Contents

Chapter 1

Maps: Total and Partial Maps

Maps (or *dictionaries*) are ubiquitous data structures both generally and in the theory of programming languages in particular; we're going to need them in many places in the coming chapters. They also make a nice case study using ideas we've seen in previous chapters, including building data structures out of higher-order functions (from *Basics* and *Poly*) and the use of reflection to streamline proofs (from *IndProp*).

We'll define two flavors of maps: *total* maps, which include a "default" element to be returned when a key being looked up doesn't exist, and *partial* maps, which return an **option** to indicate success or failure. The latter is defined in terms of the former, using None as the default element.

1.1 The Coq Standard Library

One small digression before we begin...

Unlike the chapters we have seen so far, this one does not `Require Import` the chapter before it (and, transitively, all the earlier chapters). Instead, in this chapter and from now, on we're going to import the definitions and theorems we need directly from Coq's standard library stuff. You should not notice much difference, though, because we've been careful to name our own definitions and theorems the same as their counterparts in the standard library, wherever they overlap.

```
From Coq Require Import Arith.Arith.
From Coq Require Import Bool.Bool.
Require Export Coq.Strings.String.
From Coq Require Import Logic.FunctionalExtensionality.
From Coq Require Import Lists.List.
Import ListNotations.
```

Documentation for the standard library can be found at http://coq.inria.fr/library/.

The `Search` command is a good way to look for theorems involving objects of specific types. Take a minute now to experiment with it.

1.2 Identifiers

First, we need a type for the keys that we use to index into our maps. In *Lists.v* we introduced a fresh type *id* for a similar purpose; here and for the rest of *Software Foundations* we will use the **string** type from Coq's standard library.

To compare strings, we define the function eqb_string, which internally uses the function string_dec from Coq's string library.

```
Definition eqb_string (x y : string) : bool :=
  if string_dec x y then true else false.
```

(The function string_dec comes from Coq's string library. If you check the result type of string_dec, you'll see that it does not actually return a **bool**, but rather a type that looks like {x = y} + {x ≠ y}, called a *sumbool*, which can be thought of as an "evidence-carrying boolean." Formally, an element of *sumbool* is either a proof that two things are equal or a proof that they are unequal, together with a tag indicating which. But for present purposes you can think of it as just a fancy **bool**.)

Now we need a few basic properties of string equality... Theorem eqb_string_refl : ∀ s : string, true = eqb_string s s.

```
Proof. intros s. unfold eqb_string. destruct (string_dec s s) as [|Hs].
  - reflexivity.
  - destruct Hs. reflexivity.
Qed.
```

The following useful property follows from an analogous lemma about strings:

```
Theorem eqb_string_true_iff : ∀ x y : string,
    eqb_string x y = true ↔ x = y.
Proof.
   intros x y.
   unfold eqb_string.
   destruct (string_dec x y) as [|Hs].
   - subst. split. reflexivity. reflexivity.
   - split.
     + intros contra. discriminate contra.
     + intros H. rewrite H in Hs. destruct Hs. reflexivity.
Qed.
```

Similarly:

```
Theorem eqb_string_false_iff : ∀ x y : string,
    eqb_string x y = false ↔ x ≠ y.
Proof.
  intros x y. rewrite ← eqb_string_true_iff.
  rewrite not_true_iff_false. reflexivity. Qed.
```

This handy variant follows just by rewriting:

```
Theorem false_eqb_string : ∀ x y : string,
    x ≠ y → eqb_string x y = false.
Proof.
  intros x y. rewrite eqb_string_false_iff.
  intros H. apply H. Qed.
```

1.3 Total Maps

Our main job in this chapter will be to build a definition of partial maps that is similar in behavior to the one we saw in the *Lists* chapter, plus accompanying lemmas about its behavior.

This time around, though, we're going to use *functions*, rather than lists of key-value pairs, to build maps. The advantage of this representation is that it offers a more *extensional* view of maps, where two maps that respond to queries in the same way will be represented as literally the same thing (the very same function), rather than just "equivalent" data structures. This, in turn, simplifies proofs that use maps.

We build partial maps in two steps. First, we define a type of *total maps* that return a default value when we look up a key that is not present in the map.

```
Definition total_map (A : Type) := string → A.
```

Intuitively, a total map over an element type A is just a function that can be used to look up **string**s, yielding As.

The function t_empty yields an empty total map, given a default element; this map always returns the default element when applied to any string.

```
Definition t_empty {A : Type} (v : A) : total_map A :=
  (fun _ ⇒ v).
```

More interesting is the update function, which (as before) takes a map m, a key x, and a value v and returns a new map that takes x to v and takes every other key to whatever m does.

```
Definition t_update {A : Type} (m : total_map A)
                    (x : string) (v : A) :=
  fun x' ⇒ if eqb_string x x' then v else m x'.
```

This definition is a nice example of higher-order programming: t_update takes a *function* m and yields a new function fun x' ⇒ ... that behaves like the desired map.

