains.

1.2 About Lean

The *Lean* project was launched by Leonardo de Moura at Microsoft Research Redmond in 2013. It is an ongoing, long-term effort, and much of the potential for automation will be realized only gradually over time. Lean is released under the Apache 2.0 license, a permissive open source license that permits others to use and extend the code and mathematical libraries freely.

There are currently two ways to use Lean. The first is to run it from the web: a Javascript version of Lean, a standard library of definitions and theorems, and an editor are actually downloaded to your browser and run there. This provides a quick and convenient way to begin experimenting with the system.

The second way to use Lean is to install and run it natively on your computer. The native version is much faster than the web version, and is more flexible in other ways, too. It comes with an Emacs mode that offers powerful support for writing and debugging proofs, and is much better suited for serious use. The source code, and instructions for building Lean, are available at `https://github.com/leanprover/lean/`.

At this moment, the development branch of Lean does not support homotopy type theory. If you are interested in using Lean to do homotopy type theory, you will need to use the snapshot provided at `https://github.com/leanprover/lean2/`. The tutorial for that snapshot is at `https://leanprover.github.io/tutorial/`.

1.3 About this Book

This book is designed to teach you to develop and verify proofs in Lean. Much of the background information you will need in order to do this is not specific to Lean at all. To start with, we will explain the logical system that Lean is based on, a version of *dependent type theory* that is powerful enough to prove almost any conventional mathematical theorem, and expressive enough to do it in a natural way. More specifically, Lean is based on a variant of a system known as the *Calculus of Inductive Constructions*[1, 3], or *CIC*. We will explain not only how to define mathematical objects and express mathematical assertions in dependent type theory, but also how to use it as a language for writing proofs.

Because fully detailed axiomatic proofs are so complicated, the challenge of theorem proving is to have the computer fill in as many of the details as possible. We will describe various methods to support this in dependent type theory. For example, we will discuss term rewriting, and Lean's automated methods for simplifying terms and expressions automatically. Similarly, we will discuss methods of *elaboration* and *type inference*, which can be used to support flexible forms of algebraic reasoning.

Finally, of course, we will discuss features that are specific to Lean, including the language with which you can communicate with the system, and the mechanisms Lean offers for managing complex theories and data.

If you are reading this book within Lean's online tutorial system, you will see a copy of the Lean editor at right, with an output buffer beneath it. At any point, you can type things into the editor, press the "play" button, and see Lean's response. Notice that you can resize the various windows if you would like.

Throughout the text you will find examples of Lean code like the one below:

```
theorem and_commutative (p q : Prop) : p ∧ q → q ∧ p :=
assume hpq : p ∧ q,
have hp : p, from and.left hpq,
have hq : q, from and.right hpq,
show q ∧ p, from and.intro hq hp
```

Once again, if you are reading the book online, you will see a button that reads "try it yourself." Pressing the button copies the example into the Lean editor with enough surrounding context to make the example compile correctly, and then runs Lean. We recommend running the examples and experimenting with the code on your own as you work through the chapters that follow.

1.4 Acknowledgments

This tutorial is an open access project maintained on Github. Many people have contributed to the effort, providing corrections, suggestions, examples, and text. We are grateful to Ulrik Buchholz, Nathan Carter, Amine Chaieb, Floris van Doorn, Anthony

Hart, Sean Leather, Christopher John Mazey, Assia Mahboubi, Sebastian Ullrich, Daniel Velleman, and Théo Zimmerman for their contributions, and we apologize to those whose names we have inadvertently omitted.

2

Dependent Type Theory

Dependent type theory is a powerful and expressive language, allowing us to express complex mathematical assertions, write complex hardware and software specifications, and reason about both of these in a natural and uniform way. Lean is based on a version of dependent type theory known as the *Calculus of Inductive Constructions*, with a countable hierarchy of non-cumulative universes and inductive types. By the end of this chapter, you will understand much of what this means.

2.1 Simple Type Theory

As a foundation for mathematics, set theory has a simple ontology that is rather appealing. Everything is a set, including numbers, functions, triangles, stochastic processes, and Riemannian manifolds. It is a remarkable fact that one can construct a rich mathematical universe from a small number of axioms that describe a few basic set-theoretic constructions.

But for many purposes, including formal theorem proving, it is better to have an infrastructure that helps us manage and keep track of the various kinds of mathematical objects we are working with. "Type theory" gets its name from the fact that every expression has an associated *type*. For example, in a given context, `x + 0` may denote a natural number and `f` may denote a function on the natural numbers.

Here are some examples of how we can declare objects in Lean and check their types.

```
/- declare some constants -/

constant m : nat        -- m is a natural number
constant n : nat
```

```
constants b1 b2 : bool  -- declare two constants at once

/- check their types -/

check m            -- output: nat
check n
check n + 0        -- nat
check m * (n + 0)  -- nat
check b1           -- bool
check b1 && b2     -- "&&" is boolean and
check b1 || b2     -- boolean or
check tt           -- boolean "true"

-- Try some examples of your own.
```

The /- and -/ annotations indicate that the next line is a comment block that is ignored by Lean. Similarly, two dashes indicate that the rest of the line contains a comment that is also ignored. Comment blocks can be nested, making it possible to "comment out" chunks of code, just as in many programming languages.

The `constant` and `constants` commands introduce new constant symbols into the working environment, and the `check` command asks Lean to report their types. You should test this, and try typing some examples of your own. Declaring new objects in this way is a good way to experiment with the system, but it is ultimately undesirable: Lean is a foundational system, which is to say, it provides us with powerful mechanisms to *define* all the mathematical objects we need, rather than simply postulating them to the system. We will explore these mechanisms in the chapters to come.

What makes simple type theory powerful is that one can build new types out of others. For example, if α and β are types, $\alpha \to \beta$ denotes the type of functions from α to β, and $\alpha \times \beta$ denotes the cartesian product, that is, the type of ordered pairs consisting of an element of α paired with an element of β.

```
constants m n : nat

constant f : nat → nat           -- type the arrow as "\to" or "\r"
constant f' : nat -> nat         -- alternative ASCII notation
constant f'' : ℕ → ℕ              -- \nat is alternative notation for nat
constant p : nat × nat           -- type the product as "\times"
constant q : prod nat nat        -- alternative notation
constant g : nat → nat → nat
constant g' : nat → (nat → nat)  -- has the same type as g!
constant h : nat × nat → nat

constant F : (nat → nat) → nat   -- a "functional"

check f               -- ℕ → ℕ
check f n             -- ℕ
check g m n           -- ℕ
check g m             -- ℕ → ℕ
check (m, n)          -- ℕ × ℕ
check p.1             -- ℕ
```

```
check p.2           -- ℕ
check (m, n).1      -- ℕ
check (p.1, n)      -- ℕ × ℕ
check F f           -- ℕ

-- Try it on your own: write down some types, declare some constants,
-- and check some expressions.
```

Let us dispense with some basic syntax. You can enter the unicode arrow → by typing `\to` or “`\r`. You can also use the ASCII alternative `->`, so that the expression `nat -> nat` and `nat → nat` mean the same thing. Both expressions denote the type of functions that take a natural number as input and return a natural number as output. The symbol ℕ is alternative unicode notation for `nat`; you can enter it by typing `\nat`. The unicode symbol × for the cartesian product is entered `\times`. We will generally use lower-case greek letters like α, β, and γ to range over types. You can enter these particular ones with `\a`, `\b`, and `\g`.

There are a few more things to notice here. First, the application of a function `f` to a value `x` is denoted `f x`. Second, when writing type expressions, arrows associate to the *right*; for example, the type of `g` is `nat → (nat → nat)`. Thus we can view `g` as a function that takes natural numbers and returns another function that takes a natural number and returns a natural number. In type theory, this is generally more convenient than writing `g` as a function that takes a pair of natural numbers as input, and returns a natural number as output. For example, it allows us to “partially apply” the function `g`. The example above shows that `g m` has type `nat → nat`, that is, the function that “waits” for a second argument, `n`, and then returns `g m n`. Taking a function `h` of type `nat × nat → nat` and “redefining” it to look like `g` is a process known as *currying*, something we will come back to below.

By now you may also have guessed that, in Lean, `(m, n)` denotes the ordered pair of `m` and `n`, and if `p` is a pair, `fst p` and `snd p` denote the two projections.

2.2 Types as Objects

One way in which Lean’s dependent type theory extends simple type theory is that types themselves – entities like `nat` and `bool` – are first-class citizens, which is to say that they themselves are objects of study. For that to be the case, each of them also has to have a type.

```
check nat               -- Type
check bool              -- Type
check nat → bool         -- Type
check nat × bool         -- Type
check nat → nat          -- ...
check nat × nat → nat
check nat → nat → nat
```

```
check nat → (nat → nat)
check nat → nat → bool
check (nat → nat) → nat
```

We see that each one of the expressions above is an object of type `Type`. We can also declare new constants and constructors for types:

```
constants α β : Type
constant F : Type → Type
constant G : Type → Type → Type

check α          -- Type
check F α        -- Type
check F nat      -- Type
check G α        -- Type → Type
check G α β      -- Type
check G α nat    -- Type
```

Indeed, we have already seen an example of a function of type `Type → Type → Type`, namely, the Cartesian product.

```
constants α β : Type

check prod α β        -- Type
check prod nat nat    -- Type
```

Here is another example: given any type α, the type `list α` denotes the type of lists of elements of type α.

```
constant α : Type

check list α      -- Type
check list nat    -- Type
```

For those more comfortable with set-theoretic foundations, it may be helpful to think of a type as nothing more than a set, in which case, the elements of the type are just the elements of the set. Given that every expression in Lean has a type, it is natural to ask: what type does `Type` itself have?

```
check Type      -- Type 1
```

We have actually come up against one of the most subtle aspects of Lean's typing system. Lean's underlying foundation has an infinite hierarchy of types:

```
check Type      -- Type 1
check Type 1    -- Type 2
```

```
check Type 2   -- Type 3
check Type 3   -- Type 4
check Type 4   -- Type 5
```

Think of `Type 0` as a universe of "small" or "ordinary" types. `Type 1` is then a larger universe of types, which contains `Type 0` as an element, and `Type 2` is an even larger universe of types, which contains `Type 1` as an element. The list is indefinite, so that there is a `Type n` for every natural number `n`. `Type` is an abbreviation for `Type 0`:

```
check Type
check Type 0
```

There is also `Prop`, this type has special properties, and will be discussed in the next chapter.

```
check Prop -- Type 0
```

We want some operations, however, to be *polymorphic* over type universes. For example, `list` α should make sense for any type α, no matter which type universe α lives in. This explains the type annotation of the function `list`:

```
check list    -- Type u_1 → Type u_1
```

Here `u_1` is a variable ranging over type levels. The output of the `check` command means that whenever α has type `Type n`, `list` α also has type `Type n`. The function `prod` is similarly polymorphic:

```
check prod    -- Type u_1 → Type u_2 → Type (max u_1 u_2)
```

To define polymorphic constants and variables, Lean allows us to declare universe variables explicitly:

```
universe variable u
constant α : Type u
check α
```

Throughout this book, you will see us do this in examples when we want type constructions to have as much generality as possible. We will see that the ability to treat type constructors as instances of ordinary mathematical functions is a powerful feature of dependent type theory.

2.3 Function Abstraction and Evaluation

We have seen that if we have `m n : nat`, then we have `(m, n) : nat × nat`. This gives us a way of creating pairs of natural numbers. Conversely, if we have `p : nat × nat`, then we have `fst p : nat` and `snd p : nat`. This gives us a way of "using" a pair, by extracting its two components.

We already know how to "use" a function `f : α → β`, namely, we can apply it to an element `a : α` to obtain `f a : β`. But how do we create a function from another expression?

The companion to application is a process known as "abstraction," or "lambda abstraction." Suppose that by temporarily postulating a variable `x : α` we can construct an expression `t : β`. Then the expression `fun x : α, t`, or, equivalently, `λ x : α, t`, is an object of type `α → β`. Think of this as the function from α to β which maps any value `x` to the value `t`, which depends on `x`. For example, in mathematics it is common to say "let `f` be the function which maps any natural number `x` to `x + 5`." The expression `λ x : nat, x + 5` is just a symbolic representation of the right-hand side of this assignment.

```
check fun x : nat, x + 5
check λ x : nat, x + 5
```

Here are some more abstract examples:

```
constants α β  : Type
constants a1 a2 : α
constants b1 b2 : β

constant f : α → α
constant g : α → β
constant h : α → β → α
constant p : α → α → bool

check fun x : α, f x                      -- α → α
check λ x : α, f x                        -- α → α
check λ x : α, f (f x)                    -- α → α
check λ x : α, h x b1                     -- α → α
check λ y : β, h a1 y                     -- β → α
check λ x : α, p (f (f x)) (h (f a1) b2)  -- α → bool
check λ x : α, λ y : β, h (f x) y          -- α → β → α
check λ (x : α) (y : β), h (f x) y         -- α → β → α
check λ x y, h (f x) y                    -- α → β → α
```

Lean interprets the final three examples as the same expression; in the last expression, Lean infers the type of `x` and `y` from the types of `f` and `h`.

Be sure to try writing some expressions of your own. Some mathematically common examples of operations of functions can be described in terms of lambda abstraction:

```
constants α β γ : Type
constant f : α → β
constant g : β → γ
constant b : β

check λ x : α, x        -- α → α
check λ x : α, b        -- α → β
check λ x : α, g (f x)  -- α → γ
check λ x, g (f x)

-- we can abstract any of the constants in the previous definitions

check λ b : β, λ x : α, x      -- β → α → α
check λ (b : β) (x : α), x     -- equivalent to the previous line
check λ (g : β → γ) (f : α → β) (x : α), g (f x)
                                -- (β → γ) → (α → β) → α → γ
-- we can even abstract over the type

check λ (α β : Type) (b : β) (x : α), x
check λ (α β γ : Type) (g : β → γ) (f : α → β) (x : α), g (f x)
```

Think about what these expressions mean. The expression `λ x : α, x` denotes the identity function on α, the expression `λ x : α, b` denotes the constant function that always returns `b`, and `λ x : α, g (f x)`, denotes the composition of `f` and `g`. We can, in general, leave off the type annotations on the variable and let Lean infer it for us. So, for example, we can write `λ x, g (f x)` instead of `λ x : α, g (f x)`.

We can abstract over any of the constants in the previous definitions:

```
check λ b : β, λ x : α, x      -- β → α → α
check λ (b : β) (x : α), x     -- β → α → α
check λ (g : β → γ) (f : α → β) (x : α), g (f x)
                                -- (β → γ) → (α → β) → α → γ
```

Lean lets us combine lambdas, so the second example is equivalent to the first. We can even abstract over the type:

```
check λ (α β : Type) (b : β) (x : α), x
check λ (α β γ : Type) (g : β → γ) (f : α → β) (x : α), g (f x)
```

The last expression, for example, denotes the function that takes three types, α, β, and γ, and two functions, `g : β → γ` and `f : α → β`, and returns the composition of `g` and `f`. (Making sense of the type of this function requires an understanding of dependent products, which we will explain below.) Within a lambda expression `λ x : α, t`, the variable `x` is a "bound variable": it is really a placeholder, whose "scope" does not extend beyond `t`. For example, the variable `b` in the expression `λ (b : β) (x : α), x` has nothing to do with the constant `b` declared earlier. In fact, the expression denotes the same function as `λ (u : β) (z : α), z`. Formally, the expressions that are the same up to a renaming of bound variables are called *alpha equivalent*, and are considered “the same.” Lean recognizes this equivalence.

Notice that applying a term `t : α → β` to a term `s : α` yields an expression `t s : β`. Returning to the previous example and renaming bound variables for clarity, notice the types of the following expressions:

```
constants α β γ : Type
constant f : α → β
constant g : β → γ
constant h : α → α
constants (a : α) (b : β)

check (λ x : α, x) a                -- α
check (λ x : α, b) a                -- β
check (λ x : α, b) (h a)            -- β
check (λ x : α, g (f x)) (h (h a))  -- γ

check (λ (v : β → γ) (u : α → β) x, v (u x)) g f a   -- γ

check (λ (Q R S : Type) (v : R → S) (u : Q → R) (x : Q),
       v (u x)) α β γ g f a         -- γ
```

As expected, the expression `(λ x : α, x) a` has type α. In fact, more should be true: applying the expression `(λ x : α, x)` to `a` should “return” the value `a`. And, indeed, it does:

```
constants α β γ : Type
constant f : α → β
constant g : β → γ
constant h : α → α
constants (a : α) (b : β)

eval (λ x : α, x) a                -- a
eval (λ x : α, b) a                -- b
eval (λ x : α, b) (h a)            -- b
eval (λ x : α, g (f x)) a          -- g (f a)

eval (λ (v : β → γ) (u : α → β) x, v (u x)) g f a   -- g (f a)

eval (λ (Q R S : Type) (v : R → S) (u : Q → R) (x : Q),
      v (u x)) α β γ g f a         -- g (f a)
```

The command `eval` tells Lean to *evaluate* an expression. The process of simplifying an expression `(λ x, t)s` to `t[s/x]` – that is, `t` with `s` substituted for the variable `x` – is known as *beta reduction*, and two terms that beta reduce to a common term are called *beta equivalent*. But the `eval` command carries out other forms of reduction as well:

```
constants m n : nat
constant b : bool
```

```
print "reducing pairs"
eval (m, n).1        -- m
eval (m, n).2        -- n

print "reducing boolean expressions"
eval tt && ff        -- ff
eval b && ff         -- ff

print "reducing arithmetic expressions"
eval n + 0           -- n
eval n + 2           -- succ (succ n)
eval 2 + 3           -- 5
```

In a later chapter, we will explain how these terms are evaluated. For now, we only wish to emphasize that this is an important feature of dependent type theory: every term has a computational behavior, and supports a notion of reduction, or *normalization*. In principle, two terms that reduce to the same value are called *definitionally equal*. They are considered "the same" by the underlying logical framework, and Lean does its best to recognize and support these identifications.

2.4 Introducing Definitions

As we have noted above, declaring constants in the Lean environment is a good way to postulate new objects to experiment with, but most of the time what we really want to do is *define* objects in Lean and prove things about them. The `definition` command provides one important way of defining new objects.

```
definition foo : (ℕ → ℕ) → ℕ := λ f, f 0

check foo    -- ℕ
print foo    -- λ (f : ℕ → ℕ), f 0
```

We can omit the type when Lean has enough information to infer it:

```
definition foo' := λ f : ℕ → ℕ, f 0
```

The general form of a definition is `definition foo : T := bar`. Lean can usually infer the type `T`, but it is often a good idea to write it explicitly. This clarifies your intention, and Lean will flag an error if the right-hand side of the definition does not have the right type.

Because function definitions are so common, Lean provides the shorthand `def` for `definition`, and an alternative notation, which puts the abstracted variables before the colon and omits the lambda:

```
def double (x : ℕ) : ℕ := x + x
print double
check double 3
eval double 3    -- 6

def square (x : ℕ) := x * x
print square
check square 3
eval square 3    -- 9

def do_twice (f : ℕ → ℕ) (x : ℕ) : ℕ := f (f x)

eval do_twice double 2    -- 8
```

These definitions are equivalent to the following:

```
def double : ℕ → ℕ := λ x, x + x
def square : ℕ → ℕ := λ x, x * x
def do_twice : (ℕ → ℕ) → ℕ → ℕ := λ f x, f (f x)
```

We can even use this approach to specify arguments that are types:

```
def compose (α β γ : Type) (g : β → γ) (f : α → β) (x : α) : γ :=
g (f x)
```

As an exercise, we encourage you to use `do_twice` and `double` to define functions that quadruple their input, and multiply the input by 8. As a further exercise, we encourage you to try defining a function `Do_Twice : ((ℕ → ℕ) → (ℕ → ℕ)) → (ℕ → ℕ) → (ℕ → ℕ)` which iterates *its* argument twice, so that `Do_Twice do_twice` a function which iterates *its* input four times, and evaluate `Do_Twice do_twice double 2`.

Above, we discussed the process of "currying" a function, that is, taking a function `f (a, b)` that takes an ordered pair as an argument, and recasting it as a function `f' a b` that takes two arguments successively. As another exercise, we encourage you to complete the following definitions, which "curry" and "uncurry" a function.

```
def curry (α β γ : Type) (f : α × β → γ) : α → β → γ := sorry

def uncurry (α β γ : Type) (f : α → β → γ) : α × β → γ := sorry
```

2.5 Local Definitions

Lean also allows you to introduce "local" definitions using the `let` construct. The expression `let a := t1 in t2` is definitionally equal to the result of replacing every occurrence of `a` in `t2` by `t1`.

```
check let y := 2 + 2 in y * y   -- ℕ
eval  let y := 2 + 2 in y * y   -- 16

def t (x : ℕ) : ℕ :=
let y := x + x in y * y

eval t 2   -- 16
```

Here, `t` is definitionally equal to the term `(x + x) * (x + x)`. You can combine multiple assignments in a single `let` statement:

```
check let y := 2 + 2, z := y + y in z * z   -- 16
eval  let y := 2 + 2, z := y + y in z * z   -- 64
```

Notice that the meaning of the expression `let a := t1 in t2` is very similar to the meaning of `(λ a, t2) t1`, but the two are not the same. In the first expression, you should think of every instance of `a` in `t2` as a syntactic abbreviation for `t1`. In the second expression, `a` is a variable, and the expression `λ a, t2` has to make sense independently of the value of `a`. The `let` construct is a stronger means of abbreviation, and there are expressions of the form `let a := t1 in t2` that cannot be expressed as `(λ a, t2) t1`. As an exercise, try to understand why the definition of `foo` below type checks, but the definition of `bar` does not.

```
def foo := let a := nat  in λ x : a, x + 2

/-
def bar := (λ a, λ x : a, x + 2) nat
-/
```

2.6 Variables and Sections

This is a good place to introduce some organizational features of Lean that are not a part of the axiomatic framework *per se*, but make it possible to work in the framework more efficiently.

We have seen that the `constant` command allows us to declare new objects, which then become part of the global context. Declaring new objects in this way is somewhat crass. Lean enables us to *define* all of the mathematical objects we need, and *declaring* new objects willy-nilly is therefore somewhat lazy. In the words of Bertrand Russell, it has all the advantages of theft over honest toil. We will see in the next chapter that it is also somewhat dangerous: declaring a new constant is tantamount to declaring an axiomatic extension of our foundational system, and may result in inconsistency.

So far, in this tutorial, we have used the `constant` command to create "arbitrary" objects to work with in our examples. For example, we have declared types α, β, and γ to populate our context. This can be avoided, using implicit or explicit lambda abstraction in our definitions to declare such objects "locally":

```
def compose (α β γ : Type) (g : β → γ) (f : α → β) (x : α) :
  γ := g (f x)

def do_twice (α : Type) (h : α → α) (x : α) : α := h (h x)

def do_thrice (α : Type) (h : α → α) (x : α) : α := h (h (h x))
```

Repeating declarations in this way can be tedious, however. Lean provides us with the `variable` and `variables` commands to make such declarations look global:

```
variables (α β γ : Type)

def compose (g : β → γ) (f : α → β) (x : α) : γ := g (f x)
def do_twice (h : α → α) (x : α) : α := h (h x)
def do_thrice (h : α → α) (x : α) : α := h (h (h x))
```

We can declare variables of any type, not just `Type` itself:

```
variables (α β γ : Type)
variables (g : β → γ) (f : α → β) (h : α → α)
variable x : α

def compose := g (f x)
def do_twice := h (h x)
def do_thrice := h (h (h x))

print compose
print do_twice
print do_thrice
```

Printing them out shows that all three groups of definitions have exactly the same effect.

The `variable` and `variables` commands look like the `constant` and `constants` commands we have used above, but there is an important difference: rather than creating permanent entities, the declarations simply tell Lean to insert the variables as bound variables in definitions that refer to them. Lean is smart enough to figure out which variables are used explicitly or implicitly in a definition. We can therefore proceed as though α, β, γ, `g`, `f`, `h`, and `x` are fixed objects when we write our definitions, and let Lean abstract the definitions for us automatically.

When declared in this way, a variable stays in scope until the end of the file we are working on, and we cannot declare another variable with the same name. Sometimes, however, it is useful to limit the scope of a variable. For that purpose, Lean provides the notion of a `section`:

```
section useful
  variables (α β γ : Type)
  variables (g : β → γ) (f : α → β) (h : α → α)
  variable x : α

  def compose := g (f x)
  def do_twice := h (h x)
  def do_thrice := h (h (h x))
end useful
```

When the section is closed, the variables go out of scope, and become nothing more than a distant memory.

You do not have to indent the lines within a section, since Lean treats any blocks of returns, spaces, and tabs equivalently as whitespace. Nor do you have to name a section, which is to say, you can use an anonymous `section` / `end` pair. If you do name a section, however, you have to close it using the same name. Sections can also be nested, which allows you to declare new variables incrementally.

We will see in Chapter 6 that, as a scoping mechanism, sections govern more than just variables; other commands have effects that are only operant in the current section. Similarly, if we use the `open` command inside a section, it only remains in effect until that section is closed.

2.7 Namespaces

Lean provides us with the ability to group definitions, notation, and other information into nested, hierarchical *namespaces*:

```
namespace foo
  def a : ℕ := 5
  def f (x : ℕ) : ℕ := x + 7

  def fa : ℕ := f a
  def ffa : ℕ := f (f a)

  print "inside foo"

  check a
  check f
  check fa
  check ffa
  check foo.fa
end foo

print "outside the namespace"

-- check a  -- error
-- check f  -- error
check foo.a
check foo.f
```

```
check foo.fa
check foo.ffa

open foo

print "opened foo"

check a
check f
check fa
check foo.fa
```

When we declare that we are working in the namespace `foo`, every identifier we declare has a full name with prefix "`foo.`" Within the namespace, we can refer to identifiers by their shorter names, but once we end the namespace, we have to use the longer names.

The `open` command brings the shorter names into the current context. Often, when we import a theory file, we will want to open one or more of the namespaces it contains, to have access to the short identifiers, notations, and so on. But sometimes we will want to leave this information hidden, for example, when they conflict with identifiers and notations in another namespace we want to use. Thus namespaces give us a way to manage our working environment.

For example, Lean groups definitions and theorems involving lists into a namespace `list`.

```
check list.nil
check list.cons
check list.append
```

We will discuss their types, below. The command `open list` allows us to use the shorter names:

```
open list

check nil
check cons
check append
```

Like sections, namespaces can be nested:

```
namespace foo
  def a : ℕ := 5
  def f (x : ℕ) : ℕ := x + 7

  def fa : ℕ := f a

  namespace bar
    def ffa : ℕ := f (f a)
```

```
    check fa
    check ffa
  end bar

  check fa
  check bar.ffa
end foo

check foo.fa
check foo.bar.ffa

open foo

check fa
check bar.ffa
```

Namespaces that have been closed can later be reopened, even in another file:

```
namespace foo
  def a : ℕ := 5
  def f (x : ℕ) : ℕ := x + 7

  def fa : ℕ := f a
end foo

check foo.a
check foo.f

namespace foo
  def ffa : ℕ := f (f a)
end foo
```

Like sections, nested namespaces have to be closed in the order they are opened. Also, a namespace cannot be opened within a section; namespaces have to live on the outer levels.

Namespaces and sections serve different purposes: namespaces organize data and sections declare variables for insertion in theorems. In many respects, however, a `namespace ... end` block behaves the same as a `section ... end` block. In particular, if you use the `variable` command within a namespace, its scope is limited to the namespace. Similarly, if you use an `open` command within a namespace, its effects disappear when the namespace is closed.

2.8 Dependent Types

You now have rudimentary ways of defining functions and objects in Lean, and we will gradually introduce you to many more. Our ultimate goal in Lean is to *prove* things about the objects we define, and the next chapter will introduce you to Lean's mechanisms for stating theorems and constructing proofs. Meanwhile, let us remain on the topic of defining objects in dependent type theory for just a moment longer, in order to explain what makes dependent type theory *dependent*, and why that is useful.

The short explanation is that what makes dependent type theory dependent is that types can depend on parameters. You have already seen a nice example of this: the type `list` α depends on the argument α, and this dependence is what distinguishes `list` $\mathbb{N}$ and `list bool`. For another example, consider the type `vec` α `n`, the type of vectors of elements of α of length `n`. This type depends on *two* parameters: the type α `: Type` of the elements in the vector and the length `n :` $\mathbb{N}$.

Suppose we wish to write a function `cons` which inserts a new element at the head of a list. What type should `cons` have? Such a function is *polymorphic*: we expect the `cons` function for $\mathbb{N}$, `bool`, or an arbitrary type α to behave the same way. So it makes sense to take the type to be the first argument to `cons`, so that for any type, α, `cons` α is the insertion function for lists of type α. In other words, for every α, `cons` α is the function that takes an element `a :` α and a list `l : list` α, and returns a new list, so we have `cons` α `a l : list` α.

It is clear that `cons` α should have type α → `list` α → `list` α. But what type should `cons` have? A first guess might be `Type` → α → `list` α → `list` α, but, on reflection, this does not make sense: the α in this expression does not refer to anything, whereas it should refer to the argument of type `Type`. In other words, *assuming* α `: Type` is the first argument to the function, the type of the next two elements are α and `list` α. These types vary depending on the first argument, α.

This is an instance of a *Pi type* in dependent type theory. Given α `: Type` and β `:` α → `Type`, think of β as a family of types over α, that is, a type β `a` for each `a :` α. In that case, the type Π `x :` α`,` β `x` denotes the type of functions `f` with the property that, for each `a :` α, `f a` is an element of β `a`. In other words, the type of the value returned by `f` depends on its input.

Notice that Π `x :` α`,` β makes sense for any expression β `: Type`. When the value of β depends on `x` (as does, for example, the expression β `x` in the previous paragraph), Π `x :` α`,` β denotes a dependent function type. When β doesn't depend on `x`, Π `x :` α`,` β is no different from the type α → β. Indeed, in dependent type theory (and in Lean), the Pi construction is fundamental, and α → β is nothing more than notation for Π `x :` α`,` β when β does not depend on α.

Returning to the example of lists, we can model some basic list operations as follows. We use `namespace hide` to avoid a naming conflict with the `list` type defined in the standard library.

```
namespace hide

universe variable u

constant list : Type u → Type u

constant cons : Π α : Type u, α → list α → list α
constant nil : Π α : Type u, list α
constant head : Π α : Type u, list α → α
```

```
constant tail : Π α : Type u, list α → list α
constant append : Π α : Type u, list α → list α → list α

end hide
```

You can enter the symbol Π by typing `\Pi`. Here, `nil` is intended to denote the empty list, `head` and `tail` return the first element of a list and the remainder, respectively. The constant `append` is intended to denote the function that concatenates two lists.

We emphasize that these constant declarations are only for the purposes of illustration. The `list` type and all these operations are, in fact, *defined* in Lean's standard library, and are proved to have the expected properties. In fact, as the next example shows, the types indicated above are essentially the types of the objects that are defined in the library. (We will explain the `@` symbol and the difference between the round and curly brackets momentarily.)

```
open list

check list      -- Type u_1 → Type u_1

check @cons     -- Π {α : Type u_1}, α → list α → list α
check @nil      -- Π {α : Type u_1}, list α
check @head     -- Π {α : Type u_1} [_inst_1 : inhabited α], list α → α
check @tail     -- Π {α : Type u_1}, list α → list α
check @append   -- Π {α : Type u_1}, list α → list α → list α
```

There is a subtlety in the definition of `head`: the type α is required to have at least one element, and when passed the empty list, the function must determine a default element of the relevant type. We will explain how this is done in a later chapter.

Vector operations are handled similarly:

```
universe variable u
constant vec : Type u → ℕ → Type u

namespace vec
  constant empty : Π α : Type u, vec α 0
  constant cons :
    Π (α : Type u) (n : ℕ), α → vec α n → vec α (n + 1)
  constant append :
    Π (α : Type u) (n m : ℕ),  vec α m → vec α n → vec α (n + m)
end vec
```

In the coming chapters, you will come across many instances of dependent types. Here we will mention just one more important and illustrative example, the *Sigma types*, `Σ x : α, β x`, sometimes also known as *dependent pairs*. These are, in a sense, companions to the Pi types. The type `Σ x : α, β x` denotes the type of pairs `sigma.mk a b` where `a : α` and `b : β a`.

Just as Pi types `Π x : α, β x` generalize the notion of a function type $\alpha \to \beta$ by allowing β to depend on α, Sigma types `Σ x : α, β x` generalize the cartesian product $\alpha \times \beta$ in the same way: in the expression `sigma.mk a b`, the type of the second element of the pair, `b : β a`, depends on the first element of the pair, `a : α`.

```
variable α : Type
variable β : α → Type
variable a : α
variable b : β a

check sigma.mk a b    -- Σ (a : α), β a
check (sigma.mk a b).1   -- α
check (sigma.mk a b).2   -- β (sigma.fst (sigma.mk a b))

eval  (sigma.mk a b).1   -- a
eval  (sigma.mk a b).2   -- b
```

Notice that when `p` is a dependent pair the expressions `(sigma.mk a b).1` and `(sigma.mk a b).2` are short for `sigma.fst (sigma.mk a b)` and `sigma.snd (sigma.mk a b)`, respectively, and that these reduce to `a` and `b`, respectively.

2.9 Implicit Arguments

Suppose we have an implementation of lists as described above.

```
namespace hide
universe variable u
constant list : Type u → Type u

namespace list
  constant cons : Π α : Type u, α → list α → list α
  constant nil : Π α : Type u, list α
  constant append : Π α : Type u, list α → list α → list α
end list
end hide
```

Then, given a type α, some elements of α, and some lists of elements of α, we can construct new lists using the constructors.

```
open hide.list

variable  α : Type
variable  a : α
variables l1 l2 : list α

check cons α a (nil α)
check append α (cons α a (nil α)) l1
check append α (append α (cons α a (nil α)) l1) l2
```

Because the constructors are polymorphic over types, we have to insert the type α as an argument repeatedly. But this information is redundant: one can infer the argument α in `cons α a (nil α)` from the fact that the second argument, `a`, has type α. One can similarly infer the argument in `nil α`, not from anything else in that expression, but from the fact that it is sent as an argument to the function `cons`, which expects an element of type `list α` in that position.

This is a central feature of dependent type theory: terms carry a lot of information, and often some of that information can be inferred from the context. In Lean, one uses an underscore, `_`, to specify that the system should fill in the information automatically. This is known as an "implicit argument."

```
check cons _ a (nil _)
check append _ (cons _ a (nil _)) l1
check append _ (append _ (cons _ a (nil _)) l1) l2
```

It is still tedious, however, to type all these underscores. When a function takes an argument that can generally be inferred from context, Lean allows us to specify that this argument should, by default, be left implicit. This is done by putting the arguments in curly braces, as follows:

```
namespace list
  constant cons : Π {α : Type u}, α → list α → list α
  constant nil : Π {α : Type u}, list α
  constant append : Π {α : Type u}, list α → list α → list α
end list

open hide.list

variable  α : Type
variable  a : α
variables l1 l2 : list α

check cons a nil
check append (cons a nil) l1
check append (append (cons a nil) l1) l2
```

All that has changed are the braces around `α : Type u` in the declaration of the variables. We can also use this device in function definitions:

```
universe variable u
def ident {α : Type u} (x : α) := x

variables α β : Type u
variables (a : α) (b : β)

check ident      -- ?M_1 → ?M_1
check ident a    -- α
check ident b    -- β
```

This makes the first argument to `ident` implicit. Notationally, this hides the specification of the type, making it look as though `ident` simply takes an argument of any type. In fact, the function `id` is defined in the standard library in exactly this way. We have chosen a nontraditional name here only to avoid a clash of names.

Variables can also be declared implicit when they are declared with the `variables` command:

```
universe variable u

section
  variable {α : Type u}
  variable x : α
  def ident := x
end

variables α β : Type u
variables (a : α) (b : β)

check ident
check ident a
check ident b
```

This definition of `ident` has the same effect as the one above.

Lean has very complex mechanisms for instantiating implicit arguments, and we will see that they can be used to infer function types, predicates, and even proofs. The process of instantiating these "holes," or "placeholders," in a term is often known as *elaboration*. The presence of implicit arguments means that at times there may be insufficient information to fix the meaning of an expression precisely. An expression like `id` or `list.nil` is said to be *polymorphic*, because it can take on different meanings in different contexts. One can always specify the type `T` of an expression `e` by writing `(e : T)`. This instructs Lean's elaborator to use the value `T` as the type of `e` when trying to resolve implicit arguments. The second pair of examples below use this mechanism to specify the desired types of the expressions `id` and `list.nil`:

```
check list.nil              -- list ?M1
check id                    -- ?M1 → ?M1

check (list.nil : list ℕ)  -- list ℕ
check (id : ℕ → ℕ)          -- ℕ → ℕ
```

Numerals are overloaded in Lean, but when the type of a numeral cannot be inferred, Lean assumes, by default, that it is a natural number. So the expressions in the first two `check` commands are elaborated in the same way, whereas the third `check` command interprets `2` as a raw numeral.

```
check 2            -- ℕ
check (2 : ℕ)      -- ℕ
check (2 : num)    -- num
```

Sometimes, however, we may find ourselves in a situation where we have declared an argument to a function to be implicit, but now want to provide the argument explicitly. If `foo` is such a function, the notation `@foo` denotes the same function with all the arguments made explicit.

```
check @id        -- Π {α : Type u_1}, α → α
check @id α      -- α → α
check @id β      -- β → β
check @id α a    -- α
check @id β b    -- β
```

Notice that now the first `check` command gives the type of the identifier, `id`, without inserting any placeholders. Moreover, the output indicates that the first argument is implicit.

3

Propositions and Proofs

By now, you have seen how to define some elementary notions in dependent type theory.

In this chapter, we will explain how mathematical propositions and proofs are expressed in the language of dependent type theory, so that you can start proving assertions about the objects and notations that have been defined.

3.1 Propositions as Types

One strategy for proving assertions about objects defined in the language of dependent type theory is to layer an assertion language and a proof language on top of the definition language. But there is no reason to multiply languages in this way: dependent type theory is flexible and expressive, and there is no reason we cannot represent assertions and proofs in the same general framework.

For example, we could introduce a new type, `Prop`, to represent propositions, and constructors to build new propositions from others.

```
constant and : Prop → Prop → Prop
constant or : Prop → Prop → Prop
constant not : Prop → Prop
constant implies : Prop → Prop → Prop

variables p q r : Prop
check and p q                      -- Prop
check or (and p q) r               -- Prop
check implies (and p q) (and q p)  -- Prop
```

We could then introduce, for each element `p : Prop`, another type `Proof p`, for the type of proofs of `p`. An "axiom" would be constant of such a type.

```
constant Proof : Prop → Type

constant and_comm : Π p q : Prop, Proof (implies (and p q) (and q p))

variables p q : Prop
check and_comm p q      -- Proof (implies (and p q) (and q p))
```

In addition to axioms, however, we would also need rules to build new proofs from old ones. For example, in many proof systems for propositional logic, we have the rule of modus ponens:

> From a proof of `implies p q` and a proof of `p`, we obtain a proof of `q`.

We could represent this as follows:

```
constant modus_ponens (p q : Prop) : Proof (implies p q) →  Proof p → Proof q
```

Systems of natural deduction for propositional logic also typically rely on the following rule:

> Suppose that, assuming `p` as a hypothesis, we have a proof of `q`. Then we can "cancel" the hypothesis and obtain a proof of `implies p q`.

We could render this as follows:

```
constant implies_intro (p q : Prop) : (Proof p → Proof q) → Proof (implies p q).
```

This approach would provide us with a reasonable way of building assertions and proofs. Determining that an expression `t` is a correct proof of assertion `p` would then simply be a matter of checking that `t` has type `Proof p`.

Some simplifications are possible, however. To start with, we can avoid writing the term `Proof` repeatedly by conflating `Proof p` with `p` itself. In other words, whenever we have `p : Prop`, we can interpret `p` as a type, namely, the type of its proofs. We can then read `t : p` as the assertion that `t` is a proof of `p`.

Moreover, once we make this identification, the rules for implication show that we can pass back and forth between `implies p q` and `p → q`. In other words, implication between propositions `p` and `q` corresponds to having a function that takes any element of `p` to an element of `q`. As a result, the introduction of the connective `implies` is entirely redundant: we can use the usual function space constructor `p → q` from dependent type theory as our notion of implication.

This is the approach followed in the Calculus of Inductive Constructions, and hence in Lean as well. The fact that the rules for implication in a proof system for natural deduction correspond exactly to the rules governing abstraction and application for functions is an instance of the *Curry-Howard isomorphism*, sometimes known as the *propositions-as-types* paradigm. In fact, the type `Prop` is syntactic sugar for `Sort 0`, the very bottom of the type hierarchy described in the last chapter. Moreover, `Type u` is also just syntactic sugar for `Sort (u+1)`. `Prop` has some special features, but like the other type universes, it is closed under the arrow constructor: if we have `p q : Prop`, then `p → q : Prop`.

There are at least two ways of thinking about propositions as types. To some who take a constructive view of logic and mathematics, this is a faithful rendering of what it means to be a proposition: a proposition `p` represents a sort of data type, namely, a specification of the type of data that constitutes a proof. A proof of `p` is then simply an object `t : p` of the right type.

Those not inclined to this ideology can view it, rather, as a simple coding trick. To each proposition `p` we associate a type, which is empty if `p` is false and has a single element, say `*`, if `p` is true. In the latter case, let us say that (the type associated with) `p` is *inhabited*. It just so happens that the rules for function application and abstraction can conveniently help us keep track of which elements of *Prop* are inhabited. So constructing an element `t : p` tells us that `p` is indeed true. You can think of the inhabitant of `p` as being the "fact that `p` is true." A proof of `p → q` uses "the fact that `p` is true" to obtain "the fact that `q` is true."

Indeed, if `p : Prop` is any proposition, Lean's kernel treats any two elements `t1 t2 : p` as being definitionally equal, much the same way as it treats `(λ x, t)s` and `t[s/x]` as definitionally equal. This is known as "proof irrelevance," and is consistent with the interpretation in the last paragraph. It means that even though we can treat proofs `t : p` as ordinary objects in the language of dependent type theory, they carry no information beyond the fact that `p` is true.

The two ways we have suggested thinking about the propositions-as-types paradigm differ in a fundamental way. From the constructive point of view, proofs are abstract mathematical objects that are *denoted* by suitable expressions in dependent type theory. In contrast, if we think in terms of the coding trick described above, then the expressions themselves do not denote anything interesting. Rather, it is the fact that we can write them down and check that they are well-typed that ensures that the proposition in question is true. In other words, the expressions *themselves* are the proofs.

In the exposition below, we will slip back and forth between these two ways of talking, at times saying that an expression "constructs" or "produces" or "returns" a proof of a proposition, and at other times simply saying that it "is" such a proof. This is similar to the way that computer scientists occasionally blur the distinction between syntax and semantics by saying, at times, that a program "computes" a certain function, and at other times speaking as though the program "is" the function in question.

In any case, all that really matters is that the bottom line is clear. To formally express a mathematical assertion in the language of dependent type theory, we need to exhibit a term `p : Prop`. To *prove* that assertion, we need to exhibit a term `t : p`. Lean's task, as a proof assistant, is to help us to construct such a term, `t`, and to verify that it is well-formed and has the correct type.

3.2 Working with Propositions as Types

In the propositions-as-types paradigm, theorems involving only → can be proved using lambda abstraction and application. In Lean, the `theorem` command introduces a new theorem:

```
constants p q : Prop

theorem t1 : p → q → p := λ hp : p, λ hq : q, hp
```

This looks exactly like the definition of the constant function in the last chapter, the only difference being that the arguments are elements of `Prop` rather than `Type`. Intuitively, our proof of `p → q → p` assumes `p` and `q` are true, and uses the first hypothesis (trivially) to establish that the conclusion, `p`, is true.

Note that the `theorem` command is really a version of the `definition` command: under the propositions and types correspondence, proving the theorem `p → q → p` is really the same as defining an element of the associated type. To the kernel type checker, there is no difference between the two.

There are a few pragmatic differences between definitions and theorems, however.

In normal circumstances, it is never necessary to unfold the "definition" of a theorem; by proof irrelevance, any two proofs of that theorem are definitionally equal. Once the proof of a theorem is complete, typically we only need to know that the proof exists; it doesn't matter what the proof is. In light of that fact, Lean tags proofs as *irreducible*, which serves as a hint to the parser (more precisely, the *elaborator*) that there is generally no need to unfold it when processing a file. Moreover, for efficiency purposes, Lean treats theorems as axiomatic constants within the file in which they are defined. This makes it possible to process and check theorems in parallel, since theorems later in a file do not make use of the contents of earlier proofs.

As with definitions, the `print` command will show you the proof of a theorem.

```
theorem t1 : p → q → p := λ hp : p, λ hq : q, hp

print t1
```

(To save space, the online version of Lean does not store proofs of theorems in the library, so you cannot print them in the browser interface.)

Notice that the lambda abstractions `hp : p` and `hq : q` can be viewed as temporary assumptions in the proof of `t1`. Lean provides the alternative syntax `assume` for such a lambda abstraction:

```
theorem t1 : p → q → p :=
assume hp : p,
assume hq : q,
hp
```

Lean also allows us to specify the type of the final term `hp`, explicitly, with a `show` statement.

```
theorem t1 : p → q → p :=
assume hp : p,
assume hq : q,
show p, from hp
```

Adding such extra information can improve the clarity of a proof and help detect errors when writing a proof. The `show` command does nothing more than annotate the type, and, internally, all the presentations of `t1` that we have seen produce the same term. Lean also allows you to use the alternative syntax `proposition`, `lemma`, or `corollary` instead of theorem:

```
lemma t1 : p → q → p :=
assume hp : p,
assume hq : q,
show p, from hp
```

As with ordinary definitions, one can move the lambda-abstracted variables to the left of the colon:

```
theorem t1 (hp : p) (hq : q) : p := hp

check t1 -- p → q → p
```

Now we can apply the theorem `t1` just as a function application.

```
axiom hp : p

theorem t2 : q → p := t1 hp
```

Here, the `axiom` command is alternative syntax for `constant`. Declaring a "constant" `hp : p` is tantamount to declaring that `p` is true, as witnessed by `hp`. Applying the theorem `t1 : p → q → p` to the fact `hp : p` that `p` is true yields the theorem `t2 : q → p`.

Notice, by the way, that the original theorem `t1` is true for *any* propositions `p` and `q`, not just the particular constants declared. So it would be more natural to define the theorem so that it quantifies over those, too:

```
theorem t1 (p q : Prop) (hp : p) (hq : q) : p := hp
check t1
```

The type of `t1` is now `∀ p q : Prop, p → q → p`. We can read this as the assertion "for every pair of propositions `p q`, we have `p → q → p`". The symbol ∀ is alternate syntax for Π, and later we will see how Pi types let us model universal quantifiers more generally. For the moment, however, we will focus on theorems in propositional logic, generalized over the propositions. We will tend to work in sections with variables over the propositions, so that they are generalized for us automatically.

When we generalize `t1` in that way, we can then apply it to different pairs of propositions, to obtain different instances of the general theorem.

```
theorem t1 (p q : Prop) (hp : p) (hq : q) : p := hp

variables p q r s : Prop

check t1 p q                -- p → q → p
check t1 r s                -- r → s → r
check t1 (r → s) (s → r)    -- (r → s) → (s → r) → r → s

variable h : r → s
check t1 (r → s) (s → r) h  -- (s → r) → r → s
```

Remember that under the propositions-as-types correspondence, a variable `h` of type `r → s` can be viewed as the hypothesis, or premise, that `r → s` holds. For that reason, Lean offers the alternative syntax, `premise`, for `variable`.

```
premise h : r → s
check t1 (r → s) (s → r) h
```

As another example, let us consider the composition function discussed in the last chapter, now with propositions instead of types.

```
variables p q r s : Prop

theorem t2 (h1 : q → r) (h2 : p → q) : p → r :=
assume h3 : p,
show r, from h1 (h2 h3)
```

As a theorem of propositional logic, what does `t2` say?

Lean allows the alternative syntax `premise` and `premises` for `variable` and `variables`. This makes sense, of course, for variables whose type is an element of `Prop`. It is also often useful to use numeric unicode subscripts, entered as `\0`, `\1`, `\2`, ..., for hypotheses. The following definition of `t2` has the same net effect as the preceding one.

```
variables p q r s : Prop
premises (h₁ : q → r) (h₂ : p → q)

theorem t2 : p → r :=
assume h₃ : p,
show r, from h₁ (h₂ h₃)
```

3.3 Propositional Logic

Lean defines all the standard logical connectives and notation. The propositional connectives come with the following notation:

Ascii	Unicode	Emacs shortcut for unicode	Definition
true			true
false			false
not	¬	`\not`, `\neg`	not
/\	∧	`\and`	and
\/	∨	`\or`	or
->	→	`\to`, `\r`, `\implies`	
<->	↔	`\iff`, `\lr`	iff

They all take values in `Prop`.

```
variables p q : Prop

check p → q → p ∧ q
check ¬p → p ↔ false
check p ∨ q → q ∨ p
```

The order of operations is fairly standard: unary negation ¬ binds most strongly, then ∧ and ∨, and finally → and ↔. For example, `a ∧ b → c ∨ d ∧ e` means `(a ∧ b) → (c ∨ (d ∧ e))`. Remember that → associates to the right (nothing changes now that the arguments are elements of `Prop`, instead of some other `Type`), as do the other binary connectives. So if we have `p q r : Prop`, the expression `p → q → r` reads "if `p`, then if `q`, then `r`." This is just the "curried" form of `p ∧ q → r`.

In the last chapter we observed that lambda abstraction can be viewed as an "introduction rule" for →. In the current setting, it shows how to "introduce" or establish an implication. Application can be viewed as an "elimination rule," showing how to "eliminate" or use an implication in a proof. The other propositional connectives are defined in the standard library in the file `init.datatypes`, and each comes with its canonical introduction and elimination rules.

Conjunction

The expression `and.intro h1 h2` creates a proof for `p ∧ q` using proofs `h1 : p` and `h2 : q`. It is common to describe `and.intro` as the *and-introduction* rule. In the next example we use `and.intro` to create a proof of `p → q → p ∧ q`.

```
example (hp : p) (hq : q) : p ∧ q := and.intro hp hq

check assume (hp : p) (hq : q), and.intro hp hq
```

The `example` command states a theorem without naming it or storing it in the permanent context. Essentially, it just checks that the given term has the indicated type. It is convenient for illustration, and we will use it often.

The expression `and.elim_left H` creates a proof of `p` from a proof `h : p ∧ q`. Similarly, `and.elim_right H` is a proof of `q`. They are commonly known as the right and left *and-elimination* rules.

```
example (h : p ∧ q) : p := and.elim_left h
example (h : p ∧ q) : q := and.elim_right h
```

Because they are so commonly used, the standard library provides the abbreviations `and.left` and `and.right` for `and.elim_left` and `and.elim_right`, respectively.

We can now prove `p ∧ q → q ∧ p` with the following proof term.

```
example (h : p ∧ q) : q ∧ p :=
and.intro (and.right h) (and.left h)
```

Notice that and-introduction and and-elimination are similar to the pairing and projection operations for the cartesian product. The difference is that given `hp : p` and `hq : q`, `and.intro hp hq` has type `p ∧ q : Prop`, while `pair hp hq` has type `p × q : Type`. The similarity between $\wedge$ and $\times$ is another instance of the Curry-Howard isomorphism, but in contrast to implication and the function space constructor, $\wedge$ and $\times$ are treated separately in Lean. With the analogy, however, the proof we have just constructed is similar to a function that swaps the elements of a pair.

We will see in a later chapter that certain types in Lean are *structures*, which is to say, the type is defined with a single canonical *constructor* which builds an element of the type from a sequence of suitable arguments. For every `p q : Prop`, `p ∧ q` is an example: the canonical way to construct an element is to apply `and.intro` to suitable arguments `hp : p` and `hq : q`. Lean allows us to use *anonymous constructor* notation ⟨`arg1, arg2, ...`⟩ in situations like these, when the relevant type is an inductive type and can be inferred from the context. In particular, we can often write ⟨`hp, hq`⟩ instead of `and.intro hp hq`:

```
variables p q : Prop
premises  (hp : p) (hq : q)

check (⟨hp, hq⟩ : p ∧ q)
```

These angle brackets are obtained by typing `\<` and `\>`, respectively. Alternatively, you can use ASCII equivalents `(|` and `|)`:

```
variables p q : Prop
premises  (hp : p) (hq : q)

example : p ∧ q := (|hp, hq|)
```

Lean provides another useful syntactic gadget. Given an expression `e` of an inductive type `foo` (possibly applied to some arguments), the notation `e^.bar` is shorthand for `foo.bar e`. This provides a convenient way of accessing functions without opening a namespace. For example, the following two expressions mean the same thing:

```
variable l : list ℕ

check list.head l
check l^.head
```

As a result, given `h : p ∧ q`, we can write `h^.left` for `and.left h` and `h^.right` for `and.right h`. We can therefore rewrite the sample proof above conveniently as follows:

```
example (h : p ∧ q) : q ∧ p :=
⟨h^.right, h^.left⟩
```

There is a fine line between brevity and obfuscation, and omitting information in this way can sometimes make a proof harder to read. But for straightforward constructions like the one above, when the type of `h` and the goal of the construction as salient, the notation is clean and effective.

It is common to iterate constructions like "and." Lean also allows you to flatten nested constructors that associate to the right, so that these two proofs are equivalent:

```
example (h : p ∧ q) : q ∧ p ∧ q:=
⟨h^.right, ⟨h^.left, h^.right⟩⟩

example (h : p ∧ q) : q ∧ p ∧ q:=
⟨h^.right, h^.left, h^.right⟩
```

This is often useful as well.

Disjunction

The expression `or.intro_left q hp` creates a proof of `p ∨ q` from a proof `hp : p`. Similarly, `or.intro_right p hq` creates a proof for `p ∨ q` using a proof `hq : q`. These are the left and right *or-introduction* rules.

```
example (hp : p) : p ∨ q := or.intro_left q hp
example (hq : q) : p ∨ q := or.intro_right p hq
```

The *or-elimination* rule is slightly more complicated. The idea is that we can prove `r` from `p ∨ q`, by showing that `r` follows from `p` and that `r` follows from `q`. In other words, it is a proof "by cases." In the expression `or.elim hpq hpr hqr`, `or.elim` takes three arguments, `hpq : p ∨ q`, `hpr : p → r` and `hqr : q → r`, and produces a proof of `r`. In the following example, we use `or.elim` to prove `p ∨ q → q ∨ p`.

```
example (h : p ∨ q) : q ∨ p :=
or.elim h
  (assume hp : p,
    show q ∨ p, from or.intro_right q hp)
  (assume hq : q,
    show q ∨ p, from or.intro_left p hq)
```

In most cases, the first argument of `or.intro_right` and `or.intro_left` can be inferred automatically by Lean. Lean therefore provides `or.inr` and `or.inl` as shorthands for `or.intro_right _` and `or.intro_left _`. Thus the proof term above could be written more concisely:

```
example (h : p ∨ q) : q ∨ p := or.elim h (λ hp, or.inr hp) (λ hq, or.inl hq)
```

Notice that there is enough information in the full expression for Lean to infer the types of `hp` and `hq` as well. But using the type annotations in the longer version makes the proof more readable, and can help catch and debug errors.

Because `or` has two constructors, we cannot use anonymous constructor notation. But we can still write `h^.elim` instead of `or.elim h`:

```
example (h : p ∨ q) : q ∨ p :=
h^.elim
  (assume hp : p, or.inr hp)
  (assume hq : q, or.inl hq)
```

Once again, you should exercise judgment as to whether such abbreviations enhance or diminish readability.

Negation and Falsity

The expression `not.intro h` produces a proof of `¬p` from `h : p → false`. That is, we obtain `¬p` if we can derive a contradiction from `p`. Similarly, the expression `hnp hp` produces a proof of `false` from `hp : p` and `hnp : ¬p`. The next example uses both these rules to produce a proof of `(p → q) → ¬q → ¬p`.

```
example (hpq : p → q) (hnq : ¬q) : ¬p :=
assume hp : p,
show false, from hnq (hpq hp)
```

The connective `false` has a single elimination rule, `false.elim`, which expresses the fact that anything follows from a contradiction. This rule is sometimes called *ex falso* (short for *ex falso sequitur quodlibet*), or the *principle of explosion*.

```
example (hp : p) (hnp : ¬p) : q := false.elim (hnp hp)
```

The arbitrary fact, `q`, that follows from falsity is an implicit argument in `false.elim` and is inferred automatically. This pattern, deriving an arbitrary fact from contradictory hypotheses, is quite common, and is represented by `absurd`.

```
example (hp : p) (hnp : ¬p) : q := absurd hp hnp
```

Here, for example, is a proof of `¬p → q → (q → p) → r`:

```
example (hnp : ¬p) (hq : q) (hqp : q → p) : r :=
absurd (hqp hq) hnp
```

Incidentally, just as `false` has only an elimination rule, `true` has only an introduction rule, `true.intro : true`, sometimes abbreviated `trivial : true`. In other words, `true` is simply true, and has a canonical proof, `trivial`.

Logical Equivalence

The expression `iff.intro h1 h2` produces a proof of `p ↔ q` from `h1 : p → q` and `h2 : q → p`. The expression `iff.elim_left H` produces a proof of `p → q` from `h : p ↔ q`. Similarly, `iff.elim_right H` produces a proof of `q → p` from `h : p ↔ q`. Here is a proof of `p ∧ q ↔ q ∧ p`:

```
theorem and_swap : p ∧ q ↔ q ∧ p :=
iff.intro
  (assume h : p ∧ q,
    show q ∧ p, from and.intro (and.right h) (and.left h))
  (assume h : q ∧ p,
```

```
  show p ∧ q, from and.intro (and.right h) (and.left h))

check and_swap p q    -- p ∧ q ↔ q ∧ p
```

Because they represent a form of *modus ponens*, `iff.elim_left` and `iff.elim_right` can be abbreviated `iff.mp` and `iff.mpr`, respectively. In the next example, we use that theorem to derive `q ∧ p` from `p ∧ q`:

```
premise h : p ∧ q
example : q ∧ p := iff.mp (and_swap p q) h
```

Because `iff` is defined internally from `and`, we can use the anonymous constructor notation to construct a proof of `p ↔ q` from proofs of the forward and backward directions. We can also use the `.^` notation with `mp` and `mpr`. The previous examples can therefore be written concisely as follows:

```
theorem and_swap : p ∧ q ↔ q ∧ p :=
⟨ λ h, ⟨h^.right, h^.left⟩, λ h, ⟨h^.right, h^.left⟩ ⟩

example (h : p ∧ q) : q ∧ p := (and_swap p q)^.mp h
```

3.4 Introducing Auxiliary Subgoals

This is a good place to introduce another device Lean offers to help structure long proofs, namely, the `have` construct, which introduces an auxiliary subgoal in a proof. Here is a small example, adapted from the last section:

```
variables p q : Prop

example (h : p ∧ q) : q ∧ p :=
have hp : p, from and.left h,
have hq : q, from and.right h,
show q ∧ p, from and.intro hq hp
```

Internally, the expression `have h : p, from s, t` produces the term `(λ (h : p), t) s`. In other words, `s` is a proof of `p`, `t` is a proof of the desired conclusion assuming `h : p`, and the two are combined by a lambda abstraction and application. This simple device is extremely useful when it comes to structuring long proofs, since we can use intermediate `have`'s as stepping stones leading to the final goal.

Lean also supports a structured way of reasoning backwards from a goal, which models the "suffices to show" construction in ordinary mathematics. The next example simply permutes that last two lines in the previous proof.

```
variables p q : Prop

example (h : p ∧ q) : q ∧ p :=
have hp : p, from and.left h,
suffices hq : q, from and.intro hq hp,
show q, from and.right h
```

Writing `suffices hq : q` leaves us with two goals. First, we have to show that it indeed suffices to show `q`, by proving the original goal of `q ∧ p` with the additional hypothesis `hq : q`. Finally, we have to show `q`.

3.5 Classical Logic

The introduction and elimination rules we have seen so far are all constructive, which is to say, they reflect a computational understanding of the logical connectives based on the propositions-as-types correspondence. Ordinary classical logic adds to this the law of the excluded middle, `p ∨ ¬p`. To use this principle, you have to open the classical namespace.

```
open classical

variable p : Prop
check em p
```

Intuitively, the constructive "or" is very strong: asserting `p ∨ q` amounts to knowing which is the case. If `RH` represents the Riemann hypothesis, a classical mathematician is willing to assert `RH ∨ ¬RH`, even though we cannot yet assert either disjunct.

One consequence of the law of the excluded middle is the principle of double-negation elimination:

```
theorem dne {p : Prop} (h : ¬¬p) : p :=
or.elim (em p)
  (assume hp : p, hp)
  (assume hnp : ¬p, absurd hnp h)
```

Double-negation elimination allows one to prove any proposition, `p`, by assuming `¬p` and deriving `false`, because that amounts to proving `¬¬p`. In other words, double-negation elimination allows one to carry out a proof by contradiction, something which is not generally possible in constructive logic. As an exercise, you might try proving the converse, that is, showing that `em` can be proved from `dne`.

The classical axioms also gives you access to additional patterns of proof that can be justified by appeal to `em`. For example, one can carry out a proof by cases:

```
example (h : ¬¬p) : p :=
by_cases
```

```
  (assume h1 : p, h1)
  (assume h1 : ¬p, absurd h1 h)
```

Or you can carry out a proof by contradiction:

```
example (h : ¬¬p) : p :=
by_contradiction
  (assume h1 : ¬p,
    show false, from h h1)
```

If you are not used to thinking constructively, it may take some time for you to get a sense of where classical reasoning is used. It is needed in the following example because, from a constructive standpoint, knowing that `p` and `q` are not both true does not necessarily tell you which one is false:

```
example (h : ¬ (p ∧ q)) : ¬ p ∨ ¬ q :=
or.elim (em p)
  (assume hp : p,
    or.inr
      (show ¬q, from
        assume hq : q,
        h ⟨hp, hq⟩))
  (assume hp : ¬p,
    or.inl hp)
```

We will see later that there *are* situations in constructive logic where principles like excluded middle and double-negation elimination are permissible, and Lean supports the use of classical reasoning in such contexts.

There are additional classical axioms that are not included by default in the standard library. We will discuss these in detail in a later chapter.

3.6 Examples of Propositional Validities

Lean's standard library contains proofs of many valid statements of propositional logic, all of which you are free to use in proofs of your own. In this section, we will review some common identities, and encourage you to try proving them on your own using the rules above.

The following is a long list of assertions in propositional logic. Prove as many as you can, using the rules introduced above to replace the `sorry` placeholders by actual proofs. The ones that require classical reasoning are grouped together at the end, while the rest are constructively valid.

```
open classical
```

```
variables p q r s : Prop

-- commutativity of ∧ and ∨
example : p ∧ q ↔ q ∧ p := sorry
example : p ∨ q ↔ q ∨ p := sorry

-- associativity of ∧ and ∨
example : (p ∧ q) ∧ r ↔ p ∧ (q ∧ r) := sorry
example : (p ∨ q) ∨ r ↔ p ∨ (q ∨ r) := sorry

-- distributivity
example : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) := sorry
example : p ∨ (q ∧ r) ↔ (p ∨ q) ∧ (p ∨ r) := sorry

-- other properties
example : (p → (q → r)) ↔ (p ∧ q → r) := sorry
example : ((p ∨ q) → r) ↔ (p → r) ∧ (q → r) := sorry
example : ¬(p ∨ q) ↔ ¬p ∧ ¬q := sorry
example : ¬p ∨ ¬q → ¬(p ∧ q) := sorry
example : ¬(p ∧ ¬ p) := sorry
example : p ∧ ¬q → ¬(p → q) := sorry
example : ¬p → (p → q) := sorry
example : (¬p ∨ q) → (p → q) := sorry
example : p ∨ false ↔ p := sorry
example : p ∧ false ↔ false := sorry
example : ¬(p ↔ ¬p) := sorry
example : (p → q) → (¬q → ¬p) := sorry

-- these require classical reasoning
example : (p → r ∨ s) → ((p → r) ∨ (p → s)) := sorry
example : ¬(p ∧ q) → ¬p ∨ ¬q := sorry
example : ¬(p → q) → p ∧ ¬q := sorry
example : (p → q) → (¬p ∨ q) := sorry
example : (¬q → ¬p) → (p → q) := sorry
example : p ∨ ¬p := sorry
example : (((p → q) → p) → p) := sorry
```

The `sorry` identifier magically produces a proof of anything, or provides an object of any data type at all. Of course, it is unsound as a proof method – for example, you can use it to prove `false` – and Lean produces severe warnings when files use or import theorems which depend on it. But it is very useful for building long proofs incrementally. Start writing the proof from the top down, using `sorry` to fill in subproofs. Make sure Lean accepts the term with all the `sorry`'s; if not, there are errors that you need to correct. Then go back and replace each `sorry` with an actual proof, until no more remain.

Here is another useful trick. Instead of using `sorry`, you can use an underscore `_` as a placeholder. Recall that this tells Lean that the argument is implicit, and should be filled in automatically. If Lean tries to do so and fails, it returns with an error message "don't know how to synthesize placeholder." This is followed by the type of the term it is expecting, and all the objects and hypothesis available in the context. In other words, for each unresolved placeholder, Lean reports the subgoal that needs to be filled at that point. You can then construct a proof by incrementally filling in these placeholders.

For reference, here are two sample proofs of validities taken from the list above.

```
open classical

variables p q r : Prop

-- distributivity
example : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
iff.intro
  (assume h : p ∧ (q ∨ r),
    have hp : p, from h^.left,
    or.elim (h^.right)
      (assume hq : q,
        show (p ∧ q) ∨ (p ∧ r), from or.inl ⟨hp, hq⟩)
      (assume hr : r,
        show (p ∧ q) ∨ (p ∧ r), from or.inr ⟨hp, hr⟩))
  (assume h : (p ∧ q) ∨ (p ∧ r),
    or.elim h
      (assume hpq : p ∧ q,
        have hp : p, from hpq^.left,
        have hq : q, from hpq^.right,
        show p ∧ (q ∨ r), from ⟨hp, or.inl hq⟩)
      (assume hpr : p ∧ r,
        have hp : p, from hpr^.left,
        have hr : r, from hpr^.right,
        show p ∧ (q ∨ r), from ⟨hp, or.inr hr⟩))

-- an example that requires classical reasoning
example : ¬(p ∧ ¬q) → (p → q) :=
assume h : ¬(p ∧ ¬q),
assume hp : p,
show q, from
  or.elim (em q)
    (assume hq : q, hq)
    (assume hnq : ¬q, absurd (and.intro hp hnq) h)
```

4

Quantifiers and Equality

The last chapter introduced you to methods that construct proofs of statements involving the propositional connectives. In this chapter, we extend the repertoire of logical constructions to include the universal and existential quantifiers, and the equality relation.