For example, we can build a map taking **string**s to **bool**s, where "foo" and "*bar*" are mapped to true and every other key is mapped to false, like this:

```
Definition examplemap :=
  t_update (t_update (t_empty false) "foo" true)
           "bar" true.
```

Next, let's introduce some new notations to facilitate working with maps.

First, we will use the following notation to create an empty total map with a default value. Notation "'_' '!->' v" := (t_empty v)
(at level 100, right associativity).

Example example_empty := (_ !-> false).

We then introduce a convenient notation for extending an existing map with some bindings. Notation "x '!->' v ';' m" := (t_update m x v)
(at level 100, v at *next* level, right associativity).

The examplemap above can now be defined as follows:

```
Definition examplemap' :=
  ( "bar" !-> true;
    "foo" !-> true;
    _ !-> false
  ).
```

This completes the definition of total maps. Note that we don't need to define a *find* operation because it is just function application!

```
Example update_example1 : examplemap' "baz" = false.
Proof. reflexivity. Qed.

Example update_example2 : examplemap' "foo" = true.
Proof. reflexivity. Qed.

Example update_example3 : examplemap' "quux" = false.
Proof. reflexivity. Qed.

Example update_example4 : examplemap' "bar" = true.
Proof. reflexivity. Qed.
```

To use maps in later chapters, we'll need several fundamental facts about how they behave.

Even if you don't work the following exercises, make sure you thoroughly understand the statements of the lemmas!

(Some of the proofs require the functional extensionality axiom, which is discussed in the Logic chapter.)

Exercise: 1 star, standard, optional (t_apply_empty) First, the empty map returns its default element for all keys:

```
Lemma t_apply_empty : ∀ (A : Type) (x : string) (v : A),
    (_ !-> v) x = v.
Proof.
  Admitted.
```

☐

Exercise: 2 stars, standard, optional (t_update_eq) Next, if we update a map m at a key x with a new value *v* and then look up x in the map resulting from the update, we get back *v*:

```
Lemma t_update_eq : ∀ (A : Type) (m : total_map A) x v,
    (x !-> v ; m) x = v.
Proof.
  Admitted.
```

☐

Exercise: 2 stars, standard, optional (t_update_neq) On the other hand, if we update a map m at a key *x1* and then look up a *different* key *x2* in the resulting map, we get the same result that m would have given:

```
Theorem t_update_neq : ∀ (A : Type) (m : total_map A) x1 x2 v,
    x1 ≠ x2 →
    (x1 !-> v ; m) x2 = m x2.
Proof.
  Admitted.
```

☐

Exercise: 2 stars, standard, optional (t_update_shadow) If we update a map m at a key x with a value *v1* and then update again with the same key x and another value *v2*, the resulting map behaves the same (gives the same result when applied to any key) as the simpler map obtained by performing just the second update on m:

```
Lemma t_update_shadow : ∀ (A : Type) (m : total_map A) x v1 v2,
    (x !-> v2 ; x !-> v1 ; m) = (x !-> v2 ; m).
Proof.
  Admitted.
```

☐

For the final two lemmas about total maps, it's convenient to use the reflection idioms introduced in chapter *IndProp*. We begin by proving a fundamental *reflection lemma* relating the equality proposition on *id*s with the boolean function *eqb_id*.

Exercise: 2 stars, standard, optional (eqb_stringP) Use the proof of *eqbP* in chapter *IndProp* as a template to prove the following:

```
Lemma eqb_stringP : ∀ x y : string,
    reflect (x = y) (eqb_string x y).
Proof.
  Admitted.
```

☐

Now, given **string**s *x1* and *x2*, we can use the tactic destruct (eqb_stringP *x1* *x2*) to simultaneously perform case analysis on the result of eqb_string *x1* *x2* and generate hypotheses about the equality (in the sense of =) of *x1* and *x2*.

Exercise: 2 stars, standard (t_update_same) With the example in chapter *IndProp* as a template, use eqb_stringP to prove the following theorem, which states that if we update a map to assign key x the same value as it already has in m, then the result is equal to m:

```
Theorem t_update_same : ∀ (A : Type) (m : total_map A) x,
    (x !-> m x ; m) = m.
Proof.
  Admitted.
  □
```

Exercise: 3 stars, standard, recommended (t_update_permute) Use eqb_stringP to prove one final property of the update function: If we update a map m at two distinct keys, it doesn't matter in which order we do the updates.

```
Theorem t_update_permute : ∀ (A : Type) (m : total_map A)
                                        v1 v2 x1 x2,
    x2 ≠ x1 →
    (x1 !-> v1 ; x2 !-> v2 ; m)
    =
    (x2 !-> v2 ; x1 !-> v1 ; m).
Proof.
  Admitted.
  □
```

1.4 Partial maps

Finally, we define *partial maps* on top of total maps. A partial map with elements of type A is simply a total map with elements of type **option** A and default element None.

```
Definition partial_map (A : Type) := total_map (option A).

Definition empty {A : Type} : partial_map A :=
  t_empty None.

Definition update {A : Type} (m : partial_map A)
              (x : string) (v : A) :=
  (x !-> Some v ; m).
```

We introduce a similar notation for partial maps: Notation "x '|->' v ';' m" := (update *m* *x* *v*)
(at level 100, *v* at *next* level, right associativity).