4.1 The Universal Quantifier

Notice that if α is any type, we can represent a unary predicate `p` on α as an object of type α → `Prop`. In that case, given `x :` α, `p x` denotes the assertion that `p` holds of `x`. Similarly, an object `r :` $\alpha \to \alpha \to$ `Prop` denotes a binary relation on α: given `x y :` α, `r x y` denotes the assertion that `x` is related to `y`.

The universal quantifier, `∀ x : α, p x` is supposed to denote the assertion that "for every `x : α`, `p x`" holds. As with the propositional connectives, in systems of natural deduction, "forall" is governed by an introduction and elimination rule. Informally, the introduction rule states:

> Given a proof of `p x`, in a context where `x : α` is arbitrary, we obtain a proof `∀ x : α, p x`.

The elimination rule states:

> Given a proof `∀ x : α, p x` and any term `t : α`, we obtain a proof of `p t`.

As was the case for implication, the propositions-as-types interpretation now comes into play. Remember the introduction and elimination rules for Pi types:

> Given a term `t` of type β `x`, in a context where `x :` α is arbitrary, we have `(`λ `x :` α`, t) :` Π `x :` α`,` β `x`.

The elimination rule states:

> Given a term `s :` Π `x :` α`,` β `x` and any term `t :` α, we have `s t :` β `t`.

In the case where `p x` has type `Prop`, if we replace Π `x :` α`,` β `x` with $\forall$ `x :` α`, p x`, we can read these as the correct rules for building proofs involving the universal quantifier.

The Calculus of Inductive Constructions therefore identifies Π and $\forall$ in this way. If `p` is any expression, $\forall$ `x :` α`, p` is nothing more than alternative notation for Π `x :` α`, p`, with the idea that the former is more natural than the latter in cases where where `p` is a proposition. Typically, the expression `p` will depend on `x :` α. Recall that, in the case of ordinary function spaces, we could interpret $\alpha \to \beta$ as the special case of Π `x :` α`,` β in which β does not depend on `x`. Similarly, we can think of an implication `p` $\to$ `q` between propositions as the special case of $\forall$ `x : p, q` in which the expression `q` does not depend on `x`.

Here is an example of how the propositions-as-types correspondence gets put into practice.

```
variables (α : Type) (p q : α → Prop)

example : (∀ x : α, p x ∧ q x) → ∀ y : α, p y  :=
assume h : ∀ x : α, p x ∧ q x,
take y : α,
show p y, from (h y)^.left
```

As a notational convention, we give the universal quantifier the widest scope possible, so parentheses are needed to limit the quantifier over `x` to the hypothesis in the example above. The canonical way to prove $\forall$ `y :` α`, p y` is to take an arbitrary `y`, and prove `p y`. This is the introduction rule. Now, given that `h` has type $\forall$ `x :` α`, p x` $\wedge$ `q x`, the expression `h y` has type `p y` $\wedge$ `q y`. This is the elimination rule. Taking the left conjunct gives the desired conclusion, `p y`.

Remember that expressions which differ up to renaming of bound variables are considered to be equivalent. So, for example, we could have used the same variable, `x`, in both the hypothesis and conclusion, or chosen the variable `z` instead of `y` in the proof:

```
example : (∀ x : α, p x ∧ q x) → ∀ y : α, p y  :=
assume h : ∀ x : α, p x ∧ q x,
take z : α,
show p z, from and.elim_left (h z)
```

As another example, here is how we can express the fact that a relation, `r`, is transitive:

```
variables (α : Type) (r : α → α → Prop)
variable trans_r : ∀ x y z, r x y → r y z → r x z

variables (a b c : α)
variables (hab : r a b) (hbc : r b c)

check trans_r          -- ∀ (x y z : α), r x y → r y z → r x z
check trans_r a b c
check trans_r a b c hab
check trans_r a b c hab hbc
```

Think about what is going on here. When we instantiate `trans_r` at the values `a b c`, we end up with a proof of `r a b → r b c → r a c`. Applying this to the "hypothesis" `hab : r a b`, we get a proof of the implication `r b c → r a c`. Finally, applying it to the hypothesis `hbc` yields a proof of the conclusion `r a c`.

In situations like this, it can be tedious to supply the arguments `a b c`, when they can be inferred from `hab hbc`. For that reason, it is common to make these arguments implicit:

```
variables (α : Type) (r : α → α → Prop)
variable (trans_r : ∀ {x y z}, r x y → r y z → r x z)

variables (a b c : α)
variables (hab : r a b) (hbc : r b c)

check trans_r
check trans_r hab
check trans_r hab hbc
```

The advantage is that we can simply write `trans_r hab hbc` as a proof of `r a c`. A disadvantage is that Lean does not have enough information to infer the types of the arguments in the expressions `trans_r` and `trans_r hab`. The output of the `check` command contains expressions like `?z α r trans_r a b c hab hbc`. Such an expression indicates an arbitrary value, that may depend on any of the values listed (in this case, all the variables in the local context).

Here is an example of how we can carry out elementary reasoning with an equivalence relation:

```
variables (α : Type) (r : α → α → Prop)

variable refl_r : ∀ x, r x x
variable symm_r : ∀ {x y}, r x y → r y x
variable trans_r : ∀ {x y z}, r x y → r y z → r x z

example (a b c d : α) (hab : r a b) (hcb : r c b) (hcd : r c d) : r a d :=
trans_r (trans_r hab (symm_r hcb)) hcd
```

You might want to try to prove some of these equivalences:

```
variables (α : Type) (p q : α → Prop)

example : (∀ x, p x ∧ q x) ↔ (∀ x, p x) ∧ (∀ x, q x) := sorry
example : (∀ x, p x → q x) → (∀ x, p x) → (∀ x, q x) := sorry
example : (∀ x, p x) ∨ (∀ x, q x) → ∀ x, p x ∨ q x := sorry
```

You should also try to understand why the reverse implication is not derivable in the last example.

It is often possible to bring a component outside a universal quantifier, when it does not depend on the quantified variable (one direction of the second of these requires classical logic):

```
variables (α : Type) (p q : α → Prop)
variable r : Prop

example : α → ((∀ x : α, r) ↔ r) := sorry
example : (∀ x, p x ∨ r) ↔ (∀ x, p x) ∨ r := sorry
example : (∀ x, r → p x) ↔ (r → ∀ x, p x) := sorry
```

As a final example, consider the "barber paradox", that is, the claim that in a certain town there is a (male) barber that shaves all and only the men who do not shave themselves. Prove that this implies a contradiction:

```
variables (men : Type) (barber : men) (shaves : men → men → Prop)

example (h : ∀ x : men, shaves barber x ↔ ¬ shaves x x) : false := sorry
```

It is the typing rule for Pi types, and the universal quantifier in particular, that distinguishes `Prop` from other types. Suppose we have `α : Type i` and `β : Type j`, where the expression β may depend on a variable `x : α`. Then `Π x : α, β` is an element of `Type (imax i j)`, where `imax i j` is the maximum of `i` and `j` if `j` is not 0, and 0 otherwise.

The idea is as follows. If `j` is not `0`, then `Π x : α, β` is an element of `Type (max i j)`. In other words, the type of dependent functions from α to β "lives" in the universe with smallest index greater-than or equal to the indices of the universes of α and β. Suppose, however, that β is of `Type 0`, that is, an element of `Prop`. In that case, `Π x : α, β` is an element of `Type 0` as well, no matter which type universe α lives in. In other words, if β is a proposition depending on α, then `∀ x : α, β` is again a proposition. This reflects the interpretation of `Prop` as the type of propositions rather than data, and it is what makes `Prop` *impredicative*.

The term "predicative" stems from foundational developments around the turn of the twentieth century, when logicians such as Poincaré and Russell blamed set-theoretic paradoxes on the "vicious circles" that arise when we define a property by quantifying over a collection that includes the very property being defined. Notice that if α is any type, we can form the type $\alpha \to$ `Prop` of all predicates on α (the "power type of α"). The impredicativity of Prop means that we can form propositions that quantify over $\alpha \to$ `Prop`. In particular, we can define predicates on α by quantifying over all predicates on α, which is exactly the type of circularity that was once considered problematic.

4.2 Equality

Let us now turn to one of the most fundamental relations defined in Lean's library, namely, the equality relation. In a later chapter, we will explain *how* equality is defined from the primitives of Lean's logical framework. In the meanwhile, here we explain how to use it.

Of course, a fundamental property of equality is that it is an equivalence relation:

```
check eq.refl    -- ∀ (a : ?M_1), a = a
check eq.symm    -- ?M_2 = ?M_3 → ?M_3 = ?M_2
check eq.trans   -- ?M_2 = ?M_3 → ?M_3 = ?M_4 → ?M_2 = ?M_4
```

Thus, for example, we can specialize the example from the previous section to the equality relation:

```
variables (α : Type) (a b c d : α)
premises (hab : a = b) (hcb : c = b) (hcd : c = d)

example : a = d :=
eq.trans (eq.trans hab (eq.symm hcb)) hcd
```

If we "open" the eq namespace, the names become shorter:

```
open eq

example : a = d := trans (trans hab (symm hcb)) hcd
```

We can also use the projection notation:

```
example : a = d := (hab^.trans hcb^.symm)^.trans hcd
```

Reflexivity is more powerful than it looks. Recall that terms in the Calculus of Inductive Constructions have a computational interpretation, and that the logical framework treats terms with a common reduct as the same. As a result, some nontrivial identities can be proved by reflexivity:

```
variables (α β : Type)

example (f : α → β) (a : α) : (λ x, f x) a = f a := eq.refl _
example (a : α) (b : α) : (a, b).1 = a := eq.refl _
example : 2 + 3 = 5 := eq.refl _
```

This feature of the framework is so important that the library defines a notation `rfl` for `eq.refl _`:

```
example (f : α → β) (a : α) : (λ x, f x) a = f a := rfl
example (a : α) (b : α) : (a, b).1 = a := rfl
example : 2 + 3 = 5 := rfl
```

Equality is much more than an equivalence relation, however. It has the important property that every assertion respects the equivalence, in the sense that we can substitute equal expressions without changing the truth value. That is, given `h1 : a = b` and `h2 : P a`, we can construct a proof for `P b` using substitution: `eq.subst h1 h2`.

```
example (α : Type) (a b : α) (P : α → Prop) (h1 : a = b) (h2 : P a) : P b :=
eq.subst h1 h2

example (α : Type) (a b : α) (P : α → Prop) (h1 : a = b) (h2 : P a) : P b :=
h1 ▸ h2
```

The triangle in the second presentation is nothing more than notation for `eq.subst`, and you can enter it by typing `\t`.

The rule `eq.subst` is used to define the following auxiliary rules, which carry out more explicit substitutions. They are designed to deal with applicative terms, that is, terms of form `s t`. Specifically, `congr_arg` can be used to replace the argument, `congr_fun` can be used to replace the terms that is being applied, and `congr` can be used to replace both at once.

```
variable α : Type
variables a b : α
variables f g : α → ℕ
premise h₁ : a = b
premise h₂ : f = g

example : f a = f b := congr_arg f h₁
example : f a = g a := congr_fun h₂ a
example : f a = g b := congr h₂ h₁
```

Lean's library contains a large number of common identities, such as these:

```
variables (α : Type) [comm_ring α]
variables a b c d : α

example : a + 0 = a := add_zero a
example : 0 + a = a := zero_add a
example : a * 1 = a := mul_one a
example : 1 * a = a := one_mul a
example : -a + a = 0 := neg_add_self a
example : a + -a = 0 := add_neg_self a
example : a - a = 0 := sub_self a
```

```
example : a + b = b + a := add_comm a b
example : a + b + c = a + (b + c) := add_assoc a b c
example : a * b = b * a := mul_comm a b
example : a * b * c = a * (b * c) := mul_assoc a b c
example : a * (b + c) = a * b + a * c := mul_add a b c
example : a * (b + c) = a * b + a * c := left_distrib a b c -- alternative name
example : (a + b) * c = a * c + b * c := add_mul a b c
example : (a + b) * c = a * c + b * c := right_distrib a b c -- alternative name
example : a * (b - c) = a * b - a * c := mul_sub a b c
example : (a - b) * c = a * c - b * c := sub_mul a b c
```

Identities likes these are designed to work in arbitrary instances of the relevant algebraic structures, using the type class mechanism that is described in Chapter 10. Chapter 6 provides some pointers as to how to find them in the library. Here is an example of a calculation in the natural numbers that uses substitution combined with associativity, commutativity, and distributivity of the natural numbers.

```
variables x y z : ℕ

example (x y z : ℕ) : x * (y + z) = x * y + x * z := mul_add x y z
example (x y z : ℕ) : x * (y + z) = x * y + x * z := mul_add x y z
example (x y z : ℕ) : x + y + z = x + (y + z) := add_assoc x y z

example (x y : ℕ) : (x + y) * (x + y) = x * x + y * x + x * y + y * y :=
have h1 : (x + y) * (x + y) = (x + y) * x + (x + y) * y, from mul_add (x + y) x y,
have h2 : (x + y) * (x + y) = x * x + y * x + (x * y + y * y),
  from (add_mul x y x) ▶ (add_mul x y y) ▶ h1,
h2^.trans (add_assoc (x * x + y * x)(x * y) (y * y))^.symm
```

It is often important to be able to carry out substitutions by hand, but it is tedious to prove examples like the one above in this way. Fortunately, Lean provides an environment that provides better support for such calculations, which we will turn to now.

4.3 Calculational Proofs

A calculational proof is just a chain of intermediate results that are meant to be composed by basic principles such as the transitivity of equality. In Lean, a calculation proof starts with the keyword `calc`, and has the following syntax:

```
calc
  <expr>_0  'op_1'  <expr>_1  ':'  <proof>_1
    '...'   'op_2'  <expr>_2  ':'  <proof>_2
     ...
    '...'   'op_n'  <expr>_n  ':'  <proof>_n
```

Each `<proof>_i` is a proof for `<expr>_{i-1} op_i <expr>_i`. The `<proof>_i` may also be of the form `{ <pr> }`, where `<pr>` is a proof for some equality `a = b`. The form `{ <pr>`

} is just syntactic sugar for eq.subst <pr> (eq.refl <expr>_{i-1}) In other words, we are claiming we can obtain <expr>_i by replacing a with b in <expr>_{i-1}.

Here is an example:

```
variables (a b c d e : ℕ)
variable h1 : a = b
variable h2 : b = c + 1
variable h3 : c = d
variable h4 : e = 1 + d

theorem T : a = e :=
calc
  a     = b      : h1
    ... = c + 1  : h2
    ... = d + 1  : congr_arg _ h3
    ... = 1 + d  : add_comm d (1 : ℕ)
    ... =  e     : eq.symm h4
```

The style of writing proofs is most effective when it is used in conjunction with the simp and rewrite tactics, which are discussed in greater detail in the next chapter. For example, using the abbreviation rw for rewrite, the proof above could be written as follows:

```
variables (a b c d e : ℕ)
variable h1 : a = b
variable h2 : b = c + 1
variable h3 : c = d
variable h4 : e = 1 + d

include h1 h2 h3 h4
theorem T : a = e :=
calc
  a     = b      : by rw h1
    ... = c + 1  : by rw h2
    ... = d + 1  : by rw h3
    ... = 1 + d  : by rw add_comm
    ... =  e     : by rw h4
```

In the next chapter, we will see that hypotheses can be introduced, renamed, and modified by tactics, so it is not always clear what the names in rw h1 refer to (though, in this case, it is). For that reason, section variables and premises that only appear in a tactic command or block are not automatically add to the context. The include command takes care of that. Essentially, the rewrite tactic uses a given equality (which can a hypothesis, a theorem name, or a complex term) to "rewrite" the goal. If doing so reduces the goal to an identity t = t, the tactic applies reflexivity to prove it.

Rewrites can applied sequentially, so that the proof above can be shortened to this:

```
theorem T : a = e :=
calc
  a     = d + 1  : by rw [h1, h2, h3]
```

```
    ... = 1 + d  : by rw add_comm
    ... =  e     : by rw h4
```

Or even this:

```
theorem T : a = e :=
by rw [h1, h2, h3, add_comm, h4]
```

The `simp` tactic, instead, rewrites the goal by applying the given identities repeatedly, in any order, anywhere they are applicable in a term. It also uses other rules that have been previously declared to the system, and applies associativity and commutativity wisely to put expressions in canonical forms. As a result, we can also prove the theorem as follows:

```
theorem T : a = e :=
by simp [h1, h2, h3, h4]
```

We will discuss variations of `rw` and `simp` in the next chapter.

The `calc` command can be configured for any relation that supports some form of transitivity. It can even combine different relations.

```
theorem T2 (a b c d : ℕ)
  (h1 : a = b) (h2 : b ≤ c) (h3 : c + 1 < d) : a < d :=
calc
  a     = b     : h1
    ... < b + 1 : nat.lt_succ_self b
    ... ≤ c + 1 : nat.succ_le_succ h2
    ... < d     : h3
```

With `calc`, we can write the proof in the last section in a more natural and perspicuous way.

```
example (x y : ℕ) : (x + y) * (x + y) = x * x + y * x + x * y + y * y :=
calc
  (x + y) * (x + y) = (x + y) * x + (x + y) * y  : by rw mul_add
    ... = x * x + y * x + (x + y) * y            : by rw add_mul
    ... = x * x + y * x + (x * y + y * y)        : by rw add_mul
    ... = x * x + y * x + x * y + y * y          : by rw -add_assoc
```

Here the negation before `add_assoc` tells rewrite to use the identity in the opposite direction. If brevity is what we are after, both `rw` and `simp` can do the job on their own:

```
example (x y : ℕ) : (x + y) * (x + y) = x * x + y * x + x * y + y * y :=
by rw [mul_add, add_mul, add_mul, -add_assoc]

example (x y : ℕ) : (x + y) * (x + y) = x * x + y * x + x * y + y * y :=
by simp [mul_add, add_mul]
```

4.4 The Existential Quantifier

Finally, consider the existential quantifier, which can be written as either `exists x : α, p x` or `∃ x : α, p x`. Both versions are actually notationally convenient abbreviations for a more long-winded expression, `Exists (λ x : α, p x)`, defined in Lean's library.

As you should by now expect, the library includes both an introduction rule and an elimination rule. The introduction rule is straightforward: to prove `∃ x : α, p x`, it suffices to provide a suitable term `t` and a proof of `p t`. here are some examples:

```
open nat

example : ∃ x : ℕ, x > 0 :=
have h : 1 > 0, from zero_lt_succ 0,
exists.intro 1 h

example (x : ℕ) (h : x > 0) : ∃ y, y < x :=
exists.intro 0 h

example (x y z : ℕ) (hxy : x < y) (hyz : y < z) : ∃ w, x < w ∧ w < z :=
exists.intro y (and.intro hxy hyz)

check @exists.intro
```

We can use the anonymous constructor notation `⟨t, h⟩` for `exists.intro t h`, when the type is clear from the context.

```
example : ∃ x : ℕ, x > 0 :=
⟨1, zero_lt_succ 0⟩

example (x : ℕ) (h : x > 0) : ∃ y, y < x :=
⟨0, h⟩

example (x y z : ℕ) (hxy : x < y) (hyz : y < z) : ∃ w, x < w ∧ w < z :=
⟨y, hxy, hyz⟩
```

Note that `exists.intro` has implicit arguments: Lean has to infer the predicate `p : α → Prop` in the conclusion `∃ x, p x`. This is not a trivial affair. For example, if we have have `hg : g 0 0 = 0` and write `exists.intro 0 hg`, there are many possible values for the predicate `p`, corresponding to the theorems `∃ x, g x x = x`, `∃ x, g x x = 0`, `∃ x, g x 0 = x`, etc. Lean uses the context to infer which one is appropriate. This is illustrated in the following example, in which we set the option `pp.implicit` to true to ask Lean's pretty-printer to show the implicit arguments.

```
variable g : ℕ → ℕ → ℕ
variable hg : g 0 0 = 0

theorem gex1 : ∃ x, g x x = x := ⟨0, hg⟩
theorem gex2 : ∃ x, g x 0 = x := ⟨0, hg⟩
```

```
theorem gex3 : ∃ x, g 0 0 = x := ⟨0, hg⟩
theorem gex4 : ∃ x, g x x = 0 := ⟨0, hg⟩

set_option pp.implicit true  -- display implicit arguments
check gex1
check gex2
check gex3
check gex4
```

We can view `exists.intro` as an information-hiding operation: we are "hiding" the witness to the body of the assertion. The existential elimination rule, `exists.elim`, performs the opposite operation. It allows us to prove a proposition `q` from `∃ x : α, p x`, by showing that `q` follows from `p w` for an arbitrary value `w`. Roughly speaking, since we know there is an `x` satisfying `p x`, we can give it a name, say, `w`. If `q` does not mention `w`, then showing that `q` follows from `p w` is tantamount to showing the `q` follows from the existence of any such `x`. Here is an example:

```
variables (α : Type) (p q : α → Prop)

example (h : ∃ x, p x ∧ q x) : ∃ x, q x ∧ p x :=
exists.elim h
  (take w,
    assume hw : p w ∧ q w,
    show ∃ x, q x ∧ p x, from ⟨w, hw^.right, hw^.left⟩)
```

It may be helpful to compare the exists-elimination rule to the or-elimination rule: the assertion `∃ x : α, p x` can be thought of as a big disjunction of the propositions `p a`, as `a` ranges over all the elements of α.

Notice that an existential proposition is very similar to a sigma type, as described in Section 2.8. The difference is that given `a : α` and `h : p a`, the term `exists.intro a h` has type `(∃ x : α, p x) : Prop` and `sigma.mk a h` has type `(Σ x : α, p x) : Type`. The similarity between ∃ and Σ is another instance of the Curry-Howard isomorphism.

Lean provides a more convenient way to eliminate from an existential quantifier with the `match` statement:

```
variables (α : Type) (p q : α → Prop)

example (h : ∃ x, p x ∧ q x) : ∃ x, q x ∧ p x :=
match h with ⟨w, hw⟩ :=
  ⟨w, hw^.right, hw^.left⟩
end
```

The `match` statement is part of Lean's function definition system, which provides a convenient and expressive ways of defining complex functions. Once again, it is the Curry-Howard isomorphism that allows us to co-opt this mechanism for writing proofs as well.

The `match` statement "destructs" the existential assertion into the components `w` and `hw`, which can then be used in the body of the statement to prove the proposition. We can annotate the types used in the match for greater clarity:

```
example (h : ∃ x, p x ∧ q x) : ∃ x, q x ∧ p x :=
match h with ⟨(w : α), (hw : p w ∧ q w)⟩ :=
  ⟨w, hw^.right, hw^.left⟩
end
```

We can even use the match statement to decompose the conjunction at the same time:

```
example (h : ∃ x, p x ∧ q x) : ∃ x, q x ∧ p x :=
match h with ⟨w, hpw, hqw⟩ :=
  ⟨w, hqw, hpw⟩
end
```

Lean will even allow us to use an implicit `match` in the `assume` statement:

```
example : (∃ x, p x ∧ q x) → ∃ x, q x ∧ p x :=
assume ⟨w, hpw, hqw⟩, ⟨w, hqw, hpw⟩
```

In the following example, we define `even a` as `∃ b, a = 2*b`, and then we show that the sum of two even numbers is an even number.

```
def is_even (a : nat) := ∃ b, a = 2 * b

theorem even_plus_even {a b : nat} (h1 : is_even a) (h2 : is_even b) : is_even (a + b) :=
exists.elim h1 (take w1, assume hw1 : a = 2 * w1,
exists.elim h2 (take w2, assume hw2 : b = 2 * w2,
  exists.intro (w1 + w2)
    (calc
      a + b = 2 * w1 + 2 * w2  : by rw [hw1, hw2]
        ... = 2*(w1 + w2)      : by rw mul_add)))
```

Using the various gadgets described in this chapter — the match statement, anonymous constructors, and the `rewrite` tactic, we can write this proof concisely as follows:

```
theorem even_plus_even {a b : nat} (h1 : is_even a) (h2 : is_even b) : is_even (a + b) :=
match h1, h2 with
| ⟨w1, hw1⟩, ⟨w2, hw2⟩ := ⟨w1 + w2, by rw [hw1, hw2, mul_add]⟩
end
```

Just as the constructive "or" is stronger than the classical "or," so, too, is the constructive "exists" stronger than the classical "exists". For example, the following implication requires classical reasoning because, from a constructive standpoint, knowing that it is not the case that every `x` satisfies `¬ p` is not the same as having a particular `x` that satisfies `p`.

```
open classical

variables (α : Type) (p : α → Prop)

example (h : ¬ ∀ x, ¬ p x) : ∃ x, p x :=
by_contradiction
  (assume h1 : ¬ ∃ x, p x,
    have h2 : ∀ x, ¬ p x, from
      take x,
      assume h3 : p x,
      have h4 : ∃ x, p x, from  ⟨x, h3⟩,
      show false, from h1 h4,
    show false, from h h2)
```

What follows are some common identities involving the existential quantifier. We encourage you to prove as many as you can. We are also leaving it to you to determine which are nonconstructive, and hence require some form of classical reasoning.

```
open classical

variables (α : Type) (p q : α → Prop)
variable a : α
variable r : Prop

example : (∃ x : α, r) → r := sorry
example : r → (∃ x : α, r) := sorry
example : (∃ x, p x ∧ r) ↔ (∃ x, p x) ∧ r := sorry
example : (∃ x, p x ∨ q x) ↔ (∃ x, p x) ∨ (∃ x, q x) := sorry

example : (∀ x, p x) ↔ ¬ (∃ x, ¬ p x) := sorry
example : (∃ x, p x) ↔ ¬ (∀ x, ¬ p x) := sorry
example : (¬ ∃ x, p x) ↔ (∀ x, ¬ p x) := sorry
example : (¬ ∀ x, p x) ↔ (∃ x, ¬ p x) := sorry

example : (∀ x, p x → r) ↔ (∃ x, p x) → r := sorry
example : (∃ x, p x → r) ↔ (∀ x, p x) → r := sorry
example : (∃ x, r → p x) ↔ (r → ∃ x, p x) := sorry
```

Notice that the declaration `variable a : α` amounts to the assumption that there is at least one element of type α. This assumption is needed in the second example, as well as in the last two.

Here are solutions to two of the more difficult ones:

```
example : (∃ x, p x ∨ q x) ↔ (∃ x, p x) ∨ (∃ x, q x) :=
iff.intro
  (assume ⟨a, (h1 : p a ∨ q a)⟩,
    or.elim h1
      (assume hpa : p a, or.inl ⟨a, hpa⟩)
      (assume hqa : q a, or.inr ⟨a, hqa⟩))
  (assume h : (∃ x, p x) ∨ (∃ x, q x),
    or.elim h
      (assume ⟨a, hpa⟩, ⟨a, (or.inl hpa)⟩)
```

```
    (assume ⟨a, hqa⟩, ⟨a, (or.inr hqa)⟩))

example : (∃ x, p x → r) ↔ (∀ x, p x) → r :=
iff.intro
  (assume ⟨b, (hb : p b → r)⟩,
    assume h2 : ∀ x, p x,
    show r, from  hb (h2 b))
  (assume h1 : (∀ x, p x) → r,
    show ∃ x, p x → r, from
      by_cases
        (assume hap : ∀ x, p x, ⟨a, λ h', h1 hap⟩)
        (assume hnap : ¬ ∀ x, p x,
          by_contradiction
            (assume hnex : ¬ ∃ x, p x → r,
              have hap : ∀ x, p x, from
                take x,
                by_contradiction
                  (assume hnp : ¬ p x,
                    have hex : ∃ x, p x → r,
                      from ⟨x, (assume hp, absurd hp hnp)⟩,
                    show false, from hnex hex),
              show false, from hnap hap)))
```

4.5 More on the Proof Language

We have seen that keywords like `assume`, `take`, `have`, and `show` make it possible to write formal proof terms that mirror the structure of informal mathematical proofs. In this section, we discuss some additional features of the proof language that are often convenient.

To start with, we can use anonymous "have" expressions to introduce an auxiliary goal without having to label it. We can refer to the last expression introduced in this way using the keyword `this`:

```
variable f : ℕ → ℕ
premise h : ∀ x : ℕ, f x ≤ f (x + 1)

example : f 0 ≤ f 3 :=
have f 0 ≤ f 1, from h 0,
have f 0 ≤ f 2, from le_trans this (h 1),
show f 0 ≤ f 3, from le_trans this (h 2)
```

Often proofs move from one fact to the next, so this can be effective in eliminating the clutter of lots of labels.

When the goal can be inferred, we can also ask Lean instead to fill in the proof by writing `by assumption`:

```
variable f : ℕ → ℕ
premise h : ∀ x : ℕ, f x ≤ f (x + 1)

example : f 0 ≤ f 3 :=
```

```
have f 0 ≤ f 1, from h 0,
have f 0 ≤ f 2, from le_trans (by assumption) (h 1),
show f 0 ≤ f 3, from le_trans (by assumption) (h 2)
```

This tells Lean to use the `assumption` tactic, which, in turn, proves the goal by finding a suitable hypothesis in the local context. We will learn more about the `assumption` tactic in the next chapter.

We can also ask Lean to fill in the proof by writing `‹p›`, where `p` is the proposition whose proof we want Lean to find in the context.

```
example : f 0 ≥ f 1 → f 1 ≥ f 2 → f 0 = f 2 :=
suppose f 0 ≥ f 1,
suppose f 1 ≥ f 2,
have f 0 ≥ f 2, from le_trans this ‹f 0 ≥ f 1›,
have f 0 ≤ f 2, from le_trans (h 0) (h 1),
show f 0 = f 2, from le_antisymm this ‹f 0 ≥ f 2›
```

You can type these corner quotes using `\f<` and `\f>`, respectively. The letter "f" is for "French," since the unicode symbols can also be used as French quotation marks. In fact, the notation is defined in Lean as follows:

```
notation `‹` p `›` := show p, by assumption
```

This approach is more robust than using `by assumption`, because the type of the assumption that needs to be inferred is given explicitly. It also makes proofs more readable. Here is a more elaborate example:

```
example : f 0 ≤ f 3 :=
have f 0 ≤ f 1, from h 0,
have f 1 ≤ f 2, from h 1,
have f 2 ≤ f 3, from h 2,
show f 0 ≤ f 3, from le_trans ‹f 0 ≤ f 1›
                       (le_trans ‹f 1 ≤ f 2› ‹f 2 ≤ f 3›)
```

Keep in mind that use can use the French quotation marks in this way to refer to *anything* in the context, not just things that were introduced anonymously. It use is also not limited to propositions, though using it for data is somewhat odd:

```
example (n : ℕ) : ℕ := ‹ℕ›
```

The `suppose` keyword acts as an anonymous assume:

```
example : f 0 ≥ f 1 → f 0 = f 1 :=
suppose f 0 ≥ f 1,
show f 0 = f 1, from le_antisymm (h 0) this
```

Notice that there is an asymmetry: you can use **have** with or without a label, but if you do not wish to name the assumption, you must use **suppose** rather than **assume**. The reason is that Lean allows us to write **assume h** to introduce a hypothesis without specifying it, leaving it to the system to infer to relevant assumption. An anonymous **assume** would thus lead to ambiguities when parsing expressions.

As with the anonymous **have**, when you use **suppose** to introduce an assumption, that assumption can also be invoked later in the proof by enclosing it in French quotes.

```
example : f 0 ≥ f 1 → f 1 ≥ f 2 → f 0 = f 2 :=
suppose f 0 ≥ f 1,
suppose f 1 ≥ f 2,
have f 0 ≥ f 2, from le_trans ‹f 2 ≤ f 1› ‹f 1 ≤ f 0›,
have f 0 ≤ f 2, from le_trans (h 0) (h 1),
show f 0 = f 2, from le_antisymm this ‹f 0 ≥ f 2›
```

Notice that `le_antisymm` is the assertion that if `a ≤ b` and `b ≤ a` then `a = b`, and `a ≥ b` is definitionally equal to `b ≤ a`.

5

Tactics

In this chapter, we describe an alternative approach to constructing proofs, using *tactics*. A proof term is a representation of a mathematical proof; tactics are commands, or instructions, that describe how to build such a proof. Informally, we might begin a mathematical proof by saying "to prove the forward direction, unfold the definition, apply the previous lemma, and simplify." Just as these are instructions that tell the reader how to find the relevant proof, tactics are instructions that tell Lean how to construct a proof term. They naturally support an incremental style of writing proofs, in which users decompose a proof and work on goals one step at a time.

We will describe proofs that consist of sequences of tactics as "tactic-style" proofs, to contrast with the ways of writing proof terms we have seen so far, which we will call "term-style" proofs. Each style has its own advantages and disadvantages. For example, tactic-style proofs can be harder to read, because they require the reader to predict or guess the results of each instruction. But they can also be shorter and easier to write. Moreover, tactics offer a gateway to using Lean's automation, since automated procedures are themselves tactics.

5.1 Entering Tactic Mode

Conceptually, stating a theorem or introducing a `have` statement creates a goal, namely, the goal of constructing a term with the expected type. For example, the following creates the goal of constructing a term of type `p ∧ q ∧ p`, in a context with constants `p q : Prop`, `hp : p` and `hq : q`:

```
theorem test (p q : Prop) (hp : p) (hq : q) : p ∧ q ∧ p :=
sorry
```

We can write this goal as follows:

```
p : Prop, q : Prop, hp : p, hq : q ⊢ p ∧ q ∧ p
```

Indeed, if you replace the “sorry” by an underscore in the example above, Lean will report that it is exactly this goal that has been left unsolved.

Ordinarily, we meet such a goal by writing an explicit term. But wherever a term is expected, Lean allows us to insert instead a `begin ... end` block, followed by a sequence of commands, separated by commas. We can prove the theorem above in that way:

```
theorem test (p q : Prop) (hp : p) (hq : q) : p ∧ q ∧ p :=
begin
  apply and.intro,
  exact hp,
  apply and.intro,
  exact hq,
  exact hp
end
```

The `apply` tactic applies an expression, viewed as denoting a function with zero or more arguments. It unifies the conclusion with the expression in the current goal, and creates new goals for the remaining arguments, provided that no later arguments depend on them. In the example above, the command `apply and.intro` yields two subgoals:

```
p : Prop,
q : Prop,
hp : p,
hq : q
⊢ p

⊢ q ∧ p
```

For brevity, Lean only displays the context for the first goal, which is the one addressed by the next tactic command. The first goal is met with the command `exact hp`. The `exact` command is just a variant of `apply` which signals that the expression given should fill the goal exactly. It is good form to use it in a tactic proof, since its failure signals that something has gone wrong; but otherwise `apply` would work just as well.

You can see the resulting proof term with the `print` command:

```
print test
```

You can write a tactic script incrementally. If you run Lean on an incomplete tactic proof bracketed by `begin` and `end`, the system reports all the unsolved goals that remain.

If you are running Lean with its Emacs interface, you can see this information by putting your cursor on the `end` symbol, which should be underlined. In the Emacs interface, there is another extremely useful trick: if you put your cursor on a line of a tactic proof and press “C-c C-g”, Lean will show you the goal that remains at the end of the line.

Tactic commands can take compound expressions, not just single identifiers. The following is a shorter version of the preceding proof:

```
theorem test (p q : Prop) (hp : p) (hq : q) : p ∧ q ∧ p :=
begin
  apply and.intro hp,
  exact and.intro hq hp
end
```

Unsurprisingly, it produces exactly the same proof term.

```
print test
```

Whenever a proof term is expected, instead of using a `begin...end` block, you can write the `by` keyword followed by a single tactic:

```
theorem test (p q : Prop) (hp : p) (hq : q) : p ∧ q ∧ p :=
by exact and.intro hp (and.intro hq hp)
```

In the Lean Emacs mode, if you put your cursor on the “b” in “by” and press “C-c C-g”, Lean shows you the goal that the tactic is supposed to meet.

We will see below that hypotheses can be introduced, reverted, modified, and renamed over the course of a tactic block. As a result, it is impossible for the Lean parser to detect when an identifier that occurs in a tactic block refers to a section variable that should therefore be added to the context. As a result, you need to explicitly tell Lean to include the relevant entities:

```
variables {p q : Prop} (hp : p) (hq : q)

include hp hq

example : p ∧ q ∧ p :=
begin
  apply and.intro hp,
  exact and.intro hq hp
end
```

The `include` command tells Lean to include the indicated variables (as well as any variables they depend on) from that point on, until the end of the section or file. To limit the effect of an `include`, you can use the `omit` command afterwards:

```
include hp hq

example : p ∧ q ∧ p :=
begin
  apply and.intro hp,
  exact and.intro hq hp
end

omit hp hq

-- hp and hq are no longer included by default
```

Alternatively, you can use a section to delimit the scope.

```
section
include hp hq

example : p ∧ q ∧ p :=
begin
  apply and.intro hp,
  exact and.intro hq hp
end
end

-- hp and hq are no longer included by default
```

Another workaround is to find a way to refer to the variable in question before entering a tactic block:

```
example : p ∧ q ∧ p :=
let hp := hp, hq := hq in
begin
  apply and.intro hp,
  exact and.intro hq hp
end
```

Any mention of `hp` or `hq` at all will cause it to be added to the hypotheses in the example.

5.2 Basic Tactics

In addition to `apply` and `exact`, another useful tactic is `intro`, which introduces a hypothesis. What follows is an example of an identity from propositional logic that we proved Section 3.5, now proved using tactics. We adopt the following convention regarding indentation: whenever a tactic introduces one or more additional subgoals, we indent another two spaces, until the additional subgoals are deleted. That rationale behind this convention, and other structuring mechanisms, will be discussed in Section 5.4 below.

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
  apply iff.intro,
    intro h,
    apply or.elim (and.elim_right h),
      intro hq,
      apply or.intro_left,
      apply and.intro,
        exact and.elim_left h,
      exact hq,
    intro hr,
    apply or.intro_right,
    apply and.intro,
    exact and.elim_left h,
    exact hr,
  intro h,
  apply or.elim h,
    intro hpq,
    apply and.intro,
      exact and.elim_left hpq,
    apply or.intro_left,
    exact and.elim_right hpq,
  intro hpr,
  apply and.intro,
    exact and.elim_left hpr,
  apply or.intro_right,
  exact and.elim_right hpr
end
```

The `intro` command can more generally be used to introduce a variable of any type:

```
example (α : Type) : α → α :=
begin
  intro a,
  exact a
end

example (α : Type) : ∀ x : α, x = x :=
begin
  intro x,
  exact eq.refl x
end
```

It has a plural form, `intros`, which takes a list of names.

```
example : ∀ a b c : ℕ, a = b → a = c → c = b :=
begin
  intros a b c h₁ h₂,
  exact eq.trans (eq.symm h₂) h₁
end
```

The `intros` command can also be used without any arguments, in which case, it chooses names and introduces as many variables as it can. We will see an example of this in a moment.

The `assumption` tactic looks through the assumptions in context of the current goal, and if there is one matching the conclusion, it applies it.

```
example (h₁ : x = y) (h₂ : y = z) (h₃ : z = w) : x = w :=
begin
  apply eq.trans h₁,
  apply eq.trans h₂,
  assumption    -- applied h₃
end
```

It will unify metavariables in the conclusion if necessary:

```
example (h₁ : x = y) (h₂ : y = z) (h₃ : z = w) : x = w :=
begin
  apply eq.trans,
  assumption,      -- solves x = ?m_1 with h₁
  apply eq.trans,
  assumption,      -- solves y = ?m_1 with h₂
  assumption       -- solves z = w with h₃
end
```

The following example uses the `intros` command to introduce the three variables and two hypotheses automatically:

```
example : ∀ a b c : ℕ, a = b → a = c → c = b :=
begin
  intros,
  apply eq.trans,
  apply eq.symm,
  assumption,
  assumption
end
```

There are tactics `reflexivity`, `symmetry`, and `transitivity`, which apply the corresponding operation. Using reflexivity, for example, is more general than writing `apply eq.refl`, because it works for any relation that has been tagged with the `refl` attribute.

With that tactic, the previous proof can be written more elegantly as follows:

```
example : ∀ a b c : ℕ, a = b → a = c → c = b :=
begin
  intros,
  transitivity,
  symmetry,
  assumption,
  assumption
end
```

The `repeat` combinator can be used to simplify the last two lines:

```
example : ∀ a b c : ℕ, a = b → a = c → c = b :=
begin
  intros,
  apply eq.trans,
  apply eq.symm,
  repeat { assumption }
end
```

The curly braces introduce a new tactic block; they are equivalent to a using a nested `begin ... end` pair, as discussed in the next section.

There is variant of `apply` called `fapply` that is more aggressive in creating new subgoals for arguments. Here is an example of how it is used:

```
example : ∃ a : ℕ, a = a :=
begin
  fapply exists.intro,
  exact 0,
  apply rfl
end
```

Here, the command `fapply exists.intro` creates two goals. The first is to provide a natural number, `a`, and the second is to prove that `a = a`. Notice that the second goal depends on the first; solving the first goal instantiates a metavariable in the second.

Another tactic that is sometimes useful is the `generalize` tactic, which is, in a sense, an inverse to `intro`.

```
variables x y z : ℕ

example : x = x :=
begin
  generalize x z, -- goal is x : ℕ ⊢ ∀ (z : ℕ), z = z
  intro y,        -- goal is x y : ℕ ⊢ y = y
  reflexivity
end
```

The `generalize` tactic generalizes the conclusion over the variable `x` using a universal quantifier over `z`. We can generalize any term, not just a variable:

```
example : x + y + z = x + y + z :=
begin
  generalize (x + y + z) w, -- goal is x y z : ℕ ⊢ ∀ (w : ℕ), w = w
  intro u,                  -- goal is x y z u : ℕ ⊢ u = u
  reflexivity
end
```

If the expression passed as the first argument to `generalize` is not found in the goal, `generalize` raises an error.

Notice that once we generalize over `x + y + z`, the variables `x y z : ℕ` in the context become irrelevant. The `clear` tactic throws away elements of the context, when it is safe to do so:

```
example : x + y + z = x + y + z :=
begin
  generalize (x + y + z) w, -- goal is x y z : ℕ ⊢ ∀ (w : ℕ), w = w
  clear x y z,
  intro u,                  -- goal is u : ℕ ⊢ u = u
  reflexivity
end
```

Another useful tactic is the `revert` tactic, which moves an element of the context into the goal. When applied to a variable that occurs in the goal, it has the same effect as `generalize` and `clear`:

```
example (x : ℕ) : x = x :=
begin
  revert x,     -- goal is ⊢ ∀ (x : ℕ), x = x
  intro y,      -- goal is y : ℕ ⊢ y = y
  reflexivity
end
```

Moving a hypothesis into the goal yields an implication:

```
example (x y : ℕ) (h : x = y) : y = x :=
begin
  revert h,     -- goal is x y : ℕ ⊢ x = y → y = x
  intro h₁,     -- goal is x y : ℕ, h₁ : x = y ⊢ y = x
  symmetry,
  assumption
end
```

But `revert` is even more clever, in that it will revert not only an element of the context but also all the subsequent elements of the context that depend on it. For example, reverting `x` in the example above brings `h` along with it:

```
example (x y : ℕ) (h : x = y) : y = x :=
begin
  revert x,     -- goal is y : ℕ ⊢ ∀ (x : ℕ), x = y → y = x
  intros,
  symmetry,
  assumption
end
```

You can also revert multiple elements of the context at once:

```
example (x y : ℕ) (h : x = y) : y = x :=
begin
  revert x y,     -- goal is ⊢ ∀ (x y : ℕ), x = y → y = x
  intros,
  symmetry,
  assumption
end
```

5.3 More tactics

Some additional tactics are useful for constructing and destructing propositions and data. For example, when applied to a goal of the form `p ∨ q`, the tactics `left` and `right` are equivalent to `apply or.inl` and `apply or.inr`, respectively. Conversely, the `cases` tactic can be used to decompose a disjunction.

```
example (p q : Prop) : p ∨ q → q ∨ p :=
begin
  intro h,
  cases h with hp hq,
  -- case hp : p
  right, exact hp,
  -- case hq : q
  left, exact hq
end
```

After `cases h` is applied, there are two goals. In the first, the hypothesis `h : p ∨ q` is replaced by `hp : p`, and in the second, it is replaced by `hq : q`. The `cases` can also be used to decompose a conjunction.

```
example (p q : Prop) : p ∧ q → q ∧ p :=
begin
  intro h,
  cases h with hp hq,
  constructor, exact hq, exact hp
end
```

In this case, there is only one goal after the `cases` tactic is applied, with `h : p ∧ q` replaced by a pair of assumptions, `hp : p` and `hq : q`. The constructor applies the unique constructor for conjunction, `and.intro`. With these tactics, an example from the previous section can be rewritten as follows:

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
  apply iff.intro,
  intro h,
   cases h with hp hqr,
   cases hqr with hq hr,
```

```
     left, constructor, repeat { assumption },
     right, constructor, repeat { assumption },
  intro h,
    cases h with hpq hpr,
      cases hpq with hp hq,
        constructor, exact hp, left, exact hq,
      cases hpr with hp hr,
        constructor, exact hp, right, exact hr
end
```

We will see in Chapter 7 that these tactics are quite general. The `cases` tactic can be used to decompose any element of an inductively defined type; `constructor` always applies the first constructor of an inductively defined type, and `left` and `right` can be used with inductively defined types with exactly `two` constructors. For example, we can use `cases` and `constructor` with an existential quantifier:

```
example (p q : ℕ → Prop) : (∃ x, p x) → ∃ x, p x ∨ q x :=
begin
  intro h,
  cases h with x px,
  constructor, left, exact px
end
```

Here, the `constructor` tactic leaves the first component of the existential assertion, the value of `x`, implicit. It is represented by a metavariable, which should be instantiated later on. In the previous example, the proper value of the metavariable is determine by the tactic `exact px`, since `px` has type `p x`. If you want to specify a witness to the existential quantifier explicitly, you can use the `existsi` tactic instead:

```
example (p q : ℕ → Prop) : (∃ x, p x) → ∃ x, p x ∨ q x :=
begin
  intro h,
  cases h with x px,
  existsi x, left, exact px
end
```

These tactics can be used on data just as well as propositions. In the next two example, they are used to define functions which swap the components of the product and sum types:

```
universe variables u v

def swap_pair {α : Type u} {β : Type v} : α × β → β × α :=
begin
  intro p,
  cases p with ha hb,
  constructor, exact hb, exact ha
end

def swap_sum {α : Type u} {β : Type v} : α ⊕ β → β ⊕ α :=
begin
```

```
  intro p,
  cases p with ha hb,
    right, exact ha,
    left, exact hb
end
```

Note that up to the names we have chosen for the variables, the definitions are identical to the proofs of the analogous propositions for conjunction and disjunction. The `cases` tactic will also do a case distinction on a natural number:

```
open nat

example (P : ℕ → Prop) (h₀ : P 0) (h₁ : ∀ n, P (succ n)) (m : ℕ) : P m :=
begin
  cases m with m', exact h₀, exact h₁ m'
end
```

For further discussion, see Chapter 7.

5.4 Structuring Tactic Proofs

Tactics often provide an efficient way of building a proof, but long sequences of instructions can obscure the structure of the argument. In this section, we describe some means that help provide structure to a tactic-style proof, making such proofs more readable and robust.

One thing that is nice about Lean's proof-writing syntax is that it is possible to mix term-style and tactic-style proofs, and pass between the two freely. For example, the tactics `apply` and `exact` expect arbitrary terms, which you can write using `have`, `show`, and so on. Conversely, when writing an arbitrary Lean term, you can always invoke the tactic mode by inserting a `begin...end` block. The following is a somewhat toy example:

```
example (p q r : Prop) : p ∧ (q ∨ r) → (p ∧ q) ∨ (p ∧ r) :=
begin
  intro h,
  exact
    have hp : p, from h^.left,
    have hqr : q ∨ r, from h^.right,
    show (p ∧ q) ∨ (p ∧ r),
    begin
      cases hqr with hq hr,
        exact or.inl ⟨hp, hq⟩,
      exact or.inr ⟨hp, hr⟩
    end
end
```

The following is a more natural example:

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
```

```
  apply iff.intro,
    intro h,
    cases h^.right with hq hr,
      exact
        show (p ∧ q) ∨ (p ∧ r),
          from or.inl ⟨h^.left, hq⟩,
    exact
      show (p ∧ q) ∨ (p ∧ r),
        from or.inr ⟨h^.left, hr⟩,
  intro h,
  cases h with hpq hpr,
    exact
      show p ∧ (q ∨ r),
        from ⟨hpq^.left, or.inl hpq^.right⟩,
  exact show p ∧ (q ∨ r),
    from ⟨hpr^.left, or.inr hpr^.right⟩
end
```

With the `exact` tactic, we use `show` to indicate the goal at that point in the proof. In fact, this idiom is so useful that Lean offers the following abbreviation: in a tactic block, the expression `show p, from t` abbreviates `exact (show p, from t)`. Thus we could have written the previous example more concisely as follows:

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
  apply iff.intro,
    intro h,
    cases h^.right with hq hr,
      show (p ∧ q) ∨ (p ∧ r),
        from or.inl ⟨h^.left, hq⟩,
    show (p ∧ q) ∨ (p ∧ r),
      from or.inr ⟨h^.left, hr⟩,
  intro h,
  cases h with hpq hpr,
    show p ∧ (q ∨ r),
      from ⟨hpq^.left, or.inl hpq^.right⟩,
  show p ∧ (q ∨ r),
    from ⟨hpr^.left, or.inr hpr^.right⟩
end
```

This blurs the distinction between proof-term mode and tactic-mode, and so it is important to use indentation and the structuring mechanisms discussed below to make it clear where a proof term ends. The convention we have used for indentation will be explained momentarily.

In the same way, in a tactic block, Lean interprets `have p, from t₁, t₂` as an abbreviation for `exact (have p, from t₁, t₂)`. Thus the first example in this section could have been written more concisely as follows:

```
example (p q r : Prop) : p ∧ (q ∨ r) → (p ∧ q) ∨ (p ∧ r) :=
begin
  intro h,
```

```
  have hp : p, from h^.left,
  have hqr : q ∨ r, from h^.right,
  show (p ∧ q) ∨ (p ∧ r),
  begin
    cases hqr with hq hr,
      exact or.inl ⟨hp, hq⟩,
    exact or.inr ⟨hp, hr⟩
  end
end
```

You can also nest `begin...end` blocks within other `begin...end` blocks. In a nested block, Lean focuses on the first goal, and generates an error if it has not been fully solved at the end of the block. This can be helpful in indicating the separate proofs of multiple subgoals introduced by a tactic.

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
  apply iff.intro,
  begin
    intro h,
    cases h^.right with hq hr,
    begin
      show (p ∧ q) ∨ (p ∧ r),
        from or.inl ⟨h^.left, hq⟩
    end,
    show (p ∧ q) ∨ (p ∧ r),
      from or.inr ⟨h^.left, hr⟩
  end,
  intro h,
  cases h with hpq hpr,
  begin
    show p ∧ (q ∨ r), from
      ⟨hpq^.left, or.inl hpq^.right⟩
  end,
  show p ∧ (q ∨ r), from
    ⟨hpr^.left, or.inr hpr^.right⟩
end
```

Here, we have introduced a new `begin..end` block whenever a tactic leaves more than one subgoal. You can check (using `C-c C-g` in Emacs mode, for example) that every line in this proof, there is only one goal visible. Notice that you still need to use a comma after a `begin...end` block when there are remaining goals to be discharged.

Within a `begin...end` block, you can abbreviate nested occurrences of `begin` and `end` with curly braces:

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
  apply iff.intro,
  { intro h,
    cases h^.right with hq hr,
    { show (p ∧ q) ∨ (p ∧ r),
        from or.inl ⟨h^.left, hq⟩ },
```

```
    show (p ∧ q) ∨ (p ∧ r),
      from or.inr ⟨h^.left, hr⟩ },
  intro h,
  cases h with hpq hpr,
  { show p ∧ (q ∨ r),
      from ⟨hpq^.left, or.inl hpq^.right⟩ },
  show p ∧ (q ∨ r),
    from ⟨hpr^.left, or.inr hpr^.right⟩
end
```

This helps explain the convention on indentation we have adopted here: every time a tactic leaves more than one subgoal, we separate the remaining subgoals by enclosing them in blocks and indenting, until we are back down to one subgoal. Thus if the application of theorem `foo` to a single goal produces four subgoals, one would expect the proof to look like this:

```
begin
  apply foo,
  { ... proof of first goal ... },
  { ... proof of second goal ... },
  { ... proof of third goal ... },
  proof of final goal
end
```

Another reasonable convention is to enclose *all* the remaining subgoals in indented blocks, including the last one:

```
example (p q r : Prop) : p ∧ (q ∨ r) ↔ (p ∧ q) ∨ (p ∧ r) :=
begin
  apply iff.intro,
  { intro h,
    cases h^.right with hq hr,
    { show (p ∧ q) ∨ (p ∧ r),
        from or.inl ⟨h^.left, hq⟩ },
    { show (p ∧ q) ∨ (p ∧ r),
        from or.inr ⟨h^.left, hr⟩ }},
  { intro h,
    cases h with hpq hpr,
    { show p ∧ (q ∨ r),
        from ⟨hpq^.left, or.inl hpq^.right⟩ },
    { show p ∧ (q ∨ r),
        from ⟨hpr^.left, or.inr hpr^.right⟩ }}
end
```

With this convention, the proof using `foo` described above would look like this:

```
begin
  apply foo,
  { ... proof of first goal ... },
  { ... proof of second goal ... },
  { ... proof of third goal ... },
```

```
  { ... proof of final goal ....}
end
```

Both conventions are reasonable. The second convention has the effect that the text in a long proof gradually creeps to the right. Many theorems in mathematics have side conditions that can be dispelled quickly; using the first convention means that the proofs of these side conditions are indented until we return to the "linear" part of the proof.

You can simulate the effect of the `have` construct without leaving tactic mode using the `assert` tactic.

```
example (p q : Prop) : p ∧ q ↔ q ∧ p :=
begin
  apply iff.intro,
  { intro h,
    assert hp : p, exact h^.left,
    assert hq : q, exact h^.right,
    exact ⟨hq, hp⟩ },
  intro h,
  assert hp : p, exact h^.right,
  assert hq : q, exact h^.left,
  exact ⟨hp, hq⟩
end
```

Here, the first `assert` creates a new subgoal, `p`. After that subgoal is proved, we are left with the original subgoal, with the context augmented by `hp : p`. Tactics are used are to prove both subgoals.

Another option is to use the `note` tactic, which allows you to insert a fact into the context, without having to state the proposition it proves.

```
example (p q : Prop) : p ∧ q ↔ q ∧ p :=
begin
  apply iff.intro,
  { intro h,
    note hp := h^.left,
    note hq := h^.right,
    exact ⟨hq, hp⟩ },
  intro h,
  note hp := h^.right,
  note hq := h^.left,
  exact ⟨hp, hq⟩
end
```

In general, if `e` has type `t`, then `note h := e` adds a hypothesis `h : t` to the context, without giving you access to the contents of `e`. If, instead, you need the contents of `e`, use `pose x := e` instead. This adds `x : t := e` to the context as a `let` definition that can be unfolded when needed. Here is an example:

```
example : ∃ x : ℕ, x + 3 = 8 :=
begin
```

```
  pose x := 5,
  existsi x,
  reflexivity
end
```

Just as the `assert` tactic can be used to simulate the benefits of `have`, the `change` tactic can be used to simulate the benefits of `show`. In the following example, the tactic `change q` affirms that the goal at that point is `q`, and the tactic `change p` affirms that the goal is `p`.

```
example (p q : Prop) : p ∧ q → q ∧ p :=
begin
  intro h,
  split,
  { change q,
    exact h^.right },
  change p,
  exact h^.left
end
```

The name of the `change` tactic is explained by the fact that it can be used to replace a goal by any definitionally equivalent statement.

```
example (a b : ℕ) (h : a = b) : a + 0 = b + 0 :=
begin
  change a = b,
  assumption
end
```

The `change` statement will also work if the expression you give it has metavariables, in which case, it tries to unify the expression with the goal.

```
example (a b c : ℕ) (h₁ : a = b) (h₂ : b = c) : a = c :=
begin
  transitivity,
    change _ = b, assumption,
  assumption
end
```

In this example, after the `transitivity` tactic is applied, there are two goals, `a = ?m_1` and `?m_1 = c`. After the `change`, the two goals have been specialized to `a = b` and `b = c`.

5.5 Rewriting and the Simplifier

The `rewrite` tactic (abbreviated `rw`) and the `simp` tactic were introduced in Section 4.3. In this section, we discuss them in greater detail.

The `rewrite` tactic provide a basic mechanism for applying substitutions to goals and hypotheses, providing a convenient and efficient way of working with equality. The most basic form of the tactic is `rewrite t`, where `t` is a term whose conclusion is an equality. In the following example, we use this basic form to rewrite the goal using a hypothesis.

```
variables (f : ℕ → ℕ) (k : ℕ)

example (h₁ : f 0 = 0) (h₂ : k = 0) : f k = 0 :=
begin
  rw h₂, -- replace k with 0
  rw h₁  -- replace f 0 with 0
end
```

In the example above, the first use of `rw` replaces `k` with `0` in the goal `f k = 0`. Then, the second one replaces `f 0` with `0`. The tactic automatically closes any goal of the form `t = t`.

Multiple rewrites can be combined using the notation `rw [t_1, ..., t_n]`, which is just shorthand for `rewrite t_1, ..., rewrite t_n`. The previous example can be written as follows:

```
variables (f : ℕ → ℕ) (k : ℕ)

example (h₁ : f 0 = 0) (h₂ : k = 0) : f k = 0 :=
begin
  rw [h₂, h₁]
end
```

By default, `rw` uses an equation in the forward direction, matching the left-hand side with an expression, and replacing it with the right-hand side. The notation `-t` can be used to instruct the tactic to use the equality `t` in the reverse direction.

```
variables (f : ℕ → ℕ) (a b : ℕ)

example (h₁ : a = b) (h₂ : f a = 0) : f b = 0 :=
begin
  rw [-h₁, h₂]
end
```

In this example, the term `-h₁` instructs the rewriter to replace `b` with `a`.

Sometimes the left-hand side of an identity can match more than one subterm in the pattern, in which case the `rewrite` tactic chooses the first match it finds when traversing the term. If that is not the one you want, you can use additional arguments to specify the appropriate subterm.

```
example (a b c : ℕ) : a + b + c = a + c + b :=
begin
  rw [add_assoc, add_comm b, -add_assoc]
```

```
end

example (a b c : ℕ) : a + b + c = a + c + b :=
begin
  rw [add_assoc, add_assoc, add_comm b]
end

example (a b c : ℕ) : a + b + c = a + c + b :=
begin
  rw [add_assoc, add_assoc, add_comm _ b]
end
```

In the first example above, the first step rewrites `a + b + c` to `a + (b + c)`. Then next applies commutativity to the term `b + c`; without specifying the argument, the tactic would instead rewrite `a + (b + c)` to `(b + c) + a`. Finally, the last step applies associativity in the reverse direction rewriting `a + (c + b)` to `a + c + b`. The next two examples instead apply associativity to move the parenthesis to the right on both sides, and then switch `b` and `c`. Notice that the last example specifies that the rewrite should take place on the right-hand side by specifying the second argument to `add_comm`.

By default, the `rewrite` tactic affects only the goal. The notation `rw t at h` applies the rewrite `t` at hypothesis `h`.

```
variables (f : ℕ → ℕ) (a : ℕ)

example (h : a + 0 = 0) : f a = f 0 :=
begin
  rw add_zero at h, rw h
end
```

The first step, `rw add_zero at h`, rewrites the hypothesis `a + 0 = 0` to `a = 0`. Then the new hypothesis `a = 0` is used to rewrite the goal to `f 0 = f 0`.

The `rewrite` tactic is not restricted to propositions. In the following example, we use `rw h at t` to rewrite the hypothesis `t : tuple α n` to `v : tuple α 0`.

```
universe variable u

def tuple (α : Type u) (n : ℕ) := { l : list α // list.length l = n }

variables {α : Type u} {n : ℕ}

example (h : n = 0) (t : tuple α n) : tuple α 0 :=
begin
  rw h at t,
  exact t
end
```

Note that the rewrite tactic can carry out generic calculations in any algebraic structure. The following examples involve an arbitrary ring and an arbitrary group, respectively.

```
universe variable uu

example {α : Type uu} [ring α] (a b c : α) : a * 0 + 0 * b + c * 0 + 0 * a = 0 :=
begin
  rw [mul_zero, mul_zero, zero_mul, zero_mul],
  repeat { rw add_zero }
end

example {α : Type uu} [group α] {a b : α} (h : a * b = 1) : a⁻¹ = b :=
by rw [-(mul_one a⁻¹), -h, inv_mul_cancel_left]
```

Using the type class mechanism described in Chapter 10, Lean identifies both abstract and concrete instances of the relevant algebraic structures, and instantiates the relevant facts accordingly.

Whereas `rewrite` is designed as a surgical tool for manipulating a goal, the simplifier offers a powerful form of automation. A number of identities in Lean's library have been tagged with the `[simp]` attribute, and the `simp` tactic uses them to iteratively rewrite subterms in an expression.

```
variables (x y z : ℕ) (p : ℕ → Prop)
premise   (h : p (x * y))

example : (x + 0) * (0 + y * 1 + z * 0) = x * y :=
by simp

include h
example : p ((x + 0) * (0 + y * 1 + z * 0)) :=
begin simp, assumption end
```

In the first example, the left-hand side of the equality in the goal is simplified using the usual identities involving 0 and 1, reducing the goal to `x * y = x * y`. At that point, `simp` applies reflexivity to finish it off. In the second example, `simp` reduces the goal to `p (x * y)`, at which point the assumption `h` finishes it off.

As with `rw`, you can use the keyword `at` to simplify a hypothesis:

```
example (h : p ((x + 0) * (0 + y * 1 + z * 0))) : p (x * y) :=
begin simp at h, assumption end
```

For operations that are commutative and associative, like addition on the natural numbers, the simplifier uses these two facts to rewrite an expression, as well as *left commutativity*. In the case of additition the latter is expressed as follows: `x + (y + z) = y + (x + z)`. It may seem that commutativity and left-commutativity are problematic, in that repeated application of either causes looping. But the simplifier detects identities that permute their arguments, and uses a technique known as *ordered rewriting*. This means that that the system maintains an internal ordering of terms, and only applies the identity if doing so decreases the order. With the three identities mentioned above, this has the effect that all the parentheses in an expression are associated to the right, and the expressions are ordered in a canonical (though somewhat arbitrary) way. Two expressions that are equivalent up to associativity and commutativity are then rewritten to the same canonical form.

```
variables (x y z w : ℕ) (p : ℕ → Prop)

example : x * y + z * w  * x = x * w * z + y * x :=
by simp

example (h : p (x * y + z * w  * x)) : p (x * w * z + y * x) :=
begin simp, simp at h, assumption end
```

As with the rewriter, the simplifier behaves appropriately in algebraic structures:

```
variables {α : Type} [comm_ring α]

example (x y z : α) : (x - x) * y + z = z :=
begin simp end

example (x y z w : α) : x * y + z * w  * x = x * w * z + y * x :=
by simp
```

Also as with the `rewrite` tactic, you can pass additional arguments to `simp`. These can either be names of theorems or expressions. The `simp` tactic does not recognize the `-t` syntax, so to use an identity in the other direction you need to use `eq.symm` explicitly. In any case, the additional rules are added to the collection of identities that are used to simplify a term.

```
def f (m n : ℕ) : ℕ := m + n + m

theorem f.def (m n : ℕ) : f m n = m + n + m := rfl

example {m n : ℕ} (h : n = 1) (h' : 0 = m) : (f m n) * m = m :=
by simp [h, h'^.symm, f.def]
```

If we add the attribute `[simp]` to the theorem `f.def`, we do not need to include it.

```
def f (m n : ℕ) : ℕ := m + n + m

@[simp]
theorem f.def (m n : ℕ) : f m n = m + n + m := rfl

example {m n : ℕ} (h : n = 1) (h' : 0 = m) : (f m n) * m = m :=
by simp [h, h'^.symm]
```

6

Interacting with Lean

You are now familiar with the fundamentals of dependent type theory, both as a language for defining mathematical objects and a language for constructing proofs. The one thing you are missing is a mechanism for defining new data types. We will fill this gap in the next chapter, which introduces the notion of an *inductive data type*. But first, in this chapter, we take a break from the mechanics of type theory to explore some pragmatic aspects of interacting with Lean.

Not all of the information found here will be useful to you right away. We recommend skimming this section to get a sense of Lean's features, and then returning to it as necessary.

6.1 Importing Files

The goal of Lean's front end is to interpret user input, construct formal expressions, and check that they are well formed and type correct. Lean also supports the use of various editors, which provide continuous checking and feedback. More information can be found on the Lean documentation pages.

The definitions and theorems in Lean's standard library are spread across multiple files. Users may also wish to make use of additional libraries, or develop their own projects across multiple files. When Lean starts, it automatically imports the contents of the library `init` folder, which includes a number of fundamental definitions and constructions. As a result, most of the examples we present here work "out of the box."

If you want to use additional files, however, they need to be imported manually, via an `import` statement at the beginning of a file. The command

```
import foo bar.baz.blah
```

imports the files `foo.lean` and `bar/baz/blah.lean`, where the descriptions are interpreted relative to the Lean *search path*. Information as to how the search path is determined can be found on the documentation pages. By default, it includes the standard library directory, and (in some contexts) the root of the user's local project. One can also specify imports relative to the current directory; for example,

```
import .foo ..bar.baz
```

tells Lean to import `foo.lean` from the current directory and `bar/baz.lean` relative to the parent of the current directory.

Importing is transitive. In other words, if you import `foo` and `foo` imports `bar`, then you also have access to the contents of `bar`, and do not need to import it explicitly.

6.2 More on Sections

Lean provides various sectioning mechanisms to help structure a theory. We saw in Section 2.6 that the `section` command makes it possible not only to group together elements of a theory that go together, but also to declare variables that are inserted as arguments to theorems and definitions, as necessary. Remember that the point of the variable command is to declare variables for use in theorems, as in the following example:

```
section
  variables x y : ℕ

  def double := x + x

  check double y
  check double (2 * x)

  theorem t1 : double (x + y) = double x + double y :=
  by simp [double]

  check t1 y
  check t1 (2 * x)

  theorem t2 : double (x * y) = double x * y :=
  by simp [double, mul_add]
end
```

The definition of `double` does not have to declare `x` as an argument; Lean detects the dependence and inserts it automatically. Similarly, Lean detects the occurrence of `x` in `t1` and `t2`, and inserts it automatically there, too.

Note that double does *not* have `y` as argument. Variables are only included in declarations where they are actually mentioned. More precisely, they must be mentioned outside of a tactic block; because variables can appear and can be renamed dynamically in a tactic proof, there is no reliable way of determining when a name used in a tactic proof refers to an element of the context in which the theorem is parsed, and Lean does not try to guess. You can manually ask Lean to include a variable in every definition in a section with the `include` command.

```
section
  variables (x y z : ℕ)
  variables (h₁ : x = y) (h₂ : y = z)

  include h₁ h₂
  theorem foo : x = z :=
  begin
    rw [h₁, h₂]
  end
  omit h₁ h₂

  theorem bar : x = z :=
  eq.trans h₁ h₂

  theorem baz : x = x := rfl

  check @foo
  check @bar
  check @baz
end
```

The `omit` command simply undoes the effect of the `include`; it does not prevent the arguments from being included automatically in subsequent theorems that mention them. The scope of the `include` statement can also be delimited by enclosing it in a section.

```
  section include_hs
  include h₁ h₂

  theorem foo : x = z :=
  begin
    rw [h₁, h₂]
  end

  end include_hs
```

The include command is often useful with structures that are not mentioned explicitly but meant to be inferred by type class inference, as described in Chapter 10.

It is often the case that we want to declare section variables as explicit variables but later make them implicit, or vice-versa. One can do this with a `variables` command that mentions these variables with the desired brackets, without repeating the type again. Once again, sections can be used to delimit scope. In the example below, the variables `x`, `y`, and `z` are marked implicit in `foo` but explicit in `bar`, while `x` is (somewhat perversely) marked as implicit in `baz`.

```
section
  variables (x y z : ℕ)
  variables (h₁ : x = y) (h₂ : y = z)

  section
    variables {x y z}
    include h₁ h₂
    theorem foo : x = z :=
    begin
      rw [h₁, h₂]
    end
  end

  theorem bar : x = z :=
  eq.trans h₁ h₂

  variable {x}
  theorem baz : x = x := rfl

  check @foo
  check @bar
  check @baz
end
```

Using these subsequent `variables` commands does not change the order in which variables are inserted. It only changes the explicit / implicit annotations.

In fact, Lean has two ways of introducing local elements into the sections, namely, as `variables` or as `parameters`. In the initial example in this section, the variable `x` is generalized immediately, so that even within the section `double` is a function of `x`, and `t1` and `t2` depend explicitly on `x`. This is what makes it possible to apply `double` and `t1` to other expressions, like `y` and `2 * x`. It corresponds to the ordinary mathematical locution "in this section, let `x` and `y` range over the natural numbers." Whenever `x` and `y` occur, we assume they denote natural numbers, but we do not assume they refer to the same natural number from theorem to theorem.

Sometimes, however, we wish to *fix* a value in a section. For example, following ordinary mathematical vernacular, we might say "in this section, we fix a type, α, and a binary relation `r` on α." The notion of a `parameter` captures this usage:

```
section
  parameters {α : Type} (r : α → α → Type)
  parameter  transr : ∀ {x y z}, r x y → r y z → r x z

  variables {a b c d e : α}

  theorem t1 (h₁ : r a b) (h₂ : r b c) (h₃ : r c d) : r a d :=
  transr (transr h₁ h₂) h₃

  theorem t2 (h₁ : r a b) (h₂ : r b c) (h₃ : r c d) (h₄ : r d e) :
    r a e :=
  transr h₁ (t1 h₂ h₃ h₄)
```

```
  check t1
  check t2
end

check t1
check t2
```

As with variables, the parameters α, `r`, and `transR` are inserted as arguments to definitions and theorems as needed. But there is a difference: within the section, `t1` is an abbreviation for `@t1` α `r transr`, which is to say, these arguments are held fixed until the section is closed. On the plus side, this means that you do not have to specify the explicit arguments `r` and `transr` when you write `t1` h_2 h_3 h_4, in contrast to the previous example. But it also means that you cannot specify other arguments in their place. In this example, making `r` a parameter is appropriate if `r` is the only binary relation you want to reason about in the section. In that case, it would make sense to introduce temporary infix notation like $\preceq$ for `r`, and we will see in Section 6.6 how to do that. On the other hand, if you want to apply your theorems to arbitrary binary relations within the section, you should make `r` a variable.

6.3 More on Namespaces

In Lean, identifiers are given by hierarchical *names* like `foo.bar.baz`. We saw in Section 2.7 that Lean provides mechanisms for working with hierarchical names. The command `namespace foo` causes `foo` to be prepended to the name of each definition and theorem until `end foo` is encountered. The command `open foo` then creates temporary *aliases* to definitions and theorems that begin with prefix `foo`.

```
namespace foo
def bar : ℕ := 1
end foo

open foo

check bar
check foo.bar
```

It is not important the the definition of `foo.bar` was the result of a `namespace` command:

```
def foo.bar : ℕ := 1

open foo

check bar
check foo.bar
```

Although the names of theorems and definitions have to be unique, the aliases that identify them do not. For example, the standard library defines a theorem `add_sub_cancel`, which asserts `a + b - b = a` in any additive group. The corresponding theorem on the natural numbers is named `nat.add_sub_cancel`; it is not a special case of `add_sub_cancel`, because the natural numbers do not form a group. When we open the `nat` namespace, the expression `add_sub_cancel` is overloaded, and can refer to either one. Lean tries to use type information to disambiguate the meaning in context, but you can always disambiguate by giving the full name. To that end, the string `_root_` is an explicit description of the empty prefix.

```
check add_sub_cancel
check nat.add_sub_cancel
check _root_.add_sub_cancel
```

We can prevent the shorter alias from being created by using the `protected` keyword:

```
namespace foo
protected def bar : ℕ := 1
end foo

open foo

-- check bar -- error
check foo.bar
```

This is often used for names like `nat.rec_on` and `nat.induction_on`, to prevent overloading of common names.

The `open` command admits variations. The command

```
open nat (succ add sub)
```

creates aliases for only the identifiers listed. The command

```
open nat (hiding succ add sub)
```

creates aliases for everything in the `nat` namespace *except* the identifiers lists. The command

```
open nat (renaming induction_on → induction_on) (renaming add → plus) (hiding succ sub)
```

creates aliases for everything in the `nat` namespace except `succ` and `sub`, renaming `nat.add` to `plus`, and renaming the protected definition `nat.induction_on` to `induction_on`.

It is sometimes useful to `export` aliases from one namespace to another, or to the top level. The command

```
export nat (succ add sub)
```

creates aliases for `succ`, `add`, and `sub` in the current namespace, so that whenever the namespace is open, these aliases are available. If this command is used outside a namespace, the aliases are exported to the top level. The `export` command admits all the variations described above.

6.4 Attributes

The main function of Lean is to translate user input to formal expressions that are checked by the kernel for correctness and then stored in the environment for later use. But some commands have other effects on the environment, either assigning attributes to objects in the environment, defining notation, or declaring instances of type classes, as described in Chapter 10. Most of these commands have global effects, which is to say, that they remain in effect not only in the current file, but also in any file that imports it. However, such commands can often be prefixed with the `local` modifier, which indicates that they only have effect until the current `section` or `namespace` is closed, or until the end of the current file.

In the last Chapter, we saw that theorems can be annotated with the `[simp]` attribute, which makes them available for use by the simplifier. The following example defines divisibility on the natural numbers, uses it to make the natural numbers an instance of a type for which the divisibility notation `\|` is available (the `instance` command will be explained in Chapter 10), and assign the `[simp]` attribute.

```
def nat.dvd (m n : ℕ) : Prop := ∃ k, n = m * k

instance : has_dvd nat := ⟨nat.dvd⟩

attribute [simp]
theorem nat.dvd_refl (n : ℕ) : n ∣ n :=
⟨1, by simp⟩

example : 5 ∣ 5 := by simp
```

Here the simplifier proves `5 ∣ 5` by rewriting it to `true`. Lean allows the alternative annotation `@[simp]` before a theorem to assign the attribute:

```
@[simp]
theorem nat.dvd_refl (n : ℕ) : n ∣ n :=
⟨1, by simp⟩
```

One can also assign the attribute any time after the definition takes place:

```
theorem nat.dvd_refl (n : ℕ) : n ∣ n :=
⟨1, by simp⟩

attribute [simp] nat.dvd_refl
```

In all these cases, the attribute remains in effect in any file that imports the one in which the declaration occurs. But adding the `local` modifier restricts the scope:

```
section
local attribute [simp]
theorem nat.dvd_refl (n : ℕ) : n ∣ n :=
⟨1, by simp⟩

example : 5 ∣ 5 := by simp
end

-- error:
-- example : 5 ∣ 5 := by simp
```

In fact, the `instance` command works by automatically generating a theorem name and assigning an `[instance]` attribute to it. The declaration can also be made local:

```
section
def has_dvd_nat : has_dvd nat := ⟨nat.dvd⟩

local attribute [instance] has_dvd_nat

local attribute [simp]
theorem nat.dvd_refl (n : ℕ) : n ∣ n :=
⟨1, by simp⟩

example : 5 ∣ 5 := by simp
end

-- error:
-- check 5 ∣ 5
```

For yet another example, the `reflexivity` tactic makes use of objects in the environment that have been tagged with the `[refl]` attribute:

```
@[simp,refl]
theorem nat.dvd_refl (n : ℕ) : n ∣ n :=
⟨1, by simp⟩

example : 5 ∣ 5 :=
by reflexivity
```

The scope of the `[refl]` attribute can similarly be restricted using the `local` modifier, as above.

In Section 6.6 below, we will discuss Lean's mechanisms for defining notation, and see that they also support the `local` modifier. However, in Section 6.8, we will discuss Lean's mechanisms for setting options, which does *not* follow this pattern: options can *only* be set locally, which is to say, their scope is always restricted to the current section or current file.

6.5 More on Implicit Arguments

In Section 2.9, we saw that if Lean displays the type of a term `t` as `Π {x : α}, β x`, then the curly brackets indicate that `x` has been marked as an *implicit argument* to `t`. This means that whenever you write `t`, a placeholder, or "hole," is inserted, so that `t` is replaced by `@t _`. If you don't want that to happen, you have to write `@t` instead.

Notice that implicit arugments are inserted eagerly. Suppose we define a function `f (x : ℕ) {y : ℕ} (z : ℕ)` with the arguments shown. Then, when we write the expression `f 7` without further arguments, it parsed as `f 7 _`. Lean offers a weaker annotation, `{{y : ℕ}}`, which specifies that a placeholder should only be added *before* a subsequent explicit argument. This annotation can also be written using as `⦃y : ℕ⦄`, where the unicode brackets are entered as `\{{` and `\}}`, respectively. With this annotation, the expression `f 7` would be parsed as is, whereas `f 7 3` would be parsed as `f 7 _ 3`, just as it would be with the strong annotation.

To illustrate the difference, consider the following example, which shows that a reflexive euclidean relation is both symmetric and transitive.

```
variables {α : Type} (r : α → α → Prop)

definition reflexive  : Prop := ∀ (a : α), r a a
definition symmetric  : Prop := ∀ {a b : α}, r a b → r b a
definition transitive : Prop := ∀ {a b c : α}, r a b → r b c → r a c
definition euclidean  : Prop := ∀ {a b c : α}, r a b → r a c → r b c

variable {r}

theorem th1 (reflr : reflexive r) (euclr : euclidean r) : symmetric r :=
take a b : α, suppose r a b,
show r b a, from euclr this (reflr _)

theorem th2 (symmr : symmetric r) (euclr : euclidean r) : transitive r :=
take (a b c : α), assume (rab : r a b) (rbc : r b c),
euclr (symmr rab) rbc

-- ERROR:
/-
theorem th3 (reflr : reflexive r) (euclr : euclidean r) : transitive r :=
th2 (th1 reflr euclr) euclr
-/

theorem th3 (reflr : reflexive r) (euclr : euclidean r) : transitive r :=
@th2 _ _ (@th1 _ _ reflr @euclr) @euclr
```

The results are broken down into small steps: `th1` shows that a relation that is reflexive and euclidean is symmetric, and `th2` shows that a relation that is symmetric and euclidean is transitive. Then `th3` combines the two results. But notice that we have to manually disable the implicit arguments in `th1`, `th2`, and `euclr`, because otherwise too many implicit arguments are inserted. The problem goes away if we use weak implicit arguments:

```
variables {α : Type} (r : α → α → Prop)

definition reflexive  : Prop := ∀ (a : α), r a a
definition symmetric  : Prop := ∀ ⦃a b : α⦄, r a b → r b a
definition transitive : Prop := ∀ ⦃a b c : α⦄, r a b → r b c → r a c
definition euclidean  : Prop := ∀ ⦃a b c : α⦄, r a b → r a c → r b c

variable {r}

theorem th1 (reflr : reflexive r) (euclr : euclidean r) : symmetric r :=
take a b : α, suppose r a b,
show r b a, from euclr this (reflr _)

theorem th2 (symmr : symmetric r) (euclr : euclidean r) : transitive r :=
take (a b c : α), assume (rab : r a b) (rbc : r b c),
euclr (symmr rab) rbc

theorem th3 (reflr : reflexive r) (euclr : euclidean r) : transitive r :=
th2 (th1 reflr euclr) euclr
```

There is a third kind of implicit argument that is denoted with square brackets, `[` and `]`. These are used for type classes, as explained in Chapter 10.

6.6 Notation

Lean's parser is an instance of a Pratt parser, a non-backtracking parser that is fast and flexible. You can read about Pratt parsers in a number of places online, such as here:

> http://en.wikipedia.org/wiki/Pratt_parser http://eli.thegreenplace.net/2010/01/02/top-down-operator-precedence-parsing

Identifiers can include any alphanumeric characters, including Greek characters (other than Π , Σ , and λ , which, as we have seen, have a special meaning in the dependent type theory). They can also include subscripts, which can be entered by typing `\_` followed by the desired subscripted character.

Lean's parser is extensible, which is to say, we can define new notation.

```
notation `[` a `**` b `]` := a * b + 1

def mul_square (a b : ℕ) := a * a * b * b
```

```
infix `<*>`:50 := mul_square

eval [2 ** 3]
eval 2 <*> 3
```

In this example, the `notation` command defines a complex binary notation for multiplying and adding one. The `infix` command declares a new infix operator, with precedence 50, which associates to the left. (More precisely, the token is given left-binding power 50.) The command `infixr` defines notation which associates to the right, instead.

If you declare these notations in a namespace, the notation is only available when the namespace is open. You can declare temporary notation using the keyword `local`, in which case the notation is available in the current file, and moreover, within the scope of the current `namespace` or `section`, if you are in one.

```
local notation `[` a `**` b `]` := a * b + 1
local infix `<*>`:50 := λ a b : ℕ, a * a * b * b
```

Lean's stanard library declares the left-binding powers of a number of common symbols.

https://github.com/leanprover/lean/blob/master/library/init/core.lean

You are welcome to overload these symbols for your own use, but you cannot change their binding power.

You can direct the pretty-printer to suppress notation with the command `set_option pp.notation false`. You can also declare notation to be used for input purposes only with the `[parsing_only]` attribute:

```
notation [parsing_only] `[` a `**` b `]` := a * b + 1

variables a b : ℕ
check [a ** b]
```

The output of the `check` command displays the expression as `a * b + 1`. Lean also provides mechanisms for iterated notation, such as `[a, b, c, d, e]` to denote a list with the indicated elements. See the discussion of `list` in the next chapter for an example.

The possibility of declaring parameters in a section also makes it possible to define local notation that depends on those parameters. In the example below, as long as the parameter `m` is fixed, we can write `a ≡ b` for equivalence modulo `m`. As soon as the section is closed, however, the dependence on `m` becomes explicit, and the notation `a ≡ b` is no longer valid.

```
section mod_m
  parameter (m : ℤ)
  variables (a b c : ℤ)
```

```
  definition mod_equiv := (m ∣ b - a)

  local infix ≡ := mod_equiv

  theorem mod_refl : a ≡ a :=
  show m ∣ a - a, by simp

  theorem mod_symm (h : a ≡ b) : b ≡ a :=
  by cases h with c hc; apply dvd_intro (-c); simp [eq.symm hc]

  theorem mod_trans (h₁ : a ≡ b) (h₂ : b ≡ c) : a ≡ c :=
  begin
    cases h₁ with d hd, cases h₂ with e he, apply dvd_intro (d + e),
    simp [mul_add, eq.symm hd, eq.symm he]
  end
end mod_m

check (mod_refl : ∀ (m a : ℤ), mod_equiv m a a)

check (mod_symm : ∀ (m a b : ℤ), mod_equiv m a b → mod_equiv m b a)

check (mod_trans : ∀ (m a b c : ℤ),
                    mod_equiv m a b → mod_equiv m b c → mod_equiv m a c)
```

6.7 Coercions

In Lean, the type of natural numbers, `nat`, is different from the type of integers, `int`. But there is a function `int.of_nat` that embeds the natural numbers in the integers, meaning that we can view any natural numbers as an integer, when needed. Lean has mechanisms to detect and insert *coercions* of this sort.

```
variables m n : ℕ
variables i j : ℤ

check i + m      -- i + ↑m : ℤ
check i + m + j  -- i + ↑m + j : ℤ
check i + m + n  -- i + ↑m + ↑n : ℤ
```

Notice that the output of the `check` command shows that a coercion has been inserted by printing an arrow. The latter is notation for the function `coe`; you can type the unicode arrow with `\u` or use the `coe` instead. In fact, when the order of arguments is different, you have to insert the coercion manually, because Lean does not recognize the need for a coercion until it has already parsed the earlier arguments.

```
check ↑m + i        -- ↑m + i : ℤ
check ↑(m + n) + i  -- ↑(m + n) + i : ℤ
check ↑m + ↑n + i   -- ↑m + ↑n + i : ℤ
```

In fact, Lean allows various kinds of coercions using type classes; for details, see Section 10.6.

6.8 Displaying Information

There are a number of ways in which you can query Lean for information about its current state and the objects and theorems that are available in the current context. You have already seen two of the most common ones, `check` and `eval`. Remember that `check` is often used in conjunction with the `@` operator, which makes all of the arguments to a theorem or definition explicit. In addition, you can use the `print` command to get information about any identifier. If the identifier denotes a definition or theorem, Lean prints the type of the symbol, and its definition. If it is a constant or an axiom, Lean indicates that fact, and shows the type.

```
-- examples with equality
check eq
check @eq
check eq.symm
check @eq.symm

print eq.symm

-- examples with and
check and
check and.intro
check @and.intro

-- examples with addition
check add
check @add
eval add 3 2
print add

-- a user-defined function
def foo {α : Type} (x : α) : α := x

check foo
check @foo
eval foo
eval (foo @nat.zero)
print foo
```

There are other useful `print` commands:

```
print definition             : display definition
print inductive              : display an inductive type and its constructors
print notation               : display all notation
print notation <tokens>      : display notation using any of the tokens
print axioms                 : display assumed axioms
print options                : display options set by user
```

```
print prefix <namespace>     : display all declarations in the namespace
print classes                : display all classes
print instances <class name> : display all instances of the given class
print fields <structure>     : display all "fields" of a structure
```

We will discuss inductive types, structures, classes, instances in the next four chapters. Here are examples of how these commands are used:

```
print notation
print notation + * -
print axioms
print options
print prefix nat
print prefix nat.le
print classes
print instances ring
print fields ring
```

The behavior of the generic print command is determined by its argument, so that the following pairs of commands all do the same thing.

```
print add
print definition add

print +
print notation +

print nat
print inductive nat

print group
print inductive group
```

Moreover, both `print group` and `print inductive group` recognize that a group is a structure (see Chapter 9), and so print the fields as well.

6.9 Setting Options

Lean maintains a number of internal variables that can be set by users to control its behavior. The syntax for doing so is as follows:

```
set_option <name> <value>
```

One very useful family of options controls the way Lean's *pretty- printer* displays terms. The following options take an input of true or false:

```
pp.implicit  : display implicit arguments
pp.universes : display hidden universe parameters
pp.coercions : show coercions
pp.notation  : display output using defined notations
pp.beta      : beta reduce terms before displaying them
```

As an example, the following settings yield much longer output:

```
set_option pp.implicit true
set_option pp.universes true
set_option pp.notation false
set_option pp.numerals false

check 2 + 2 = 4
eval (λ x, x + 2) = (λ x, x + 3)
check (λ x, x + 1) 1
```

The command `set_option pp.all true` carries out these settings all at once, whereas `set_option pp.all false` reverts to the previous values. Pretty printing additional information is often very useful when you are debugging a proof, or trying to understand a cryptic error message. Too much information can be overwhelming, though, and Lean's defaults are generally sufficient for ordinary interactions.

By default, the pretty-printer does not reduce applied lambda-expressions, but this is sometimes useful. The `pp.beta` option controls this feature.

```
set_option pp.beta true
check (λ x, x + 1) 1
```

6.10 Elaboration Hints

When you ask Lean to process an expression like `λ x y z, f (x + y) z`, you are leaving information implicit. For example, the types of `x`, `y`, and `z` have to be inferred from the context, the notation `+` may be overloaded, and there may be implicit arguments to `f` that need to be filled in as well. Moreover, we will see in Chapter 10 that some implicit arguments are synthesized by a process known as *type class resolution*. And we have also already seen in the last chapter that some parts of an expression can be constructed by the tactic framework.

Inferring some implicit arguments is straightforward. For example, suppose a function `f` has type `Π {α : Type}, α → α → α` and Lean is trying to parse the expression `f n`, where `n` can be inferred to have type `nat`. Then it is clear that the implicit argument α has to be `nat`. However, some inference problems are *higher order*. For example, the substitution operation for equality, `eq.subst`, has the following type:

```
eq.subst : ∀ {α : Sort u} {P : α → Prop} {a b : α}, a = b → P a → P b
```

Now suppose we are given `a b : ℕ` and h_1 `: a = b` and h_2 `: a * b > a`. Then, in the expression `eq.subst` h_1 h_2, `P` could be any of the following:

- `λ x, x * b > x`
- `λ x, x * b > a`
- `λ x, a * b > x`
- `λ x, a * b > a`

In other words, our intent may be to replace either the first or second `a` in h_2, or both, or neither. Similar ambiguities arise in inferring induction predicates, or inferring function arguments. Even second-order unification is known to be undecidable. Lean therefore relies on heuristics to fill in such arguments, and when it fails to guess the right ones, they need to be provided explicity.

To make matters worse, sometimes definitions need to be unfolded, and sometimes expressions need to be reduced according to the computational rules of the underlying logical framework. Once again, Lean has to rely on heuristics to determine what to unfold or reduce, and when.

There are attributes, however, that can be used to provide hints to the elaborator. One class of attributes determines how eagerly definitions are unfolded: constants can be marked with the attribute `[reducible]`, `[semireducible]`, or `[irreducible]`. Definitions are marked `[semireducible]` by default. A definition with the `[reducible]` attribute is unfolded eagerly; if you think of a definition are serving as an abbreviation, this attribute would be appropriate. The elaborator avoids unfolding definitions with the `[irreducible]` attribute. Theorems are marked `[irreducible]` by default, because typically proofs are not relevant to the elaboration process.

It is worth emphasizing that these attributes are only hints to the elaborator. When checking an elaborated term for correctness, Lean's kernel will unfold whatever definitions it needs to unfold. As with other attributes, the ones above can be assigned with the `local` modifier, so that they are in effect only in the current section or file.

Lean also has a family of attributes that control the elaboration strategy. A definition or theorem can be marked `[elab_with_expected_type]`, `[elab_simple]`. or `[elab_as_eliminator]`. When applied to a definition `f`, these bear on elaboration of an expression `f a b c ...` in which `f` is applied to arguments. With the default attribute, `[elab_with_expected_type]`, the arguments `a`, `b`, `c`, … are elaborating using information about their expected type, inferred from `f` and the previous arguments. In contrast, with `[elab_simple]`, the arguments are elaborated from left to right without propagating information about their types. The last attribute, `[elab_as_eliminator]`, is commonly used for eliminators like recursors, induction principles, and eq.subst. It uses a separate heuristic to infer higher-order parameters. We will consider such operations in more detail in the next chapter.

Once again, these attributes can assigned and reassigned after an object is defined, and you can use the local modifier to limit their scope. Moreover, using the @ simple in front of an identifier in an expression instructs the elaborator to use the [elab_simple] strategy; the idea is that, when you provide the tricky parameters explicitly, you want the elaborator to weigh that information heavily. In fact, Lean offers an alterantive annotation, @@, which leaves parameters before the first higher-order parameter explicit. For example, @@eq.subst leaves the type of the equation implicit, but makes the context of the substitution explicit.

6.11 Using the Library

To use Lean effectively you will inevitably need to make use of definitions and theorems in the library. Recall that the import command at the beginning of a file imports previously compiled results from other files, and that importing is transitive; if you import foo and foo imports bar, then the definitions and theorems from bar are available to you as well. But the act of opening a namespace, which provides shorter names, does not carry over. In each file, you need to open the namespaces you wish to use.

In general, it is important for you to be familiar with the library and its contents, so you know what theorems, definitions, notations, and resources are available to you. Below we will see that Lean's editor modes can also help you find things you need, but studying the contents of the library directly is often unavoidable. Lean's standard library can be found online, on github:

https://github.com/leanprover/lean/tree/master/library

You can see the contents of the directories and files using github's browser interface. If you have installed Lean on your own computer, you can find the library in the lean folder, and explore it with your file manager. Comment headers at the top of each file provide additional information.

Lean's library developers follow general naming guidelines to make it easier to guess the name of a theorem you need, or to find it using tab completion in editors with a Lean mode that supports this, which is discussed in the next section. Identifiers are generally snake_case, which is to say, they are composed of words written in lower case separated by underscores. For the most part, we rely on descriptive names. Often the name of theorem simply describes the conclusion:

```
open nat

check succ_ne_zero
```

```
check @mul_zero
check @mul_one
check @sub_add_eq_add_sub
check @le_iff_lt_or_eq
```

If only a prefix of the description is enough to convey the meaning, the name may be made even shorter:

```
check @neg_neg
check pred_succ
```

Sometimes, to disambiguate the name of theorem or better convey the intended reference, it is necessary to describe some of the hypotheses. The word "of" is used to separate these hypotheses:

```
check @nat.lt_of_succ_le
check @lt_of_not_ge
check @lt_of_le_of_ne
check @add_lt_add_of_lt_of_le
```

Sometimes the word "left" or "right" is helpful to describe variants of a theorem.

```
check @add_le_add_left
check @add_le_add_right
```

We can also use the word "self" to indicate a repeated argument:

```
check mul_inv_self
check neg_add_self
```

Remember that identifiers in Lean can be organized into hierarchical namespaces. For example, the theorem named `lt_of_succ_le` in the namespace `nat` has full name `nat.lt_of_succ_le`, but the shorter name is made available by the command `open nat`. We will see in Chapter 7 and Chapter 9 that defining structures and inductive data types in Lean generates associated operations, and these are stored in a namespace with the same name as the type under definition. For example, the product type comes with the following opens:

```
check @prod.mk
check @prod.fst
check @prod.snd
check @prod.rec
```

The first is used to construct a pair, whereas the next two, `prod.fst` and `prod.snd`, project the two elements. The last, `prod.rec`, provides another mechanism for defining functions on a product in terms of a function on the two components. Names like `prod.rec` are *protected*, which means that one has to use the full name even when the `prod` namespace is open.

With the propositions as types correspondence, logical connectives are also instances of inductive types, and so we tend to use dot notation for them as well:

```
check @and.intro
check @and.elim
check @and.left
check @and.right
check @or.inl
check @or.inr
check @or.elim
check @exists.intro
check @exists.elim
check @eq.refl
check @eq.subst
```

7

Inductive Types

We have seen that Lean's formal foundation includes basic types, `Prop, Type 1, Type 2, ...`, and allows for the formation of dependent function types, Π `x :` α. β. In the examples, we have also made use of additional types like `bool`, `nat`, and `int`, and type constructors, like `list`, and product, $\times$. In fact, in Lean's library, every concrete type other than the universes and every type constructor other than Pi is an instance of a general family of type constructions known as *inductive types*. It is remarkable that it is possible to construct a substantial edifice of mathematics based on nothing more than the type universes, Pi types, and inductive types; everything else follows from those.

Intuitively, an inductive type is built up from a specified list of constructors. In Lean, the syntax for specifying such a type is as follows:

```
inductive foo : Type
| constructor₁ : ... → foo
| constructor₂ : ... → foo
...
| constructorₙ : ... → foo
```

The intuition is that each constructor specifies a way of building new objects of `foo`, possibly from previously constructed values. The type `foo` consists of nothing more than the objects that are constructed in this way. The first character `|` in an inductive declaration is optional. We can also separate constructors using a comma instead of `|`.

We will see below that the arguments to the constructors can include objects of type `foo`, subject to a certain "positivity" constraint, which guarantees that elements of `foo` are built from the bottom up. Roughly speaking, each `...` can be any Pi type constructed from `foo` and previously defined types, in which `foo` appears, if at all, only as the "target" of the Pi type. For more details, see [2].

We will provide a number of examples of inductive types. We will also consider slight generalizations of the scheme above, to mutually defined inductive types, and so-called *inductive families*.

As with the logical connectives, every inductive type comes with introduction rules, which show how to construct an element of the type, and elimination rules, which show how to "use" an element of the type in another construction. The analogy to the logical connectives should not come as a surprise; as we will see below, they, too, are examples of inductive type constructions. You have already seen the introduction rules for an inductive type: they are just the constructors that are specified in the definition of the type. The elimination rules provide for a principle of recursion on the type, which includes, as a special case, a principle of induction as well.

In the next chapter, we will describe Lean's function definition package, which provides even more convenient ways to define functions on inductive types and carry out inductive proofs. But because the notion of an inductive type is so fundamental, we feel it is important to start with a low-level, hands-on understanding. We will start with some basic examples of inductive types, and work our way up to more elaborate and complex examples.

7.1 Enumerated Types

The simplest kind of inductive type is simply a type with a finite, enumerated list of elements.

```
inductive weekday : Type
| sunday : weekday
| monday : weekday
| tuesday : weekday
| wednesday : weekday
| thursday : weekday
| friday : weekday
| saturday : weekday
```

The `inductive` command creates a new type, `weekday`. The constructors all live in the `weekday` namespace.

```
check weekday.sunday
check weekday.monday

open weekday

check sunday
check monday
```

Think of `sunday`, `monday`, …, `saturday` as being distinct elements of `weekday`, with no other distinguishing properties. The elimination principle, `weekday.rec`, is defined along with the type weekday and its constructors. It is also known as a *recursor*, and it is what makes the type "inductive": it allows us to define a function on weekday by assigning values corresponding to each constructor. The intuition is that an inductive type is exhaustively generated by the constructors, and has no elements beyond those they construct.

We will use a slight variant of weekday.rec, weekday.rec_on (also generated automatically), which takes its arguments in a more convenient order. (Note that the shorter names rec and rec_on are not made available by default when we open the weekday namespace. This avoids clashes with the functions of the same names for other inductive types.) If we import nat, we can use weekday.rec_on to define a function from weekday to the natural numbers:

```
def number_of_day (d : weekday) : ℕ :=
weekday.rec_on d 1 2 3 4 5 6 7

eval number_of_day weekday.sunday
eval number_of_day weekday.monday
eval number_of_day weekday.tuesday
```

The first (explicit) argument to rec_on is the element being "analyzed." The next seven arguments are the values corresponding to the seven constructors. Note that number_of_day weekday.sunday evaluates to 1: the computation rule for rec_on recognizes that sunday is a constructor, and returns the appropriate argument.

Below we will encounter a more restricted variant of rec_on, namely, cases_on. When it comes to enumerated types, rec_on and cases_on are the same. You may prefer to use the label cases_on, because it emphasizes that the definition is really a definition by cases.

```
def number_of_day (d : weekday) : ℕ :=
weekday.cases_on d 1 2 3 4 5 6 7
```

It is often useful to group definitions and theorems related to a structure in a namespace with the same name. For example, we can put the number_of_day function in the weekday namespace. We are then allowed to use the shorter name when we open the namespace.

The names rec_on, cases_on, induction_on, and so on are generated automatically. As noted above, they are *protected* to avoid name clashes. In other words, they are not provided by default when the namespace is opened. However, you can explicitly declare abbreviations for them using the renaming option when you open a namespace.

```
namespace weekday
  @[reducible]
  private def cases_on := @weekday.cases_on

  def number_of_day (d : weekday) : nat :=
  cases_on d 1 2 3 4 5 6 7
end weekday
```

```
eval weekday.number_of_day weekday.sunday

open weekday (renaming cases_on → cases_on)

eval number_of_day sunday
check cases_on
```

We can define functions from `weekday` to `weekday`:

```
namespace weekday
  def next (d : weekday) : weekday :=
  weekday.cases_on d monday tuesday wednesday thursday friday saturday sunday

  def previous (d : weekday) : weekday :=
  weekday.cases_on d saturday sunday monday tuesday wednesday thursday friday

  eval next (next tuesday)
  eval next (previous tuesday)

  example : next (previous tuesday) = tuesday := rfl
end weekday
```

How can we prove the general theorem that `next (previous d) = d` for any weekday `d`? The induction principle parallels the recursion principle: we simply have to provide a proof of the claim for each constructor:

```
  theorem next_previous (d: weekday) : next (previous d) = d :=
  weekday.induction_on d
    (show next (previous sunday) = sunday, from rfl)
    (show next (previous monday) = monday, from rfl)
    (show next (previous tuesday) = tuesday, from rfl)
    (show next (previous wednesday) = wednesday, from rfl)
    (show next (previous thursday) = thursday, from rfl)
    (show next (previous friday) = friday, from rfl)
    (show next (previous saturday) = saturday, from rfl)
```

In fact, `induction_on` is just a special case of `rec_on` where the target type is an element of `Prop`. In other words, under the propositions-as-types correspondence, the principle of induction is a type of definition by recursion, where what is being "defined" is a proof instead of a piece of data. We could equally well have used `cases_on`:

```
  theorem next_previous (d: weekday) : next (previous d) = d :=
  weekday.cases_on d
    (show next (previous sunday) = sunday, from rfl)
    (show next (previous monday) = monday, from rfl)
    (show next (previous tuesday) = tuesday, from rfl)
    (show next (previous wednesday) = wednesday, from rfl)
    (show next (previous thursday) = thursday, from rfl)
    (show next (previous friday) = friday, from rfl)
    (show next (previous saturday) = saturday, from rfl)
```

While the `show` commands make the proof clearer and more readable, they are not necessary:

```
theorem next_previous (d: weekday) : next (previous d) = d :=
weekday.cases_on d rfl rfl rfl rfl rfl rfl rfl
```

Some fundamental data types in the Lean library are instances of enumerated types.

```
inductive empty : Type

inductive unit : Type
| star : unit

inductive bool : Type
| ff : bool
| tt : bool
```

(To run these examples, we put them in a namespace called `hide`, so that a name like `bool` does not conflict with the `bool` in the standard library. This is necessary because these types are part of the Lean "prelude" that is automatically imported with the system is started.)

The type `empty` is an inductive data type with no constructors. The type `unit` has a single element, `star`, and the type `bool` represents the familiar boolean values. As an exercise, you should think about what the introduction and elimination rules for these types do. As a further exercise, we suggest defining boolean operations `band`, `bor`, `bnot` on the boolean, and verifying common identities. Note that defining a binary operation like `band` will require nested cases splits:

```
def band (b1 b2 : bool) : bool :=
bool.cases_on b1
  ff
  (bool.cases_on b2 ff tt)
```

Similarly, most identities can be proved by introducing suitable case splits, and then using `rfl`.

7.2 Constructors with Arguments

Enumerated types are a very special case of inductive types, in which the constructors take no arguments at all. In general, a "construction" can depend on data, which is then represented in the constructed argument. Consider the definitions of the product type and sum type in the library:

```
universe variables u v

inductive prod (α : Type u) (β : Type v)
| mk : α → β → prod

inductive sum (α : Type u) (β : Type v)
| inl {} : α → sum
| inr {} : β → sum
```

Notice that we do not include the types α and β in the target of the constructors. For the moment, ignore the annotation `{}` after the constructors `inl` and `inr`; we will explain that below. In the meanwhile, think about what is going on in these examples. The product type has one constructor, `prod.mk`, which takes two arguments. To define a function on `prod` α β, we can assume the input is of the form `prod.mk a b`, and we have to specify the output, in terms of `a` and `b`. We can use this to define the two projections for prod; remember that the standard library defines notation $\alpha \times \beta$ for `prod` α β and `(a, b)` for `prod.mk a b`.

```
def fst {α : Type u} {β : Type v} (p : α × β) : α :=
prod.rec_on p (λ a b, a)

def snd {α : Type u} {β : Type v} (p : α × β) : β :=
prod.rec_on p (λ a b, b)
```

The function `fst` takes a pair, `p`. Applying the recursor `prod.rec_on p (fun a b, a)` interprets `p` as a pair, `prod.mk a b`, and then uses the second argument to determine what to do with `a` and `b`. Remember that you can enter the symbol for a product by typing `\times`. Recall also from Section 2.8 that to give these definitions the greatest generality possible, we allow the types α and β to belong to any universe.

Here is another example:

```
def prod_example (p : bool × ℕ) : ℕ :=
prod.rec_on p (λ b n, cond b (2 * n) (2 * n + 1))

eval prod_example (tt, 3)
eval prod_example (ff, 3)
```

The `cond` function is a boolean conditional: `cond b t1 t2` return `t1` if `b` is true, and `t2` otherwise. (It has the same effect as `bool.rec_on b t2 t1`.) The function `prod_example` takes a pair consisting of a boolean, `b`, and a number, `n`, and returns either `2 * n` or `2 * n + 1` according to whether `b` is true or false.

In contrast, the sum type has *two* constructors, `inl` and `inr` (for "insert left" and "insert right"), each of which takes *one* (explicit) argument. To define a function on `sum` α β, we have to handle two cases: either the input is of the form `inl a`, in which case we have to specify an output value in terms of `a`, or the input is of the form `inr b`, in which case we have to specify an output value in terms of `b`.

```
def sum_example (s : ℕ ⊕ ℕ) : ℕ :=
sum.cases_on s (λ n, 2 * n) (λ n, 2 * n + 1)

eval sum_example (sum.inl 3)
eval sum_example (sum.inr 3)
```

This example is similar to the previous one, but now an input to `sum_example` is implicitly either of the form `inl n` or `inr n`. In the first case, the function returns `2 * n`, and the second case, it returns `2 * n + 1`. You can enter the symbol for the sum by typing `\oplus`.

In the section after next we will see what happens when the constructor of an inductive type takes arguments from the inductive type itself. What characterizes the examples we consider in this section is that this is not the case: each constructor relies only on previously specified types.

Notice that a type with multiple constructors is disjunctive: an element of `sum` α β is either of the form `inl a` *or* of the form `inl b`. A constructor with multiple arguments introduces conjunctive information: from an element `prod.mk a b` of `prod` α β we can extract `a` *and* `b`. An arbitrary inductive type can include both features, by having any number of constructors, each of which takes any number of arguments.

A type, like `prod`, with only one constructor is purely conjunctive: the constructor simply packs the list of arguments into a single piece of data, essentially a tuple where the type of subsequent arguments can depend on the type of the initial argument. We can also think of such a type as a “record” or a “structure”. In Lean, these two words are synonymous, and provide alternative syntax for inductive types with a single constructor.

```
structure prod (α β : Type) :=
mk :: (fst : α) (snd : β)
```

The `structure` command simultaneously introduces the inductive type, `prod`, its constructor, `mk`, the usual eliminators (`rec`, `rec_on`), as well as the projections, `fst` and `snd`, as defined above.

If you do not name the constructor, Lean uses `mk` as a default. For example, the following defines a record to store a color as a triple of RGB values:

```
record color := (red : nat) (green : nat) (blue : nat)
def yellow := color.mk 255 255 0
eval color.red yellow
```

The definition of `yellow` forms the record with the three values shown, and the projection `color.red` returns the red component. The `structure` command is especially useful for defining algebraic structures, and Lean provides substantial infrastructure to support working with them. Here, for example, is the definition of a semigroup:

```
universe variable u

structure Semigroup :=
(carrier : Type u)
(mul : carrier → carrier → carrier)
(mul_assoc : ∀ a b c, mul (mul a b) c = mul a (mul b c))
```

We will see more examples in Chapter 9.

Notice that the product type depends on parameters α β `: Type` which are arguments to the constructors as well as `prod`. Lean detects when these arguments can be inferred from later arguments to a constructor, and makes them implicit in that case. Sometimes an argument can only be inferred from the return type, which means that it could not be inferred by parsing the expression from bottom up, but may be inferrable from context. In that case, Lean does not make the argument implicit by default, but will do so if we add the annotation `{}` after the constructor. We used that option, for example, in the definition of `sum`:

```
inductive sum (α : Type u) (β : Type v)
| inl {} : α → sum
| inr {} : β → sum
```

αs a result, the argument α to `inl` and the argument β to `inr` are left implicit.

We have already discussed sigma types, also known as the dependent product:

```
inductive sigma {α : Type u} (β : α → Type v)
| dpair : Π a : α, β a → sigma
```

Two more examples of inductive types in the library are the following:

```
inductive option (α : Type u)
| none {} : option
| some    : α → option

inductive inhabited (α : Type u)
| mk : α → inhabited
```

In the semantics of dependent type theory, there is no built-in notion of a partial function. Every element of a function type $\alpha \to \beta$ or a Pi type Π `x :` α`,` β is assumed to have a value at every input. The `option` type provides a way of representing partial functions. An element of `option` β is either `none` or of the form `some b`, for some value `b :` β. Thus we can think of an element `f` of the type $\alpha \to$ `option` β as being a partial function from α to β: for every `a :` α, `f a` either returns `none`, indicating the `f a` is “undefined”, or `some b`.

An element of `inhabited α` is simply a witness to the fact that there is an element of α. Later, we will see that `inhabited` is an example of a *type class* in Lean: Lean can be instructed that suitable base types are inhabited, and can automatically infer that other constructed types are inhabited on that basis.

As exercises, we encourage you to develop a notion of composition for partial functions from α to β and β to γ, and show that it behaves as expected. We also encourage you to show that `bool` and `nat` are inhabited, that the product of two inhabited types is inhabited, and that the type of functions to an inhabited type is inhabited.

7.3 Inductively Defined Propositions

Inductively defined types can live in any type universe, including the bottom-most one, `Prop`. In fact, this is exactly how the logical connectives are defined.

```
inductive false : Prop

inductive true : Prop
| intro : true

inductive and (a b : Prop) : Prop
| intro : a → b → and

inductive or (a b : Prop) : Prop
| intro_left  : a → or
| intro_right : b → or
```

You should think about how these give rise to the introduction and elimination rules that you have already seen. There are rules that govern what the eliminator of an inductive type can eliminate *to*, that is, what kinds of types can be the target of a recursor. Roughly speaking, what characterizes inductive types in `Prop` is that one can only eliminate to other types in `Prop`. This is consistent with the understanding that if `p : Prop`, an element `hp : p` carries no data. There is a small exception to this rule, however, which we will discuss below, in the section on inductive families.

Even the existential quantifier is inductively defined:

```
inductive Exists {α : Type u} (p : α → Prop) : Prop
| intro : ∀ (a : α), p a → Exists

def exists.intro := @Exists.intro
```

Keep in mind that the notation `∃ x : α, p` is syntactic sugar for `Exists (λ x : α, p)`.

The definitions of `false`, `true`, `and`, and `or` are perfectly analogous to the definitions of `empty`, `unit`, `prod`, and `sum`. The difference is that the first group yields elements of `Prop`, and the second yields elements of `Type i` for `i` greater than 0. In a similar way, `∃ x : α, p` is a `Prop`-valued variant of `Σ x : α, p`.

This is a good place to mention another inductive type, denoted `{x : α | p}`, which is sort of a hybrid between `∃ x : α, P` and `Σ x : α, P`.

```
inductive subtype {α : Type u} (p : α → Prop)
| tag : Π x : α, p x → subtype
```

The notation `{x : α | p}` is syntactic sugar for `subtype (λ x : α, p)`. It is modeled after subset notation in set theory: the idea is that `{x : α | p}` denotes the collection of elements of α that have property `p`.

7.4 Defining the Natural Numbers

The inductively defined types we have seen so far are "flat": constructors wrap data and insert it into a type, and the corresponding recursor unpacks the data and acts on it. Things get much more interesting when the constructors act on elements of the very type being defined. A canonical example is the type `nat` of natural numbers:

```
inductive nat : Type
| zero : nat
| succ : nat → nat
```

There are two constructors. We start with `zero : nat`; it takes no arguments, so we have it from the start. In contrast, the constructor `succ` can only be applied to a previously constructed `nat`. Applying it to `zero` yields `succ zero : nat`. Applying it again yields `succ (succ zero) : nat`, and so on. Intuitively, `nat` is the "smallest" type with these constructors, meaning that it is exhaustively (and freely) generated by starting with `zero` and applying `succ` repeatedly.

As before, the recursor for `nat` is designed to define a dependent function `f` from `nat` to any domain, that is, an element `f` of `Π n : nat, C n` for some `C : nat → Type`. It has to handle two cases: the case where the input is `zero`, and the case where the input is of the form `succ n` for some `n : nat`. In the first case, we simply specify a target value with the appropriate type, as before. In the second case, however, the recursor can assume that a value of `f` at `n` has already been computed. As a result, the next argument to the recursor specifies a value for `f (succ n)` in terms of `n` and `f n`. If we check the type of the recursor,

```
check @nat.rec_on
```

we find the following:

```
Π {C : nat → Type} (n : nat),
  C nat.zero → (Π (a : nat), C a → C (nat.succ a)) → C n
```

The implicit argument, `C`, is the codomain of the function being defined. In type theory it is common to say `C` is the `motive` for the elimination/recursion. The next argument, `n : nat`, is the input to the function. It is also known as the `major premise`. Finally, the two arguments after specify how to compute the zero and successor cases, as described above. They are also known as the `minor premises`.

Consider, for example, the addition function `add m n` on the natural numbers. Fixing `m`, we can define addition by recursion on `n`. In the base case, we set `add m zero` to `m`. In the successor step, assuming the value `add m n` is already determined, we define `add m (succ n)` to be `succ (add m n)`.

```
namespace nat

def add (m n : nat) : nat :=
nat.rec_on n m (λ n add_m_n, succ add_m_n)

-- try it out
eval add (succ zero) (succ (succ zero))

end nat
```

It is useful to put such definitions into a namespace, `nat`. We can then go on to define familiar notation in that namespace. The two defining equations for addition now hold definitionally:

```
instance : has_zero nat := has_zero.mk zero
instance : has_add nat := has_add.mk add

theorem add_zero (m : nat) : m + 0 = m := rfl
theorem add_succ (m n : nat) : m + succ n = succ (m + n) := rfl
```

We will explain how the `instance` command works in Chapter 10. In the examples below, we will henceforth use Lean's version of the natural numbers.

Proving a fact like `0 + m = m`, however, requires a proof by induction. As observed above, the induction principle is just a special case of the recursion principle, when the codomain `C n` is an element of `Prop`. It represents the familiar pattern of an inductive proof: to prove `∀ n, C n`, first prove `C 0`, and then, for arbitrary `n`, assume `ih : C n` and prove `C (succ n)`.

```
theorem zero_add (n : ℕ) : 0 + n = n :=
nat.induction_on n
  (show 0 + 0 = 0, from rfl)
  (take n,
    assume ih : 0 + n = n,
    show 0 + succ n = succ n, from
      calc
        0 + succ n = succ (0 + n) : rfl
          ... = succ n : by rw ih)
```

In the example above, we encourage you to replace `induction_on` with `rec_on` and observe that the theorem is still accepted by Lean. As we have seen above, `induction_on` is just a special case of `rec_on`.

For another example, let us prove the associativity of addition, `∀ m n k, m + n + k = m + (n + k)`. (The notation `+`, as we have defined it, associates to the left, so `m + n + k` is really `(m + n) + k`.) The hardest part is figuring out which variable to do the induction on. Since addition is defined by recursion on the second argument, `k` is a good guess, and once we make that choice the proof almost writes itself:

```
theorem add_assoc (m n k : ℕ) : m + n + k = m + (n + k) :=
nat.induction_on k
  (show m + n + 0 = m + (n + 0), from rfl)
  (take k,
    assume ih : m + n + k = m + (n + k),
    show m + n + succ k = m + (n + succ k), from
      calc
        m + n + succ k = succ (m + n + k) : rfl
          ... = succ (m + (n + k)) : by rw ih
          ... = m + succ (n + k) : rfl
          ... = m + (n + succ k) : rfl)
```

For another example, suppose we try to prove the commutativity of addition. Choosing induction on the second argument, we might begin as follows:

```
theorem add_comm (m n : nat) : m + n = n + m :=
nat.induction_on n
  (show m + 0 = 0 + m, by rw [nat.zero_add, nat.add_zero])
  (take n,
    assume ih : m + n = n + m,
    calc
      m + succ n = succ (m + n) : rfl
        ... = succ (n + m) : by rw ih
        ... = succ n + m : sorry)
```

At this point, we see that we need another supporting fact, namely, that `succ (n + m) = succ n + m`. We can prove this by induction on `m`:

```
theorem succ_add (m n : nat) : succ m + n = succ (m + n) :=
nat.induction_on n
  (show succ m + 0 = succ (m + 0), from rfl)
  (take n,
    assume ih : succ m + n = succ (m + n),
    show succ m + succ n = succ (m + succ n), from
      calc
        succ m + succ n = succ (succ m + n) : rfl
          ... = succ (succ (m + n)) : by rw ih
          ... = succ (m + succ n) : rfl)
```

We can then replace the `sorry` in the previous proof with `succ_add`.

As an exercise, try defining other operations on the natural numbers, such as multiplication, the predecessor function (with `pred 0 = 0`), truncated subtraction (with `n - m = 0` when `m` is greater than or equal to `n`), and exponentiation. Then try proving some of their basic properties, building on the theorems we have already proved.

7.5 Tactics

Given the fundamental importance of inductive types in Lean, it should not be surprising that there are a number of tactics described to work with them effectively. We describe some of them here.

The `cases` tactic works on elements of an inductively defined type, and does what the name suggests: it decomposes the element according to each of the possible constructors. In its most basic form, it is applied to an element `x` in the local context. It then reduces the goal to cases in which `x` is replaced by each of the constructions.

```
open nat
variable p : ℕ → Prop

example (hz : p 0) (hs : ∀ n, p (succ n)) : ∀ n, p n :=
begin
  intro n,
  cases n,
  { exact hz },  -- goal is p 0
  apply hs       -- goal is a : ℕ ⊢ p (succ a)
end
```

There are extra bells and whistles. For one thing, `cases` allows you to choose the names for the arguments to the constructors using a `with` clause. In the next example, for example, we choose the name `m` for the argument to `succ`, so that the second case refers to `succ m`. More importantly, the cases tactic will detect any items in the local context that depend on the target variable. It reverts these elements, does the split, and reintroduces them. In the example below, notice that the hypothesis `h : n ≠ 0` becomes `h : 0 ≠ 0` in the first branch, and `h : succ m ≠ 0` in the second.

```
open nat

example (n : ℕ) (h : n ≠ 0) : succ (pred n) = n :=
begin
  cases n with m,
  -- first goal: h : 0 ≠ 0 ⊢ succ (pred 0) = 0
    { apply (absurd rfl h) },
  -- second goal: h : succ m ≠ 0 ⊢ succ (pred (succ a)) = succ a
  reflexivity
end
```

Notice that `cases` can be used to produce data as well as prove propositions.

```
def f (n : ℕ) : ℕ :=
begin
  cases n, exact 3, exact 7
end

example : f 0 = 3 := rfl
example : f 5 = 7 := rfl
```

Once again, cases will revert and depedencies in the context, split, and then reintroduce them.

```
universe variable u

def tuple (α : Type u) (n : ℕ) := { l : list α // list.length l = n }

variables {α : Type u} {n : ℕ}

def f {n : ℕ} (t : tuple α n) : ℕ :=
begin
  cases n, exact 3, exact 7
end

def my_tuple : tuple ℕ 3 := ⟨[0, 1, 2], rfl⟩

example : f my_tuple = 7 := rfl
```

If there are multiple constructors with arguments, you can provide `cases` with a list of all the names, arranged sequentially:

```
inductive foo : Type
| bar1 : ℕ → ℕ → foo
| bar2 : ℕ → ℕ → ℕ → foo

def silly (x : foo) : ℕ :=
begin
  cases x with a b c d e,
  exact b,    -- a, b, c are in the context
  exact e     -- d, e    are in the context
end
```

You can also use `cases` with an arbitrary expression. Assuming that expression occurs in the goal, the cases tactic will generalize over the expression, introduce the resulting universally quantified variable, and case on that.

```
open nat
variable p : ℕ → Prop

example (hz : p 0) (hs : ∀ n, p (succ n)) (m k : ℕ) : p (m + 3 * k) :=
begin
  cases (m + 3 * k),
  { exact hz },  -- goal is p 0
```

```
  apply hs        -- goal is a : ℕ ⊢ p (succ a)
end
```

Think of this as saying "split on cases as to whether `m + 3 * k` is zero or the successor of some number." The result is functionally equivalent to the following:

```
example (hz : p 0) (hs : ∀ n, p (succ n)) (m k : ℕ) : p (m + 3 * k) :=
begin
  generalize (m + 3 * k) n,
  intro n,
  cases n,
  { exact hz },  -- goal is p 0
  apply hs        -- goal is a : ℕ ⊢ p (succ a)
end
```

Notice that the expression `m + 3 * k` is erased by generalize; all that matters is whether it is of the form `0` or `succ a`. This form of `cases` will *not* revert any hypotheses that also mention the expression in equation (in this case, `m + 3 * k`). If such a term appears in a hypothesis and you want to generalize over that as well, you need to `revert` it explicitly.

If the expression you case on does not appear in the goal, the `cases` tactic uses `assert` to put the type of the expression into the context. Here is an example:

```
example (p : Prop) (m n : ℕ) (h₁ : m < n → p) (h₂ : m ≥ n → p) : p :=
begin
  cases lt_or_ge m n with hlt hge,
  { exact h₁ hlt },
  exact h₂ hge
end
```

The theorem `lt_or_ge m n` says `m < n ∨ m ≥ n`, and it is natural to think of the proof above as splitting on these two cases. In the first branch, we have the hypothesis `h₁ : m < n`, and in the second we have the hypothesis `h₂ : m ≥ n`. The proof above is functionally equivalent to the following:

```
example (p : Prop) (m n : ℕ) (h₁ : m < n → p) (h₂ : m ≥ n → p) : p :=
begin
  assert h : m < n ∨ m ≥ n,
  { exact lt_or_ge m n },
  cases h with hlt hge,
  { exact h₁ hlt },
  exact h₂ hge
end
```

After the first two lines, we have `h : m < n ∨ m ≥ n` as a hypothesis, and we simply do cases on that.

Here is another example, where we use the decidability of equality on the natural numbers to split on the cases `m = n` and `m ≠ n`.

```
check nat.sub_self

example (m n : ℕ) : m - n = 0 ∨ m ≠ n :=
begin
  cases decidable.em (m = n) with heq hne,
  { rw heq,
    left, exact nat.sub_self n },
  right, exact hne
end
```

Remember that if you `open classical`, you can use the law of the excluded middle for any proposition at all. But using type class inference (see Chapter 10), Lean can actually find the relevant decision procedure, which means that you can use the case split in a computable function.

```
def f (m k : ℕ) : ℕ :=
begin
  cases m - k, exact 3, exact 7
end

example : f 5 7 = 3 := rfl
example : f 10 2 = 7 := rfl
```

Aspects of computability will be discussed in a later chapter.

7.6 Other Inductive Types

Let us consider some more examples of inductively defined types. For any type, α, the type `list` α of lists of elements of α is defined in the library.

```
inductive list (α : Type u)
| nil {} : list
| cons : α → list → list

namespace list

variable {α : Type}

notation h :: t  := cons h t

def append (s t : list α) : list α :=
list.rec t (λ x l u, x::u) s

notation s ++ t := append s t

theorem nil_append (t : list α) : nil ++ t = t := rfl

theorem cons_append (x : α) (s t : list α) : x::s ++ t = x::(s ++ t) := rfl

end list
```

A list of elements of type α is either the empty list, `nil`, or an element `h : α` followed by a list `t : list α`. We define the notation `h :: t` to represent the latter. The first element, `h`, is commonly known as the "head" of the list, and the remainder, `t`, is known as the "tail." Recall that the notation `{}` in the definition of the inductive type ensures that the argument to `nil` is implicit. In most cases, it can be inferred from context. When it cannot, we have to write `@nil α` to specify the type α.

Lean allows us to define iterative notation for lists:

```
inductive list (α : Type u)
| nil {} : list
| cons : α → list → list

namespace list

notation `[` l:(foldr `,` (h t, cons h t) nil) `]` := l

section
  open nat
  check [1, 2, 3, 4, 5]
  check ([1, 2, 3, 4, 5] : list num)
end

end list
```

In the first `check`, Lean assumes that `[1, 2, 3, 4, 5]` is a list of natural numbers. The `(t : list num)` expression forces Lean to interpret `t` as a list of numerals.

As an exercise, prove the following:

```
theorem append_nil (t : list α) : t ++ nil = t := sorry

theorem append_assoc (r s t : list α) : r ++ s ++ t = r ++ (s ++ t) := sorry
```

Try also defining the function `length : Π α : Type, list α → nat` that returns the length of a list, and prove that it behaves as expected (for example, `length (s ++ t) = length s + length t`).

For another example, we can define the type of binary trees:

```
inductive binary_tree
| leaf : binary_tree
| node : binary_tree → binary_tree → binary_tree
```

In fact, we can even define the type of countably branching trees:

```
inductive cbtree
| leaf : cbtree
| sup : (ℕ → cbtree) → cbtree
```

```
namespace cbtree

def succ (t : cbtree) : cbtree :=
sup (λ n, t)

def omega : cbtree :=
sup (λ n, nat.rec_on n leaf (λ n t, succ t))

end cbtree
```

8

Induction and Recursion

Other than the type universes and Pi types, inductively defined types provide the only means of defining new types in the Calculus of Inductive Constructions. We have also seen that, fundamentally, the constructors and the recursors provide the only means of defining functions on these types. By the propositions-as-types correspondence, this means that induction is the fundamental method of proof for these types.

Working with induction and recursion is therefore fundamental to working in the Calculus of Inductive Constructions. For that reason Lean provides more natural ways of defining recursive functions, performing pattern matching, and writing inductive proofs. Behind the scenes, these are "compiled" down to recursors.

Thus, the function definition package, which performs this reduction, is not part of the trusted code base.

8.1 Pattern Matching

The `cases_on` recursor can be used to define functions and prove theorems by cases. But complicated definitions may use several nested `cases_on` applications, and may be hard to read and understand. Pattern matching provides a more convenient and standard way of defining functions and proving theorems. Lean supports a very general form of pattern matching called *dependent pattern matching*.

A pattern-matching definition is of the following form:

```
definition [name] [parameters] : [domain] → [codomain]
| [patterns_1] := [value_1]
...
| [patterns_n] := [value_n]
```

The parameters are fixed, and each assignment defines the value of the function for a different case specified by the given pattern. As a first example, we define the function `sub2` for natural numbers:

```
open nat

def sub2 : nat → nat
| 0     := 0
| 1     := 0
| (a+2) := a

example : sub2 5 = 3 := rfl
```

The default compilation method guarantees that the pattern matching equations hold definitionally.

```
example : sub2 0 = 0 := rfl

example : sub2 1 = 0 := rfl

example (a : nat) : sub2 (a + 2) = a := rfl
```

We can use the command `print sub2` to see how our definition was compiled into recursors.

```
print sub2
```

We will say a term is a *constructor application* if it is of the form `c a_1 ... a_n` where `c` is the constructor of some inductive data type. Note that in the definition `sub2`, the terms `1` and `a+2` are not constructor applications. However, the compiler normalizes them at compilation time, and obtains the constructor applications `succ zero` and `succ (succ a)` respectively. This normalization step is just a convenience that allows us to write definitions resembling the ones found in textbooks. There is no magic here: the compiler simply uses the kernel's ordinary evaluation mechanism. If we had written `2+a`, the definition would be rejected since `2+a` does not normalize into a constructor application.

In the next example, we use pattern-matching to define Boolean negation `bnot`, and proving `bnot (bnot b) = b`.

```
def bnot : bool → bool
| tt := ff
| ff := tt

theorem bnot_bnot : ∀ (b : bool), bnot (bnot b) = b
| tt := rfl    -- proof that bnot (bnot tt) = tt
| ff := rfl    -- proof that bnot (bnot ff) = ff
```

As described in Chapter 7, Lean inductive data types can be parametric. The following example defines the `tail` function using pattern matching. The argument `α : Type` is a parameter and occurs before the colon to indicate it does not participate in the pattern matching. Lean allows parameters to occur after `:`, but it cannot pattern match on them.

```
open list

def tail1 {α : Type} : list α → list α
| nil      := nil
| (h :: t) := t

-- Parameter α may occur after ':'
def tail2 : Π {α : Type}, list α → list α
| α nil      := nil
| α (h :: t) := t
```

8.2 Structural Recursion and Induction

The function definition package supports structural recursion, that is, recursive applications where one of the arguments is a subterm of the corresponding term on the left-hand-side. Later, we describe how to compile recursive equations using well-founded recursion. The main advantage of the default compilation method is that the recursive equations hold definitionally.

Here are some examples from the last chapter, written in the new style:

```
def add : nat → nat → nat
| m zero     := m
| m (succ n) := succ (add m n)

local infix `+` := add

theorem add_zero (m : nat) : m + zero = m := rfl
theorem add_succ (m n : nat) : m + succ n = succ (m + n) := rfl

theorem zero_add : ∀ n, zero + n = n
| zero     := rfl
| (succ n) := congr_arg succ (zero_add n)

def mul : nat → nat → nat
| n zero     := zero
| n (succ m) := mul n m + m
```

The "definition" of `zero_add` makes it clear that proof by induction is really a form of induction in Lean.

As with definition by pattern matching, parameters to a structural recursion or induction may appear before the colon. Such parameters are simply added to the local context before the definition is processed. For example, the definition of addition may be written as follows:

```
def add (m : nat) : nat → nat
| zero     := m
| (succ n) := succ (add n)
```

This may seem a little odd, but you should read the definition as follows: "Fix `m`, and define the function which adds something to `m` recursively, as follows. To add zero, return `m`. To add the successor of `n`, first add `n`, and then take the successor." The mechanism for adding parameters to the local context is what makes it possible to process match expressions within terms, as described below.

A more interesting example of structural recursion is given by the Fibonacci function `fib`.

```
def fib : nat → nat
| 0     := 1
| 1     := 1
| (a+2) := fib (a+1) + fib a

-- the defining equations hold definitionally
example : fib 0 = 1 := rfl
example : fib 1 = 1 := rfl
example (a : nat) : fib (a+2) = fib (a+1) + fib a := rfl
```

Another classic example is the list `append` function.

```
def append {α : Type} : list α → list α → list α
| []     l := l
| (h::t) l := h :: append t l

example : append [(1 : ℕ), 2, 3] [4, 5] = [1, 2, 3, 4, 5] := rfl
```

8.3 Dependent Pattern-Matching

All the examples we have seen so far can be easily written using `cases_on` and `rec_on`. However, this is not the case with indexed inductive families, such as `vector α n`. A lot of boilerplate code needs to be written to define very simple functions such as `map`, `zip`, and `unzip` using recursors.

To understand the difficulty, consider what it would take to define a function `tail` which takes a vector `v : vector α (succ n)` and deletes the first element. A first thought might be to use the `cases_on` function:

```
open nat

inductive vector (α : Type) : nat → Type
| nil {} : vector 0
| cons   : Π {n}, α → vector n → vector (succ n)
```

```
open vector
local notation h :: t := cons h t

check @vector.cases_on
-- Π {α : Type}
--   {C : Π (a : ℕ), vector α a → Type}
--   {a : ℕ}
--   (n : vector α a),
--   (e1 : C 0 nil)
--   (e2 : Π {n : ℕ} (a : α) (a_1 : vector α n), C (succ n) (cons a a_1)),
--   C a n
```

But what value should we return in the `nil` case? Something funny is going on: if `v` has type `vector α (succ n)`, it *can't* be nil, but it is not clear how to tell that to `cases_on`.

One standard solution is to define an auxiliary function:

```
def tail_aux {α : Type} {n m : nat} (v : vector α m) :
    m = succ n → vector α n :=
vector.cases_on v
  (assume H : 0 = succ n, nat.no_confusion H)
  (take m (a : α) w : vector α m,
    assume H : succ m = succ n,
      nat.no_confusion H (λ H1 : m = n, eq.rec_on H1 w))

def tail {α : Type} {n : nat} (v : vector α (succ n)) : vector α n :=
tail_aux v rfl
```

In the `nil` case, `m` is instantiated to `0`, and `no_confusion` makes use of the fact that `0 = succ n` cannot occur. Otherwise, `v` is of the form `a :: w`, and we can simply return `w`, after casting it from a vector of length `m` to a vector of length `n`.

The difficulty in defining `tail` is to maintain the relationships between the indices. The hypothesis `e : m = succ n` in `tail_aux` is used to "communicate" the relationship between `n` and the index associated with the minor premise. Moreover, the `zero = succ n` case is "unreachable," and the canonical way to discard such a case is to use `no_confusion`.

The `tail` function is, however, easy to define using recursive equations, and the function definition package generates all the boilerplate code automatically for us.

Here are a number of examples:

```
def head {α : Type} : Π {n}, vector α (succ n) → α
| n (h :: t) := h

def tail {α : Type} : Π {n}, vector α (succ n) → vector α n
| n (h :: t) := t

lemma eta {α : Type} : ∀ {n} (v : vector α (succ n)), head v :: tail v = v
| n (h::t) := rfl

def map {α β γ : Type} (f : α → β → γ)
               : Π {n : nat}, vector α n → vector β n → vector γ n
```

```
| 0        nil     nil     := nil
| (succ n) (a::va) (b::vb) := f a b :: map va vb

def zip {α β : Type} : Π {n}, vector α n → vector β n → vector (α × β) n
| 0        nil nil         := nil
| (succ n) (a::va) (b::vb) := (a, b) :: zip va vb
```

Note that we can omit recursive equations for "unreachable" cases such as `head nil`. The automatically generated definitions for indexed families are far from straightforward. For example:

```
print map
print map._main
```

The `map` function is even more tedious to define by hand than the `tail` function. We encourage you to try it, using `rec_on`, `cases_on` and `no_confusion`.

8.4 Variations on Pattern Matching

We say that a set of recursive equations *overlaps* when there is an input that more than one left-hand-side can match. In the following definition the input `0 0` matches the left-hand-side of the first two equations. Should the function return `1` or `2`?

```
def f : nat → nat → nat
| 0     y     := 1
| x     0     := 2
| (x+1) (y+1) := 3
```

Overlapping patterns are often used to succinctly express complex patterns in data, and they are allowed in Lean. Lean handles the ambiguity by using the first applicable equation. In the example above, the following equations hold definitionally:

```
variables (a b : nat)

example : f 0     0     = 1 := rfl
example : f 0     (a+1) = 1 := rfl
example : f (a+1) 0     = 2 := rfl
example : f (a+1) (b+1) = 3 := rfl
```

Lean also supports *wildcard patterns*, also known as *anonymous variables*. They are used to create patterns where we don't care about the value of a specific argument. In the function `f` defined above, the values of `x` and `y` are not used in the right-hand-side. Here is the same example using wildcards:

```
def f : nat → nat → nat
| 0  _  := 1
| _  0  := 2
| _  _  := 3
variables (a b : nat)
example : f 0     0     = 1 := rfl
example : f 0     (a+1) = 1 := rfl
example : f (a+1) 0     = 2 := rfl
example : f (a+1) (b+1) = 3 := rfl
```

Some functional languages support *incomplete patterns*. In these languages, the interpreter produces an exception or returns an arbitrary value for incomplete cases. We can simulate the arbitrary value approach using the `inhabited` type class, discussed in Chapter 10 Roughly, an element of `inhabited` α is simply a witness to the fact that there is an element of α; in Chapter 10 we will see that Lean can be instructed that suitable base types are inhabited, and can automatically infer that other constructed types are inhabited on that basis. On this basis, the standard library provides an arbitrary element, `arbitrary` α, of any inhabited type.

We can also use the type `option` α to simulate incomplete patterns. The idea is to return `some a` for the provided patterns, and use `none` for the incomplete cases. The following example demonstrates both approaches.

```
def f1 : nat → nat → nat
| 0  _  := 1
| _  0  := 2
| _  _  := arbitrary nat   -- the "incomplete" case

variables (a b : nat)

example : f1 0     0     = 1 := rfl
example : f1 0     (a+1) = 1 := rfl
example : f1 (a+1) 0     = 2 := rfl
example : f1 (a+1) (b+1) = arbitrary nat := rfl

def f2 : nat → nat → option nat
| 0  _  := some 1
| _  0  := some 2
| _  _  := none            -- the "incomplete" case

example : f2 0     0     = some 1 := rfl
example : f2 0     (a+1) = some 1 := rfl
example : f2 (a+1) 0     = some 2 := rfl
example : f2 (a+1) (b+1) = none   := rfl
```

8.5 Inaccessible Terms

Sometimes an argument in a dependent matching pattern is not essential to the definition, but nonetheless has to be included to specialize the type of the expression appropriately.

Lean allows users to mark such subterms as *inaccessible* for pattern matching. These annotations are essential, for example, when a term occurring in the left-hand side is neither a variable nor a constructor application, because these are not suitable targets for pattern matching. We can view such inaccessible terms as "don't care" components of the patterns. You can declare a subterm inaccessible by writing `.t`. If `t` is a composite term, we need to write `.(t)`.

The following example can be found in cite{goguen:et:al:06}. We declare an inductive type that defines the property of "being in the image of `f`". You can view an element of the type `image_of f b` as evidence that `b` is in the image of `f`, whereby the constructor `imf` is used to build such evidence. We can then define any function `f` with an "inverse" which takes anything in the image of `f` to an element that is mapped to it. The typing rules forces us to write `f a` for the first argument, but this term is not a variable nor a constructor application, and plays no role in the pattern-matching definition. To define the function `inverse` below, we *have to* mark `f a` inaccessible.

```
variables {α β : Type}
inductive image_of (f : α → β) : β → Type
| imf : Π a, image_of (f a)

open image_of

def inverse {f : α → β} : Π b, image_of f b → α
| .(f a) (imf .f a) := a
```

In the example above, the inaccessible annotation makes it clear that `f` is *not* a pattern matching variable.

8.6 Match Expressions

Lean also provides a compiler for *match-with* expressions found in many functional languages. It uses essentially the same infrastructure used to compile recursive equations.

```
def is_not_zero (a : nat) : bool :=
match a with
| 0     := ff
| (n+1) := tt
end

-- We can use recursive equations and match
variable {α : Type}
variable p : α → bool

def filter : list α → list α
| []       := []
| (a :: l) :=
  match p a with
  |  tt := a :: filter l
```

```
  | ff := filter l
  end

example : filter is_not_zero [1, 0, 0, 3, 0] = [1, 3] := rfl
```

8.7 Well-Founded Recursion

[TODO: write this section.]

9

Structures and Records

We have seen that Lean's foundational system includes inductive types. We have, moreover, noted that it is a remarkable fact that it is possible to construct a substantial edifice of mathematics based on nothing more than the type universes, Pi types, and inductive types; everything else follows from those. The Lean standard library contains many instances of inductive types (e.g., `nat`, `prod`, `list`), and even the logical connectives are defined using inductive types.

Remember that a non-recursive inductive type that contains only one constructor is called a *structure* or *record*. The product type is a structure, as is the dependent product type, that is, the Sigma type. In general, whenever we define a structure `S`, we usually define *projection* functions that allow us to "destruct" each instance of `S` and retrieve the values that are stored in its fields. The functions `prod.pr1` and `prod.pr2`, which return the first and second elements of a pair, are examples of such projections.

When writing programs or formalizing mathematics, it is not uncommon to define structures containing many fields. The `structure` command, available in Lean, provides infrastructure to support this process. When we define a structure using this command, Lean automatically generates all the projection functions. The `structure` command also allows us to define new structures based on previously defined ones. Moreover, Lean provides convenient notation for defining instances of a given structure.

9.1 Declaring Structures

The structure command is essentially a "front end" for defining inductive data types. Every `structure` declaration introduces a namespace with the same name. The general form is as follows:

```
structure <name> <parameters> <parent-structures> : Type :=
  <constructor> :: <fields>
```

Most parts are optional. Here is an example:

```
structure point (α : Type) :=
mk :: (x : α) (y : α)
```

Values of type `point` are created using `point.mk a b`, and the fields of a point `p` are accessed using `point.x p` and `point.y p`. The structure command also generates useful recursors and theorems. Here are some of the constructions generated for the declaration above.

```
check point              -- a Type
check point.rec_on       -- the recursor
check point.induction_on -- then recursor to Prop
check point.x            -- a projection / field accessor
check point.y            -- a projection / field accessor
```

You can obtain the complete list of generated constructions using the command `print prefix`.

```
print prefix point
```

Here are some simple theorems and expressions that use the generated constructions. As usual, you can avoid the prefix `point` by using the command `open point`.

```
eval point.x (point.mk 10 20)
eval point.y (point.mk 10 20)

open point

example (α : Type) (a b : α) : x (mk a b) = a :=
rfl

example (α : Type) (a b : α) : y (mk a b) = b :=
rfl
```

Given `p : point nat`, the notation `p^.x` is shorthand for `point.x p`. This provides a convenient way of accessing the fields of a structure.

```
def p := point.mk 10 20

check p^.x -- nat
eval  p^.x -- 10
eval  p^.y -- 20
```

If the constructor is not provided, then a constructor is named `mk` by default.

```
structure prod (α : Type) (β : Type) :=
(pr1 : α) (pr2 : β)

check prod.mk
```

The keyword `record` is an alias for `structure`.

```
record point (α : Type) :=
mk :: (x : α) (y : α)
```

You can provide universe levels explicitly. The annotations in the next example force the parameters α and β to be types from the same universe, and set the return type to also be in the same universe.

```
structure {u} prod (α : Type u) (β : Type u) : Type (max 1 u) :=
(pr1 : α) (pr2 : β)

set_option pp.universes true
check prod.mk
```

The `set_option` command above instructs Lean to display the universe levels.

We use `max 1 l` as the resultant universe level to ensure the universe level is never `0` even when the parameter α and β are propositions. Recall that in Lean, `Type 0` is `Prop`, which is impredicative and proof irrelevant.

We can use the anonymous constructor notation to build structure values whenever the expected type is known.

```
structure {u} prod (α : Type u) (β : Type u) : Type (max 1 u) :=
(pr1 : α) (pr2 : β)

example : prod nat nat :=
⟨1, 2⟩

check (⟨1, 2⟩ : prod nat nat)
```

9.2 Objects

We have been using constructors to create elements of a structure (or record) type. For structures containing many fields, this is often inconvenient, because we have to remember the order in which the fields were defined. Lean therefore provides the following alternative notations for defining elements of a structure type.

```
{ structure-name . <field-name> := <expr>)* }
or
{<field-name> := <expr>)*}
```

The prefix `structure-name .` can be omitted whenever the name of the structure can be inferred from the expected type. For example, we use this notation to define "points." The order that the fields are specified does not matter, so all the expressions below define the same point.

```
structure point (α : Type) :=
mk :: (x : α) (y : α)

check { point . x := 10, y := 20 }   -- point ℕ
check { point . y := 20, x := 10 }
check ({x := 10, y := 20} : point nat)

example : point nat :=
{ y := 20, x := 10 }
```

If the value of a field is not specified, Lean tries to infer it. If the unspecified fields cannot be inferred, Lean signs an error indicating the corresponding placeholder could not be synthesized.

```
structure my_struct :=
mk :: {α : Type} {β : Type} (a : α) (b : β)

check { my_struct . a := 10, b := true }
```

Record update is another common operation. It consists in creating a new record object by modifying the value of one or more fields. Lean provides a variation of the notation described above for record updates.

```
{ record-obj with <field-name> := <expr>)* }
```

The semantics is simple: record objects `<record-obj>` provide the values for the unspecified fields. If more than one record object is provided, then they are visited in order until Lean finds one the contains the unspecified field. Lean raises an error if any of the field names remain unspecified after all the objects are visited.

```
structure point (α : Type) :=
mk :: (x : α) (y : α)

def p : point nat :=
{x := 1, y := 2}

eval {p with y := 3}
eval {p with x := 3}
```

9.3 Inheritance

We can *extend* existing structures by adding new fields. This feature allow us to simulate a form of *inheritance*.

```
structure point (α : Type) :=
mk :: (x : α) (y : α)

inductive color
| red | green | blue

structure color_point (α : Type) extends point α :=
mk :: (c : color)
```

We can "rename" fields inherited from parent structures using the `renaming` clause.

```
structure prod (α : Type) (β : Type) :=
pair :: (pr1 : α) (pr2 : β)

-- Rename fields pr1 and pr2 to x and y respectively.
structure point3 (α : Type) extends prod α α renaming pr1→x pr2→y :=
mk :: (z : α)

check point3.x
check point3.y
check point3.z
```

In the next example, we define a structure using multiple inheritance, and then define an object using objects of the parent structures.

```
structure point (α : Type) :=
(x : α) (y : α) (z : α)

structure rgb_val :=
(red : nat) (green : nat) (blue : nat)

structure red_green_point (α : Type) extends point α, rgb_val :=
(no_blue : blue = 0)

def p   : point nat := {x := 10, y := 10, z := 20}
def rgp : red_green_point nat :=
{p with red := 200, green := 40, blue := 0, no_blue := rfl}

example : rgp^.x   = 10 := rfl
example : rgp^.red = 200 := rfl
```

10

Type Classes

We have seen that Lean's elaborator provides helpful automation, filling in information that is tedious to enter by hand. In this section we will explore a simple but powerful technical device known as *type class inference*, which provides yet another mechanism for the elaborator to supply missing information.

The notion of a *type class* originated with the *Haskell* programming language. Many of the original uses carry over, but, as we will see, the realm of interactive theorem proving raises even more possibilities for their use.

10.1 Type Classes and Instances

Any family of types can be marked as a *type class*. Then we can declare particular elements of a type class to be *instances*. These provide hints to the elaborator: any time the elaborator is looking for an element of a type class, it can consult a table of declared instances to find a suitable element.

More precisely, there are three steps involved:

- First, we declare a family of inductive types to be a type class.
- Second, we declare instances of the type class.
- Finally, we mark some implicit arguments with square brackets instead of curly brackets, to inform the elaborator that these arguments should be inferred by the type class mechanism.

Here is a somewhat frivolous example:

```
attribute [class] nat

instance nat_one : ℕ := 1
/- The command instance is syntax sugar for
def nat_one : ℕ := 1
attribute [instance, reducible] nat_one
-/

def foo [x : ℕ] : nat := x

check @foo
eval foo

example : foo = 1 := rfl
```

Here we declare `nat` to be a class with a "canonical" instance `1`. Then we declare `foo` to be, essentially, the identity function on the natural numbers, but we mark the argument implicit, and indicate that it should be inferred by type class inference. When we write `foo`, the preprocessor interprets it as `foo ?x`, where `?x` is an implicit argument. But when the elaborator gets hold of the expression, it sees that `?x : ℕ` is supposed to be solved by type class inference. It looks for a suitable element of the class, and it finds the instance `one`. Thus, when we evaluate `foo`, we simply get `1`.

It is tempting to think of `foo` as defined to be equal to `1`, but that is misleading. Every time we write `foo`, the elaborator searches for a value. If we declare other instances of the class, that can change the value that is assigned to the implicit argument. This can result in seemingly paradoxical behavior. For example, we might continue the development above as follows:

```
instance nat_two : ℕ := 2

eval foo

example : foo ≠ 1 :=
λ h : 2 = 1, nat.no_confusion h (λ h : 1 = 0, nat.no_confusion h)
```

Now the "same" expression `foo` evaluates to `2`. Whereas before we could prove `foo = 1`, now we can prove `foo ≠ 1`, because the inferred implicit argument has changed. When searching for a suitable instance of a type class, the elaborator tries the most recent instance declaration first, by default. We will see below, however, that it is possible to give individual instances higher or lower priority.

As with other attributes, you can assign the `class` or `instance` attributes in a definition, or after the fact, with an `attribute` command. As usual, the assignments `attribute [class] foo` and `attribute [instance] foo`. To limit the scope of an assignment to the current file, use the `local attribute` variant.

The reason the example is frivolous is that there is rarely a need to "infer" a natural number; we can just hard-code the choice of `1` or `2` into the definition of `foo`. Type classes become useful when they depend on parameters, in which case, the value that is inferred depends on these parameters.

Let us work through a simple example. Many theorems hold under the additional assumption that a type is inhabited, which is to say, it has at least one element. For example, if α is a type, `∃ x : α, x = x` is true only if α is inhabited. Similarly, it often happens that we would like a definition to return a default element in a "corner case." For example, we would like the expression `head l` to be of type α when `l` is of type `list α`; but then we are faced with the problem that `head l` needs to return an "arbitrary" element of α in the case where `l` is the empty list, `nil`.

For purposes like this, the standard library defines a type class `inhabited : Type → Type`, to enable type class inference to infer a "default" or "arbitrary" element of an inhabited type. We will carry out a similar development in the examples that follow, using a namespace `hide` to avoid conflicting with the definitions in the standard library.

Let us start with the first step of the program above, declaring an appropriate class:

```
class inhabited (α : Type) :=
(value : α)
/- The command 'class' above is shorthand for

@[class] structure inhabited (α : Type) :=
(value : α)
-/
```

An element of the class `inhabited α` is simply an expression of the form `inhabited.mk a`, for some element `a : α`. The projection `inhabited.value` will allow us to "extract" such an element of α from an element of `inhabited α`.

The second step of the program is to populate the class with some instances:

```
instance Prop_inhabited : inhabited Prop :=
inhabited.mk true

instance bool_inhabited : inhabited bool :=
inhabited.mk tt

instance nat_inhabited : inhabited nat :=
inhabited.mk 0

instance unit_inhabited : inhabited unit :=
inhabited.mk ()
```

In the Lean standard library, we regularly use the anonymous constructor when defining instances. It is particularly useful when the class name is long.

```
instance Prop_inhabited : inhabited Prop :=
⟨true⟩

instance bool_inhabited : inhabited bool :=
```

```
⟨tt⟩

instance nat_inhabited : inhabited nat :=
⟨0⟩

instance unit_inhabited : inhabited unit :=
⟨()⟩
```

This arranges things so that when type class inference is asked to infer an element `?M : Prop`, it can find the element `true` to assign to `?M`, and similarly for the elements `tt`, `0`, and `()` of the types `bool`, `nat`, and `unit`, respectively.

The final step of the program is to define a function that infers an element `s : inhabited α` and puts it to good use. The following function simply extracts the corresponding element `a : α`:

```
definition default (α : Type) [s : inhabited α] : α :=
@inhabited.value α s
```

This has the effect that given a type expression α, whenever we write `default α`, we are really writing `default α ?s`, leaving the elaborator to find a suitable value for the metavariable `?s`. When the elaborator succeeds in finding such a value, it has effectively produced an element of type α, as though by magic.

```
check default Prop    -- Prop
check default nat     -- ℕ
check default bool    -- bool
check default unit    -- unit
```

In general, whenever we write `default α`, we are asking the elaborator to synthesize an element of type α.

Notice that we can "see" the value that is synthesized with `eval`:

```
eval default Prop    -- true
eval default nat     -- 0
eval default bool    -- tt
eval default unit    -- ()
```

Sometimes we want to think of the default element of a type as being an *arbitrary* element, whose specific value should not play a role in our proofs. For that purpose, we can write `arbitrary α` instead of `default α`. The definition of `arbitrary` is the same as that of default, but is marked `irreducible` to discourage the elaborator from unfolding it. This does not preclude proofs from making use of the value, however, so the use of `arbitrary` rather than `default` functions primarily to signal intent.

10.2 Chaining Instances

If that were the extent of type class inference, it would not be all the impressive; it would be simply a mechanism of storing a list of instances for the elaborator to find in a lookup table. What makes type class inference powerful is that one can *chain* instances. That is, an instance declaration can in turn depend on an implicit instance of a type class. This causes class inference to chain through instances recursively, backtracking when necessary, in a Prolog-like search.

For example, the following definition shows that if two types α and β are inhabited, then so is their product:

```
instance prod_inhabited {α β : Type} [inhabited α] [inhabited β]
                        : inhabited (prod α β) :=
⟨(default α, default β)⟩
```

With this added to the earlier instance declarations, type class instance can infer, for example, a default element of `nat × bool × unit`:

```
check default (nat × bool)
eval default (nat × bool)
```

Given the expression `default (nat × bool)`, the elaborator is called on to infer an implicit argument `?M : inhabited (nat × bool)`. The instance `prod_inhabited` reduces this to inferring `?M1 : inhabited nat` and `?M2 : inhabited bool`. The first one is solved by the instance `nat_inhabited`. The second uses `bool_inhabited`.

Similarly, we can inhabit function spaces with suitable constant functions:

```
instance inhabited_fun (α : Type) {β : Type} [inhabited β] : inhabited (α → β) :=
⟨(λ a : α, default β)⟩

check default (nat → nat × bool)
eval default (nat → nat × bool)
```

In this case, type class inference finds the default element `λ (a : nat), (0, tt)`.

As an exercise, try defining default instances for other types, such as sum types and the list type.

10.3 Decidable Propositions

Let us consider another example of a type class defined in the standard library, namely the type class of `decidable` propositions. Roughly speaking, an element of `Prop` is said to be decidable if we can decide whether it is true or false. The distinction is only useful in constructive mathematics; classically, every proposition is decidable. Nonetheless, as we will see, the implementation of the type class allows for a smooth transition between constructive and classical logic.

In the standard library, `decidable` is defined formally as follows:

```
class inductive decidable (p : Prop) : Type
| is_false : ¬p → decidable
| is_true  :  p → decidable
```

Logically speaking, having an element `t : decidable p` is stronger than having an element `t : p ∨ ¬p`; it enables us to define values of an arbitrary type depending on the truth value of `p`. For example, for the expression `if p then a else b` to make sense, we need to know that `p` is decidable. That expression is syntactic sugar for `ite p a b`, where `ite` is defined as follows:

```
def ite (c : Prop) [d : decidable c] {α : Type} (t e : α) : α :=
decidable.rec_on d (λ hnc, e) (λ hc, t)
```

The standard library also contains a variant of `ite` called `dite`, the dependent if-then-else expression. It is defined as follows:

```
def dite (c : Prop) [d : decidable c] {α : Type} (t : c → α) (e : ¬ c → α) : α :=
decidable.rec_on d (λ hnc : ¬ c, e hnc) (λ hc : c, t hc)
```

That is, in `dite c t e`, we can assume `hc : c` in the "then" branch, and `hnc : ¬ c` in the "else" branch. To make `dite` more convenient to use, Lean allows us to write `if h : c then t else e` instead of `dite c (λ h : c, t) (λ h : ¬ c, e)`.

In the standard library, we cannot prove that every proposition is decidable. But we can prove that *certain* propositions are decidable. For example, we can prove that basic operations like equality and comparisons on the natural numbers and the integers are decidable. Moreover, decidability is preserved under propositional connectives:

```
check @and.decidable
-- Π {p q : Prop} [hp : decidable p] [hq : decidable q], decidable (p ∧ q)

check @or.decidable
check @not.decidable
check @implies.decidable
```

Thus we can carry out definitions by cases on decidable predicates on the natural numbers:

```
open nat

def step (a b x : ℕ) : ℕ :=
if x < a ∨ x > b then 0 else 1
```

```
set_option pp.implicit true
print definition step
```

Turning on implicit arguments shows that the elaborator has inferred the decidability of the proposition `x < a ∨ x > b`, simply by applying appropriate instances.

With the classical axioms, we can prove that every proposition is decidable. When you import the classical axioms, then, `decidable p` has an instance for every `p`, and the elaborator infers that value quickly. Thus all theorems in the standard library that rely on decidability assumptions are freely available in the classical library.

10.4 Overloading with Type Classes

We now consider the application of type classes that motivates their use in functional programming languages like Haskell, namely, to overload notation in a principled way. In Lean, a symbol like `+` can be given entirely unrelated meanings, a phenomenon that is sometimes called "ad-hoc" overloading. Typically, however, we use the `+` symbol to denote a binary function from a type to itself, that is, a function of type $\alpha \to \alpha \to \alpha$ for some type α. We can use type classes to infer an appropriate addition function for suitable types α. We will see in the next section that this is especially useful for developing algebraic hierarchies of structures in a formal setting.

We can declare a type class `has_add α` as follows:

```
universe variables u

class has_add (α : Type u) :=
(add : α → α → α)

def add {α : Type u} [has_add α] : α → α → α := has_add.add

local notation a `+` b := add a b
```

The class `has_add α` is supposed to be inhabited exactly when there is an appropriate addition function for α. The `add` function is designed to find an instance of `has_add α` for the given type, α, and apply the corresponding binary addition function. The notation `a + b` thus refers to the addition that is appropriate to the type of `a` and `b`. We can the declare instances for `nat`, and `bool`:

```
instance nat_has_add : has_add nat :=
⟨nat.add⟩

instance bool_has_add : has_add bool :=
⟨bor⟩

check 2 + 2    -- nat
check tt + ff  -- bool
```

As with **inhabited** and **decidable**, the power of type class inference stems not only from the fact that the class enables the elaborator to look up appropriate instances, but also from the fact that it can chain instances to infer complex addition operations. For example, assuming that there are appropriate addition functions for types α and β, we can define addition on $\alpha \times \beta$ pointwise:

```
instance prod_has_add {α : Type u} {β : Type v} [has_add α] [has_add β] : has_add (α × β) :=
⟨λ ⟨a₁, b₁⟩ ⟨a₂, b₂⟩, ⟨a₁+a₂, b₁+b₂⟩⟩

check (1, 2) + (3, 4)    -- ℕ × ℕ
eval  (1, 2) + (3, 4)    -- (4, 6)
```

We can similarly define pointwise addition of functions:

```
instance fun_has_add {α : Type u} {β : Type v} [has_add β] : has_add (α → β) :=
⟨λ f g x, f x + g x⟩

check (λ x : nat, 1) + (λ x, 2)   -- ℕ → ℕ
eval (λ x : nat, 1) + (λ x, 2)    -- λ (x : ℕ), 3
```

As an exercise, try defining instances of `has_add` for lists, and show that they have the work as expected.

10.5 Managing Type Class Inference

You can ask Lean for information about the classes and instances that are currently in scope:

```
print classes
print instances inhabited
```

At times, you may find that the type class inference fails to find an expected instance, or, worse, falls into an infinite loop and times out. To help debug in these situations, Lean enables you to request a trace of the search:

```
set_option trace.class_instances true
```

If you add this to your file in Emacs mode and use `C-c C-x` to run an independent Lean process on your file, the output buffer will show a trace every time the type class resolution procedure is subsequently triggered.

You can also limit the search depth (the default is 32):

```
set_option class.instance_max_depth 5
```

Remember also that in the Emacs Lean mode, tab completion works in `set_option`, to help you find suitable options.

As noted above, the type class instances in a given context represent a Prolog-like program, which gives rise to a backtracking search. Both the efficiency of the program and the solutions that are found can depend on the order in which the system tries the instance. Instances which are declared last are tried first. Moreover, if instances are declared in other modules, the order in which they are tried depends on the order in which namespaces are opened. Instances declared in namespaces which are opened later are tried earlier.

You can change the order that type classes instances are tried by assigning them a *priority*. When an instance is declared, it is assigned a priority value `std.priority.default`, defined to be 1000 in module `init.core` in the standard library. You can assign other priorities when defining an instance, and you can later change the priority with the `attribute` command. The following example illustrates how this is done:

```
class foo :=
(a : nat) (b : nat)

@[priority std.priority.default+1]
instance i1 : foo :=
⟨1, 1⟩

instance i2 : foo :=
⟨2, 2⟩

example : foo.a = 1 := rfl

@[priority std.priority.default+20]
instance i3 : foo :=
⟨3, 3⟩

example : foo.a = 3 := rfl

attribute [instance, priority 10] i3

example : foo.a = 1 := rfl

attribute [instance, priority std.priority.default-10] i1

example : foo.a = 2 := rfl
```

10.6 Coercions using Type Classes

The most basic type of coercion maps elements of one type to another. For example, a coercion from `nat` to `int` allows us to view any element `n : nat` as an element of `int`. But some coercions depend on parameters; for example, for any type α, we can view any element `l : list` α as an element of `set` α, namely, the set of elements occurring in the list. The corresponding coercion is defined on the "family" of types `list α`, parameterized by α.

Lean allows us to declare three kinds of coercions:

- from a family of types to another family of types
- from a family of types to the class of sorts
- from a family of types to the class of function types

The first kind of coercion allows us to view any element of a member of the source family as an element of a corresponding member of the target family. The second kind of coercion allows us to view any element of a member of the source family as a type. The third kind of coercion allows us to view any element of the source family as a function. Let us consider each of these in turn.

In Lean, coercions are implemented on top of the type class resolution framework. We define a coercion from α to β by declaring an instance of `has_coe α β`. For example, we can define a coercion from `bool` to `Prop` as follows:

```
instance bool_to_Prop : has_coe bool Prop :=
⟨λ b, b = tt⟩
-- Now, we can use bool terms in if-then-else terms
vm_eval if tt then "hello" else "world"
vm_eval if ff then "hello" else "world"
```

We can define a coercion from `list α` to `set α` as follows:

```
universe variable u

def list.to_set {α : Type u} : list α → set α
| []     := ∅
| (h::t) := {h} ∪ list.to_set t

instance list_to_set_coe (α : Type u) : has_coe (list α) (set α) :=
⟨list.to_set⟩

def s : set nat  := {1, 2}
def l : list nat := [3, 4]

check s ∪ l -- set nat
```

Coercions are only considered if the given and expected types do not contain metavariables at elaboration time. In the following example, when we elaborate the union operator, the type of `[3, 2]` is `list ?m`, and a coercion will not be considered since it contains metavariables.

```
/- The following check command produces an error. -/
-- check s ∪ [3, 2]
```

We can workaround this issue by using a type ascription.

```
check s ∪ [(3:nat), 2]
-- or
check s ∪ ([3, 2] : list nat)
```

In the examples above, you may have noticed the symbol ↑ produced by the `check` commands. It is the lift operator, `↑t` is notation for `coe t`. We can use this operator to force a coercion to be introduced in a particular place. It is also helpful to make our intent clear, and workaround limitations of the coercion resolution system.

```
check s ∪ ↑[3, 2]

variables n m : nat
variable i : int
check i + ↑n + ↑m
check i + ↑(n + m)

/- In the above two examples, the coercions are not
   strictly necessary since Lean will insert implicit
   nat → int coercions. However, the following example
   doesn't work because the expected type of i is a nat
   to match the type of n (and no int → nat coercion
   exists). -/
-- check n + i

/- Assuming we want the more general type, we can
   insert an explicit ↑ to coerce n to int. -/
check ↑n + i
```

The standard library defines a coercion from subtype `{x : α // p x}` to α as follows:

```
instance coe_subtype {α : Type u} {p : α → Prop} : has_coe {x // p x} α :=
⟨λ s, subtype.elt_of s⟩
```

Lean will also chain coercions as necessary. Actually, the type class `has_coe_t` is the transitive closure of `has_coe`. You may have noticed that the type of `coe` depends on `has_lift_t`, the transitive closure of the type class `has_lift`, instead of `has_coe_t`. Every instance of `has_coe_t` is also an instance of `has_lift_t`, but the elaborator only introduces automatically instances of `has_coe_t`. That is, to be able to coerce using an instance of `has_lift_t`, we must use the operator ↑. In the standard library, we have the following instance:

```
namespace hide
universe variables u v

instance lift_list {a : Type u} {b : Type v} [has_lift_t a b] : has_lift (list a) (list b) :=
⟨λ l, list.map (@coe a b _) l⟩
```

```
variables s : list nat
variables r : list int
check ↑s ++ r

end hide
```

It is not an instance of `has_coe` because lists are frequently used for writing programs, and we do not want a linear-time operation to be silently introduced by Lean, and potentially mask mistakes performed by the user. By forcing the user to write ↑, she is making her intent clear to Lean.

Let us now consider the second kind of coercion. By the *class of sorts*, we mean the collection of universes `Type u`. A coercion of the second kind is of the form

```
c : Π x1 : A1, ..., xn : An, F x1 ... xn → Type u
```

where `F` is a family of types as above. This allows us to write `s : t` whenever `t` is of type `F a1 ... an`. In other words, the coercion allows us to view the elements of `F a1 ... an` as types. This is very useful when defining algebraic structures in which one component, the carrier of the structure, is a `Type`. For example, we can define a semigroup as follows:

```
universe variable u

structure Semigroup : Type (u+1) :=
(carrier : Type u)
(mul : carrier → carrier → carrier)
(mul_assoc : ∀ a b c : carrier, mul (mul a b) c = mul a (mul b c))

instance Semigroup_has_mul (S : Semigroup) : has_mul (S^.carrier) :=
⟨S^.mul⟩
```

In other words, a semigroup consists of a type, `carrier`, and a multiplication, `mul`, with the property that the multiplication is associative. The `instance` command allows us to write `a * b` instead of `Semigroup.mul S a b` whenever we have `a b : S^.carrier`; notice that Lean can infer the argument `S` from the types of `a` and `b`. The function `Semigroup.carrier` maps the class `Semigroup` to the sort `Type u`:

```
check Semigroup.carrier
```

If we declare this function to be a coercion, then whenever we have a semigroup `S : Semigroup`, we can write `a : S` instead of `a : S^.carrier`:

```
instance Semigroup_to_sort : has_coe_to_sort Semigroup :=
{S := Type u, coe := λ S, S^.carrier}
```

```
example (S : Semigroup) (a b c : S) : (a * b) * c = a * (b * c) :=
Semigroup.mul_assoc _ a b c
```

It is the coercion that makes it possible to write `(a b c : S)`. Note that, we define an instance of `has_coe_to_sort Semigroup` instead of `has_coe Semigroup Type`. The reason is that when Lean needs a coercion to sort, it only knows it needs a type, but, in general, the universe is not known. The field `S` in the class `has_coe_to_sort` is used to specify the universe we are coercing too.

By the *class of function types*, we mean the collection of Pi types `Π z : B, C`. The third kind of coercion has the form

```
c : Π x1 : A1, ..., xn : An, y : F x1 ... xn, Π z : B, C
```

where `F` is again a family of types and `B` and `C` can depend on `x1, ..., xn, y`. This makes it possible to write `t s` whenever `t` is an element of `F a1 ... an`. In other words, the coercion enables us to view elements of `F a1 ... an` as functions. Continuing the example above, we can define the notion of a morphism between semigroups `S1` and `S2`. That is, a function from the carrier of `S1` to the carrier of `S2` (note the implicit coercion) that respects the multiplication. The projection `morphism.mor` takes a morphism to the underlying function:

```
instance Semigroup_to_sort : has_coe_to_sort Semigroup :=
{S := _, coe := λ S, S^.carrier}

structure morphism (S1 S2 : Semigroup) :=
(mor : S1 → S2)
(resp_mul : ∀ a b : S1, mor (a * b) = (mor a) * (mor b))

check @morphism.mor
```

As a result, it is a prime candidate for the third type of coercion.

```
instance morphism_to_fun (S1 S2 : Semigroup) : has_coe_to_fun (morphism S1 S2) :=
{ F   := λ _, S1 → S2,
  coe := λ m, m^.mor }

lemma resp_mul {S1 S2 : Semigroup} (f : morphism S1 S2) (a b : S1) : f (a * b) = f a * f b :=
f^.resp_mul a b

example (S1 S2 : Semigroup) (f : morphism S1 S2) (a : S1) :
  f (a * a * a) = f a * f a * f a :=
calc
  f (a * a * a) = f (a * a) * f a : by rw [resp_mul f]
            ... = f a * f a * f a : by rw [resp_mul f]
```

With the coercion in place, we can write `f (a * a * a)` instead of `morphism.mor f (a * a * a)`. When the `morphism`, `f`, is used where a function is expected, Lean inserts the coercion. Similar to `has_coe_to_sort`, we have yet another class `has_coe_to_fun` for the this class of coercions. The field `F` is used to specify function type we are coercing too. This type may depend on the type we are coercing from.

Finally, `f` and `S` are notations for `coe_fn f` and `coe_sort S`. They are the coercion operators for the function and sort classes.

We can instruct Lean's pretty-printer to hide the operators ↑ and with `set_option`.

```
theorem test (S1 S2 : Semigroup) (f : morphism S1 S2) (a : S1) :
  f (a * a * a) = f a * f a * f a :=
calc
  f (a * a * a) = f (a * a) * f a : by rw [resp_mul f]
            ... = f a * f a * f a : by rw [resp_mul f]

check @test
set_option pp.coercions false
check @test
```

11

Axioms and Computation

We have seen that the version of the Calculus of Inductive Constructions that has been implemented in Lean includes dependent function types, inductive types, and a hierarchy of universes that starts with an impredicative, proof-irrelevant `Prop` at the bottom. In this chapter, we consider ways of extending the CIC with additional axioms and rules. Extending a foundational system in such a way is often convenient; it can make it possible to prove more theorems, as well as make it easier to prove theorems that could have been proved otherwise. But there can be negative consequences of adding additional axioms, consequences which may go beyond concerns about their correctness. In particular, the use of axioms bears on the computational content of definitions and theorems, in ways we will explore here.

Lean is designed to support both computational and classical reasoning. Users that are so inclined can stick to a "computationally pure" fragment, which guarantees that closed expressions in the system evaluate to canonical normal forms. In particular, any closed computationally pure expression of type ℕ, for example, will reduce to a numeral.

Lean's standard library defines an additional axiom, propositional extensionality, and a quotient construction which in turn implies the principle of function extensionality. These extensions are used, for example, to develop theories of sets and finite sets. We will see below that using these theorems can block evaluation in Lean's kernel, so that closed terms of type ℕ no longer evaluate to numerals. But Lean erases types and propositional information when compiling definitions to bytecode for its virtual machine evaluator, and since these axioms only add new propositions, they are compatible with that computational interpretation. Even computationally inclined users may wish to use the classical law of the excluded middle to reason about computation. This also blocks evaluation in the kernel, but it is compatible with compilation to bytecode.

The standard library also defines a choice principle that is entirely antithetical to a computational interpretation, since it magically produces "data" from a proposition asserting its existence. Its use is essential to some classical constructions, and users can import it when needed. But expressions that use this construction to produce data do not have computational content, and in Lean we are required to mark such definitions as `noncomputable` to flag that fact.

Using a clever trick (known as Diaconescu's theorem), one can use propositional extensionality, function extensionality, and choice to derive the law of the excluded middle. As noted above, however, use of the law of the excluded middle is still compatible with bytecode compilation and code extraction, as are other classical principles, as long as they are not used to manufacture data.

To summarize, then, on top of the underlying framework of universes, dependent function types, and inductive types, the standard library adds three additional components:

- the axiom of propositional extensionality
- a quotient construction, which implies function extensionality
- a choice principle, which produces data from an existential proposition.

The first two of these block normalization within Lean, but are compatible with bytecode evaluation, whereas the third is not amenable to computational interpretation. We will spell out the details more precisely below.

11.1 Historical and Philosophical Context

For most of its history, mathematics was essentially computational: geometry dealt with constructions of geometric objects, algebra was concerned with algorithmic solutions to systems of equations, and analysis provided means to compute the future behavior of systems evolving over time. From the proof of a theorem to the effect that "for every `x`, there is a `y` such that …", it was generally straightforward to extract an algorithm to compute such a `y` given `x`.

In the nineteenth century, however, increases in the complexity of mathematical arguments pushed mathematicians to develop new styles of reasoning that suppress algorithmic information and invoke descriptions of mathematical objects that abstract away the details of how those objects are represented. The goal was to obtain a powerful "conceptual" understanding without getting bogged down in computational details, but this had the effect of admitting mathematical theorems that are simply *false* on a direct computational reading.

There is still fairly uniform agreement today that computation is important to mathematics. But there are different views as to how best to address computational concerns. From a *constructive* point of view, it is a mistake to separate mathematics from its computational roots; every meaningful mathematical theorem should have a direct computational interpretation. From a *classical* point of view, it is more fruitful to maintain a separation of concerns: we can use one language and body of methods to write computer programs, while maintaining the freedom to use a nonconstructive theories and methods to reason about them. Lean is designed to support both of these approaches. Core parts of the library are developed constructively, but the system also provides support for carrying out classical mathematical reasoning.

Computationally, the purest part of dependent type theory avoids the use of `Prop` entirely. Inductive types and dependent function types can be viewed as data types, and terms of these types can be "evaluated" by applying reduction rules until no more rules can be applied. In principle, any closed term (that is, term with no free variables) of type $\mathbb{N}$ should evaluate to a numeral, `succ (... (succ zero)...)`.

Introducing a proof-irrelevant `Prop` and marking theorems irreducible represents a first step towards separation of concerns. The intention is that elements of a type `p : Prop` should play no role in computation, and so the particular construction of a term `t : p` is "irrelevant" in that sense. One can still define computational objects that incorporate elements of type `Prop`; the point is that these elements can help us reason about the effects of the computation, but can be ignored when we extract "code" from the term. Elements of type `Prop` are not entirely innocuous, however. They include equations `s = t :` α for any type α, and such equations can be used as casts, to type check terms. Below, we will see examples of how such casts can block computation in the system. However, computation is still possible under an evaluation scheme that erases propositional content, ignores intermediate typing constraints, and reduces terms until they reach a normal form. This is precisely what Lean's virtual machine does.

Having adopted a proof-irrelevant `Prop`, one might consider it legitimate to use, for example, the law of the excluded middle, `p ∨ ¬p`, where `p` is any proposition. Of course, this, too, can block computation according to the rules of CIC, but it does not block bytecode evaluation, as described above. It is only the choice principles discussed in Section 11.5 that completely erasing the distinction between the proof-irrelevant and data-relevant parts of the theory.

11.2 Propositional Extensionality

Propositional extensionality is the following axiom:

```
axiom propext {a b : Prop} : (a ↔ b) → a = b
```

It asserts that when two propositions imply one another, they are actually equal. This is consistent with set-theoretic interpretations in which any element `a : Prop` is either empty or the singleton set `{*}`, for some distinguished element `*`. The axiom has the effect that equivalent propositions can be substituted for one another in any context:

```
section
  variables a b c d e : Prop
  variable p : Prop → Prop

  example (h : a ↔ b) : (c ∧ a ∧ d → e) ↔ (c ∧ b ∧ d → e) :=
  propext h ▸ iff.refl _

  example (h : a ↔ b) (h₁ : p a) : p b :=
  propext h ▸ h₁
end
```

The first example could be proved more laboriously without `propext` using the fact that the propositional connectives respect propositional equivalence. The second example represents a more essential use of `propext`. In fact, it is equivalent to `propext` itself, a fact which we encourage you to prove.

Given any definition or theorem in Lean, you can use the `print axioms` command to display the axioms it depends on.

```
print axioms thm₁  -- propext
print axioms thm₂  -- propext
```

11.3 Function Extensionality

Similar to propositional extensionality, function extensionality asserts that any two functions of type `Π x : α, β x` that agree on all their inputs are equal.

```
universes u₁ u₂

check (@funext : ∀ {α : Type u₁} {β : α → Type u₂} {f₁ f₂ : Π (x : α), β x},
        (∀ (x : α), f₁ x = f₂ x) → f₁ = f₂)
```

From a classical, set-theoretic perspective, this is exactly what it means for two functions to be equal. This is known as an "extensional" view of functions. From a constructive perspective, however, it is sometimes more natural to think of functions as algorithms, or computer programs, that are presented in some explicit way. It is certainly the case that two computer programs can compute the same answer for every input despite the fact that they are syntactically quite different. In much the same way, you might want to maintain a view of functions that does not force you to identify two functions that have the same input / output behavior. This is known as an "intensional" view of functions.

In fact, function extensionality follows from the existence of quotients, which we describe in the next section. In the Lean standard library, therefore, `funext` is thus proved from the quotient construction.

Suppose that for `α : Type` we define the `set α := α → Prop` to denote the type of subsets of α, essentially identifying subsets with predicates. By combining `funext` and `propext`, we obtain an extensional theory of such sets:

```
universe u

def set (α : Type u) := α → Prop

namespace set

variable {α : Type u}

definition mem (x : α) (a : set α) := a x
notation e ∈ a := mem e a

theorem setext {a b : set α} (h : ∀ x, x ∈ a ↔ x ∈ b) : a = b :=
funext (take x, propext (h x))

end set
```

We can then proceed to define the empty set and set intersection, for example, and prove set identities:

```
definition empty : set α := λ x, false
local notation `∅` := empty

definition inter (a b : set α) : set α := λ x, x ∈ a ∧ x ∈ b
notation a ∩ b := inter a b

theorem inter_self (a : set α) : a ∩ a = a :=
setext (take x, and_self _)

theorem inter_empty (a : set α) : a ∩ ∅ = ∅ :=
setext (take x, and_false _)

theorem empty_inter (a : set α) : ∅ ∩ a = ∅ :=
setext (take x, false_and _)

theorem inter.comm (a b : set α) : a ∩ b = b ∩ a :=
setext (take x, and_comm _ _)
```

The following is an example of how function extensionality blocks computation inside the Lean kernel.

```
def f₁  (x : ℕ) := x
def f₂ (x : ℕ) := 0 + x

theorem feq : f₁ = f₂ := funext (take x, (zero_add x)^.symm)

def val : ℕ := eq.rec_on feq (0 : ℕ)

-- complicated!
```

```
eval val

-- evaluates to 0
vm_eval val
```

First, we show that the two functions f_1 and f_2 are equal using function extensionality, and then we cast `0` of type ℕ by replacing f_1 by f_2 in the type. Of course, the cast is vacuous, because ℕ does not depend on f_1. But that is enough to do the damage: under the computational rules of the system, we now have a closed term of ℕ that does not reduce to a numeral. In this case, we may be tempted to reduce the expression to `0`. But in nontrivial examples, eliminating cast changes the type of the term, which might make an ambient expression type incorrect. The virtual machine, however, has no trouble evaluating the expression to `0`. Here is a similarly contrived example that shows how `propext` can get in the way.

```
theorem tteq : (true ∧ true) = true := propext (and_true true)

def val : ℕ := eq.rec_on tteq 0

-- complicated!
eval val

-- evaluates to 0
vm_eval val
```

Current research programs, including work on *observational type theory* and *cubical type theory*, aim to extend type theory in ways that permit reductions for casts involving function extensionality, quotients, and more. But the solutions are not so clear cut, and the rules of Lean's underlying calculus do not sanction such reductions.

In a sense, however, a cast does not change the meaning of an expression. Rather, it is a mechanism to reason about the expression's type. Given an appropriate semantics, it then makes sense to reduce terms in ways that preserve their meaning, ignoring the intermediate bookkeeping needed to make the reductions type correct. In that case, adding new axioms in `Prop` does not matter; by proof irrelevance, an expression in `Prop` carries no information, and can be safely ignored by the reduction procedures.

11.4 Quotients

Let α be any type, and let `r` be an equivalence relation on α. It is mathematically common to form the "quotient" α `/ r`, that is, the type of elements of α "modulo" `r`. Set theoretically, one can view α `/ r` as the set of equivalence classes of α modulo `r`. If `f :` $\alpha \to \beta$ is any function that respects the equivalence relation in the sense that for every `x y :` α, `r x y` implies `f x = f y`, then `f` "lifts" to a function `f' :` α `/ r` $\to \beta$ defined on each equivalence class `⟦x⟧` by `f' ⟦x⟧ = f x`. Lean's standard library extends the

Calculus of Inductive Constructions with additional constants that perform exactly these constructions, and installs this last equation as a definitional reduction rule.

In its most basic form, the quotient construction does not even require `r` to be an equivalence relation. The following constants are built into Lean:

```
universes u v

constant quot      : Π {α : Sort u}, (α → α → Prop) → Sort u
constant quot.mk   : Π {α : Sort u} (r : α → α → Prop), α → quot r

axiom    quot.ind  : ∀ {α : Sort u} {r : α → α → Prop} {β : quot r → Prop},
                      (∀ a, β (quot.mk r a)) → ∀ (q : quot r), β q

constant quot.lift : Π {α : Sort u} {r : α → α → Prop} {β : Sort u} (f : α → β),
                      (∀ a b, r a b → f a = f b) → quot r → β
```

The first one forms a type `quot r` given a type α by any binary relation `r` on α. The second maps α to `quot` α, so that for any `a :` α, `quot.mk a` is an element of `quot r`. The third principle, `quot.ind`, says that every element of `quot.mk a` is of this form. Given any function `f` and a proof `h` that respects the relation `r`, `quot.lift f h` is the corresponding function on `quot r`. The idea is that for any element `a` in α, `quot.lift f h` is the function which maps `quot.mk r a` to `f a`, wherein `h` shows that this fuction is well defined. In fact, the computation principle is declared as a reduction rule, as the proof below makes clear.

```
variables α β : Type
variable  r : α → α → Prop
variable  a : α

-- the quotient type
check (quot r : Type)

-- the class of a
check (quot.mk r a : quot r)

variable  f : α → β
premise   h : ∀ a₁ a₂, r a₁ a₂ → f a₁ = f a₂

-- the corresponding function on quot r
check (quot.lift f h : quot r → β)

-- the computation principle
theorem thm : quot.lift f h (quot.mk r a) = f a := rfl
```

The four constants, `quot`, `quot.mk`, `quot.ind`, and `quot.lift` in and of themselves are not very strong. You can check that the `quot.ind` is satisfied if we take `quot r` to be simply α, and take `quot.lift` to be the identity function (ignoring `h`). For that reason, these four constants are not viewed as additional axioms:

```
print axioms thm   -- no axioms
```

They are, like inductively defined types and the associated constructors and recursors, viewed as part of the logical framework.

What makes the quot construction into a bona fide quotient is the following additional axiom:

```
axiom quot.sound : ∀ {α : Type u} {r : α → α → Prop} {a b : α},
                   r a b → quot.mk r a = quot.mk r b
```

This is the axiom that asserts that any two element of α that are related by `r` become identified in the quotient. If a theorem or definition makes use of `quot.sound`, it will show up in the `print axioms` command.

Of course, the quotient construction is most commonly used in situations when `r` is an equivalence relation. Given `r` as above, it is not hard to see that the relation `r' a b` defined by `quot.mk r a = quot.mk r b` is an equivalence relation. The axiom `quot.sound` says that `r a b` implies `r' a b`. Using `quot.lift` and `quot.ind`, we can show that the latter is the smallest equivalence relation containing `r`, in the sense that if `r'' a b` is any equivalence relation containing `r`, the `r' a b` implies `r'' a b`. In particular, if `r` was an equivalence relation to start with, then `r a b` holds iff `r' a b` for every `a` and `b`.

To support this common use case, the standard library defines the notion of a *setoid*, which is simply a type with an associated equivalence relation:

```
class setoid (α : Type u) :=
(r : α → α → Prop) (iseqv : equivalence r)

namespace setoid
  infix `≈` := setoid.r

  variable {α : Type u}
  variable [s : setoid α]
  include s

  theorem refl (a : α) : a ≈ a :=
  (@setoid.iseqv α s)^.left a

  theorem symm {a b : α} : a ≈ b → b ≈ a :=
  λ h, (@setoid.iseqv α s)^.right^.left h

  theorem trans {a b c : α} : a ≈ b → b ≈ c → a ≈ c :=
  λ h₁ h₂, (@setoid.iseqv α s)^.right^.right h₁ h₂
end setoid
```

Given a type α, a relation `r` on α, and a proof `p` that `r` is an equivalence relation, we can define `setoid.mk p` as an instance of the setoid class.

```
def quotient {α : Type u} (s : setoid α) :=
@quot α setoid.r
```

The constants `quotient.mk`, `quotient.ind`, `quotient.lift`, and `quotient.sound` are nothing more than the specializations of the corresponding elements of `quot`. The fact that type class inference can find the setoid associated to a type α brings a number of benefits. First, we can use the notation `a ≈ b` (entered with `\eq` in Emacs) for `setoid.r a b`, where the instance of `setoid` is implicit in the notation `setoid.r`. We can use the generic theorems `setoid.refl`, `setoid.symm`, `setoid.trans` to reason about the relation. Specifically with quotients we can use the generic notation `⟦a⟧` for `quot.mk setoid.r` where the instance of `setoid` is implicit in the notation `setoid.r`, as well as the theorem `quotient.exact`:

```
check (@quotient.exact : ∀ {α : Type u} [setoid α] {a b : α}, ⟦a⟧ = ⟦b⟧ → a ≈ b)
```

Together with `quotient.sound`, this implies that the elements of the quotient correspond exactly to the equivalence classes of elements in α.

Recall that in the standard library, $\alpha \times \beta$ represents the Cartesian product of the types α and β. To illustrate the use of quotients, let us define the type of *unordered* pairs of elements of a type α as a quotient of the type $\alpha \times \alpha$. First, we define the relevant equivalence relation:

```
universe u

private definition eqv {α : Type u} (p₁ p₂ : α × α) : Prop :=
(p₁.1 = p₂.1 ∧ p₁.2 = p₂.2) ∨ (p₁.1 = p₂.2 ∧ p₁.2 = p₂.1)

infix `~` := eqv
```

The next step is to prove that `eqv` is in fact an equivalence relation, which is to say, it is reflexive, symmetric and transitive. We can prove these three facts in a convenient and readable way by using dependent pattern matching to perform case-analysis and break the hypotheses into pieces that are then reassembled to produce the conclusion.

```
open or eq

private theorem eqv.refl {α : Type u} : ∀ p : α × α, p ~ p :=
take p, inl ⟨rfl, rfl⟩

private theorem eqv.symm {α : Type u} : ∀ p₁ p₂ : α × α, p₁ ~ p₂ → p₂ ~ p₁
| (a₁, a₂) (b₁, b₂) (inl ⟨a₁b₁, a₂b₂⟩) := inl ⟨symm a₁b₁, symm a₂b₂⟩
| (a₁, a₂) (b₁, b₂) (inr ⟨a₁b₂, a₂b₁⟩) := inr ⟨symm a₂b₁, symm a₁b₂⟩

private theorem eqv.trans {α : Type u} : ∀ p₁ p₂ p₃ : α × α, p₁ ~ p₂ → p₂ ~ p₃ → p₁ ~ p₃
| (a₁, a₂) (b₁, b₂) (c₁, c₂) (inl ⟨a₁b₁, a₂b₂⟩) (inl ⟨b₁c₁, b₂c₂⟩) :=
  inl ⟨trans a₁b₁ b₁c₁, trans a₂b₂ b₂c₂⟩
| (a₁, a₂) (b₁, b₂) (c₁, c₂) (inl ⟨a₁b₁, a₂b₂⟩) (inr ⟨b₁c₂, b₂c₁⟩) :=
  inr ⟨trans a₁b₁ b₁c₂, trans a₂b₂ b₂c₁⟩
| (a₁, a₂) (b₁, b₂) (c₁, c₂) (inr ⟨a₁b₂, a₂b₁⟩) (inl ⟨b₁c₁, b₂c₂⟩) :=
  inr ⟨trans a₁b₂ b₂c₂, trans a₂b₁ b₁c₁⟩
```

```
| (a₁, a₂) (b₁, b₂) (c₁, c₂) (inr ⟨a₁b₂, a₂b₁⟩) (inr ⟨b₁c₂, b₂c₁⟩) :=
  inl ⟨trans a₁b₂ b₂c₁, trans a₂b₁ b₁c₂⟩

private theorem is_equivalence (α : Type u) : equivalence (@eqv α) :=
mk_equivalence (@eqv α) (@eqv.refl α) (@eqv.symm α) (@eqv.trans α)
```

We open the namespaces `or` and `eq` to be able to use `or.inl`, `or.inr`, and `eq.trans` more conveniently.

Now that we have proved that `eqv` is an equivalence relation, we can construct a `setoid (α × α)`, and use it to define the type `uprod α` of unordered pairs.

```
instance uprod.setoid (α : Type u) : setoid (α × α) :=
setoid.mk (@eqv α) (is_equivalence α)

definition uprod (α : Type u) : Type u :=
quotient (uprod.setoid α)

namespace uprod
  definition mk {α : Type u} (a₁ a₂ : α) : uprod α :=
  ⟦(a₁, a₂)⟧

  local notation `{` a₁ `,` a₂ `}` := mk a₁ a₂
end uprod
```

Notice that we locally define the notation `{a₁, a₂}` for ordered pairs as `⟦(a₁, a₂)⟧`. This is useful for illustrative purposes, but it is not a good idea in general, since the notation will shadow other uses of curly brackets, such as for records and sets.

We can easily prove that `{a₁, a₂} = {a₂, a₁}` using `quot.sound`, since we have `(a₁, a₂) ~ (a₂, a₁)`.

```
  theorem mk_eq_mk {α : Type} (a₁ a₂ : α) : {a₁, a₂} = {a₂, a₁} :=
  quot.sound (inr ⟨rfl, rfl⟩)
```

To complete the example, given `a : α` and `u : uprod α`, we define the proposition `a ∈ u` which should hold if `a` is one of the elements of the unordered pair `u`. First, we define a similar proposition `mem_fn a u` on (ordered) pairs; then we show that `mem_fn` respects the equivalence relation `eqv` with the lemma `mem_respects`. This is an idiom that is used extensively in the Lean standard library.

```
  private definition mem_fn {α : Type} (a : α) : α × α → Prop
  | (a₁, a₂) := a = a₁ ∨ a = a₂

  -- auxiliary lemma for proving mem_respects
  private lemma mem_swap {α : Type} {a : α} : ∀ {p : α × α}, mem_fn a p = mem_fn a (⟨p.2, p.1⟩)
  | (a₁, a₂) := propext (iff.intro
      (λ l : a = a₁ ∨ a = a₂, or.elim l (λ h₁, inr h₁) (λ h₂, inl h₂))
      (λ r : a = a₂ ∨ a = a₁, or.elim r (λ h₁, inr h₁) (λ h₂, inl h₂)))
```

```
private lemma mem_respects {α : Type} : ∀ {p₁ p₂ : α × α} (a : α),  p₁ ~ p₂ → mem_fn a p₁ = mem_fn a p₂
| (a₁, a₂) (b₁, b₂) a (inl ⟨a₁b₁, a₂b₂⟩) :=
  begin dsimp at a₁b₁, dsimp at a₂b₂, rewrite [a₁b₁, a₂b₂] end
| (a₁, a₂) (b₁, b₂) a (inr ⟨a₁b₂, a₂b₁⟩) :=
  begin dsimp at a₁b₂, dsimp at a₂b₁, rewrite [a₁b₂, a₂b₁], apply mem_swap end

def mem {α : Type} (a : α) (u : uprod α) : Prop :=
quot.lift_on u (λ p, mem_fn a p) (λ p₁ p₂ e, mem_respects a e)

local infix `∈` := mem

theorem mem_mk_left {α : Type} (a b : α) : a ∈ {a, b} :=
inl rfl

theorem mem_mk_right {α : Type} (a b : α) : b ∈ {a, b} :=
inr rfl

theorem mem_or_mem_of_mem_mk {α : Type} {a b c : α} : c ∈ {a, b} → c = a ∨ c = b :=
λ h, h
```

For convenience, the standard library also defines `quotient.lift₂` for lifting binary functions, and `quotient.ind₂` for induction on two variables.

We close this section with some hints as to why the quotient construction implies function extenionality. It is not hard to show that extensional equality on the `Π x : α, β x` is an equivalence relation, and so we can consider the type `extfun α β` of functions "up to equivalence." Of course, application respects that equivalence in the sense that if f_1 is equivalent to f_2, then f_1 `a` is equal to f_2 `a`. Thus application gives rise to a function `extfun_app : extfun α β → Π x : α, β x`. But for every `f`, `extfun_app ⟦f⟧` is definitionally equal to `λ x, f x`, which is in turn definitionally equal to `f`. So, when f_1 and f_2 are extenionally equal, we have the following chain of equalities:

```
f₁ = extfun_app ⟦f₁⟧ = extfun_app ⟦f₂⟧ = f₂
```

As a result, f_1 is equal to f_2.

11.5 Choice

To state the final axiom defined in the standard library, we need the `nonempty` type, which is defined as follows:

```
class inductive nonempty (α : Sort u) : Prop
| intro : α → nonempty
```

Because `nonempty α` has type `Prop` and its constructor contains data, it can only eliminate to `Prop`. In fact, `nonempty α` is equivalent to `∃ x : α, true`:

```
example (α : Type u) : nonempty α ↔ ∃ x : α, true :=
iff.intro (λ ⟨a⟩, ⟨a, trivial⟩) (λ ⟨a, h⟩, ⟨a⟩)
```

Our axiom of choice is now expressed simply as follows:

```
axiom choice {α : Sort u} : nonempty α → α
```

Given only the assertion `h` that α is nonempty, `choice h` magically produces an element of α. Of course, this blocks any meaningful computation: by the interpretation of `Prop`, `h` contains no information at all as to how to find such an element.

This is found in the `classical` namespace, so the full name of the theorem is `classical.choice`. The choice principle is equivalent to the principle of *indefinite description*, which can be expressed with subtypes as follows:

```
noncomputable theorem indefinite_description {α : Sort u} (p : α → Prop) :
  (∃ x, p x) → {x // p x} :=
λ h, choice (let ⟨x, px⟩ := h in ⟨⟨x, px⟩⟩)
```

Because it depends on `choice`, Lean cannot generate bytecode for `indefinite_description`, and so requires us to mark the definition as `noncomputable`. Also in the `classical` namespace, the function `some` and the property `some_spec` decompose the two parts of the output of `indefinite_description`:

```
noncomputable def some {a : Sort u} {p : a → Prop} (h : ∃ x, p x) : a :=
subtype.elt_of (indefinite_description p h)

theorem some_spec {a : Sort u} {p : a → Prop} (h : ∃ x, p x) : p (some h) :=
subtype.has_property (indefinite_description p h)
```

The `choice` principle also erases the distinction between the property of being `nonempty` and the more constructive property of being `inhabited`:

```
noncomputable theorem inhabited_of_nonempty {α : Type u} : nonempty α → inhabited α :=
λ h, choice (let ⟨a⟩ := h in ⟨⟨a⟩⟩)
```

In the next section, we will see that `propext`, `funext`, and `choice`, taken together, imply the law of the excluded middle and the decidability of all propositions. Using those, one can strengthen the principle of indefinite description as follows:

```
check (@strong_indefinite_description :
       Π {α : Sort u} (p : α → Prop), nonempty α → {x // (∃ (y : α), p y) → p x})
```

Assuming the ambient type α is nonempty, `strong_indefinite_description p` produces an element of α satisfying `p` if there is one. The data component of this definition is conventionally known as *Hilbert's epsilon function*:

```
check (@epsilon : Π {α : Sort u} [nonempty α], (α → Prop) → α)

check (@epsilon_spec : ∀ {a : Sort u} {p : a → Prop} (hex : ∃ (y : a), p y),
        p (@epsilon _ (nonempty_of_exists hex) p))
```

11.6 The Law of the Excluded Middle

The law of the excluded middle is the following

```
check (@em : ∀ (p : Prop), p ∨ ¬p)
```

Diaconescu's theorem states that the axiom of choice is sufficient to derive the law of excluded middle. More precisely, it shows that the law of the excluded middle follows from `classical.choice`, `propext`, and `funext`. We sketch the proof that is found in the standard library.

First, we import the necessary axioms, fix a parameter, `p`, and define two predicates `U` and `V`:

```
open classical

section diaconescu
parameter  p : Prop

def U (x : Prop) : Prop := x = true ∨ p
def V (x : Prop) : Prop := x = false ∨ p

lemma exU : ∃ x, U x := ⟨true, or.inl rfl⟩
lemma exV : ∃ x, V x := ⟨false, or.inl rfl⟩

end diaconescu
```

If `p` is true, then every element of `Prop` is in both `U` and `V`. If `p` is false, then `U` is the singleton `true`, and `V` is the singleton `false`.

Next, we use `some` to choose an element from each of `U` and `V`:

```
noncomputable def u := some exU
noncomputable def v := some exV

lemma u_def : U u := some_spec exU
lemma v_def : V v := some_spec exV
```

Each of U and V is a disjunction, so u_def and v_def represent four cases. In one of these cases, u = true and v = false, and in all the other cases, p is true. Thus we have:

```
lemma not_uv_or_p : u ≠ v ∨ p :=
or.elim u_def
  (assume hut : u = true,
    or.elim v_def
      (assume hvf : v = false,
        have hne : u ≠ v, from eq.symm hvf ▶ eq.symm hut ▶ true_ne_false,
        or.inl hne)
      (assume hp : p, or.inr hp))
  (assume hp : p, or.inr hp)
```

On the other hand, if p is true, then, by function extensionality and propositional extensionality, U and V are equal. By the definition of u and v, this implies that they are equal as well.

```
lemma p_implies_uv : p → u = v :=
assume hp : p,
have hpred : U = V, from
  funext (take x : Prop,
    have hl : (x = true ∨ p) → (x = false ∨ p), from
      assume a, or.inr hp,
    have hr : (x = false ∨ p) → (x = true ∨ p), from
      assume a, or.inr hp,
    show (x = true ∨ p) = (x = false ∨ p), from
      propext (iff.intro hl hr)),
have h₀ : ∀ exU exV,
    @classical.some _ U exU = @classical.some _ V exV,
  from hpred ▶ λ exU exV, rfl,
show u = v, from h₀ _ _
```

Putting these last two facts together yields the desired conclusion:

```
theorem em : p ∨ ¬p :=
have h : ¬(u = v) → ¬p, from mt p_implies_uv,
  or.elim not_uv_or_p
    (assume hne : ¬(u = v), or.inr (h hne))
    (assume hp : p, or.inl hp)
```

Consequences of excluded middle include double-negation elimination, proof by cases, and proof by contradiction, all of which are described in Section 3.5. The law of the excluded middle and propositional extensionality imply propositional completeness:

```
theorem prop_complete (a : Prop) : a = true ∨ a = false :=
or.elim (em a)
  (λ t, or.inl (propext (iff.intro (λ h, trivial) (λ h, t))))
  (λ f, or.inr (propext (iff.intro (λ h, absurd h f) (λ h, false.elim h))))
```

Together with choice, we also get the stronger principle that every proposition is decidable. Recall that the class of `decidable` propositions is defined as follows:

```
class inductive decidable (p : Prop)
| is_false : ¬ p → decidable
| is_true :  p → decidable
```

In contrast to `p ∨ ¬ p`, which can only eliminate to `Prop`, the type `decidable p` is equivalent to the sum type `p ⊕ ¬ p`, which can elminate to any type. It is this data that is needed to write an if-then-else expression.

As an example of classical reasoning, we use `some` to show that if `f : α → β` is injective and α is inhabited, then `f` has a left inverse. To define the left inverse `linv`, we use a dependent if-then-else expression. Recall that `if h : c then t else e` is notation for `dite c (λ h : c, t) (λ h : ¬ c, e)`. In the definition of `linv`, choice is used twice: first, to show that `(∃ a : A, f a = b)` is "decidable," and then to choose an `a` such that `f a = b`. Notice that we make `prop_decidable` a local instance to justify the if-then-else expression.

```
open classical function
local attribute [instance] prop_decidable

noncomputable definition linv {α β : Type} [h : inhabited α] (f : α → β) : β → α :=
λ b : β, if ex : (∃ a : α, f a = b) then some ex else arbitrary α

theorem linv_comp_self {α β : Type} {f : α → β}
    [inhabited α] (inj : injective f) :
  linv f ∘ f = id :=
funext (take a,
  have ex  : ∃ a₁ : α, f a₁ = f a, from exists.intro a rfl,
  have   feq : f (some ex) = f a, from some_spec ex,
  calc linv f (f a) = some ex :  dif_pos ex
             ...    = a       :  inj feq)
```

From a classical point of view, `linv` is a function. From a constructive point of view, it is unacceptable; because there is no way to implement such a function in general, the construction is not informative.

Bibliography

[1] Thierry Coquand and Gerard Huet. The calculus of constructions. *Inf. Comput.*, 76(2-3):95–120, February 1988.

[2] Peter Dybjer. Inductive families. *Formal Asp. Comput.*, 6(4):440–465, 1994.

[3] Frank Pfenning and Christine Paulin-Mohring. Inductively defined types in the calculus of constructions. In Michael G. Main, Austin Melton, Michael W. Mislove, and David A. Schmidt, editors, *Mathematical Foundations of Programming Semantics, 5th International Conference, Tulane University, New Orleans, Louisiana, USA, March 29 - April 1, 1989, Proceedings*, volume 442 of *Lecture Notes in Computer Science*, pages 209–228. Springer, 1989.