We can also hide the last case when it is empty. Notation "x '|->' v" := (update empty x v)
(at level 100).

```
Example examplepmap :=
  ("Church" |-> true ; "Turing" |-> false).
```

We now straightforwardly lift all of the basic lemmas about total maps to partial maps.

```
Lemma apply_empty : ∀ (A : Type) (x : string),
    @empty A x = None.
Proof.
  intros. unfold empty. rewrite t_apply_empty.
  reflexivity.
Qed.

Lemma update_eq : ∀ (A : Type) (m : partial_map A) x v,
    (x |-> v ; m) x = Some v.
Proof.
  intros. unfold update. rewrite t_update_eq.
  reflexivity.
Qed.

Theorem update_neq : ∀ (A : Type) (m : partial_map A) x1 x2 v,
    x2 ≠ x1 →
    (x2 |-> v ; m) x1 = m x1.
Proof.
  intros A m x1 x2 v H.
  unfold update. rewrite t_update_neq. reflexivity.
  apply H. Qed.

Lemma update_shadow : ∀ (A : Type) (m : partial_map A) x v1 v2,
    (x |-> v2 ; x |-> v1 ; m) = (x |-> v2 ; m).
Proof.
  intros A m x v1 v2. unfold update. rewrite t_update_shadow.
  reflexivity.
Qed.

Theorem update_same : ∀ (A : Type) (m : partial_map A) x v,
    m x = Some v →
    (x |-> v ; m) = m.
Proof.
  intros A m x v H. unfold update. rewrite ← H.
  apply t_update_same.
Qed.

Theorem update_permute : ∀ (A : Type) (m : partial_map A)
                                x1 x2 v1 v2,
```

```
    x2 ≠ x1 →
    (x1 |-> v1 ; x2 |-> v2 ; m) = (x2 |-> v2 ; x1 |-> v1 ; m).
Proof.
  intros A m x1 x2 v1 v2. unfold update.
  apply t_update_permute.
Qed.
```

Chapter 2

Imp: Simple Imperative Programs

In this chapter, we take a more serious look at how to use Coq to study other things. Our case study is a *simple imperative programming language* called Imp, embodying a tiny core fragment of conventional mainstream languages such as C and Java. Here is a familiar mathematical function written in Imp.

Z ::= X;; Y ::= 1;; WHILE ~(Z = 0) DO Y ::= Y * Z;; Z ::= Z - 1 END

We concentrate here on defining the *syntax* and *semantics* of Imp; later chapters in *Programming Language Foundations* (*Software Foundations*, volume 2) develop a theory of *program equivalence* and introduce *Hoare Logic*, a widely used logic for reasoning about imperative programs.

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Bool.Bool.
From Coq Require Import Init.Nat.
From Coq Require Import Arith.Arith.
From Coq Require Import Arith.EqNat.
From Coq Require Import omega.Omega.
From Coq Require Import Lists.List.
From Coq Require Import Strings.String.
Import ListNotations.

From PLF Require Import Maps.
```

2.1 Arithmetic and Boolean Expressions

We'll present Imp in three parts: first a core language of *arithmetic and boolean expressions*, then an extension of these expressions with *variables*, and finally a language of *commands* including assignment, conditions, sequencing, and loops.

2.1.1 Syntax

```
Module AExp.
```

These two definitions specify the *abstract syntax* of arithmetic and boolean expressions.

```
Inductive aexp : Type :=
  | ANum (n : nat)
  | APlus (a1 a2 : aexp)
  | AMinus (a1 a2 : aexp)
  | AMult (a1 a2 : aexp).

Inductive bexp : Type :=
  | BTrue
  | BFalse
  | BEq (a1 a2 : aexp)
  | BLe (a1 a2 : aexp)
  | BNot (b : bexp)
  | BAnd (b1 b2 : bexp).
```

In this chapter, we'll mostly elide the translation from the concrete syntax that a programmer would actually write to these abstract syntax trees – the process that, for example, would translate the string "1 + 2 × 3" to the AST

APlus (ANum 1) (AMult (ANum 2) (ANum 3)).

The optional chapter *ImpParser* develops a simple lexical analyzer and parser that can perform this translation. You do *not* need to understand that chapter to understand this one, but if you haven't already taken a course where these techniques are covered (e.g., a compilers course) you may want to skim it.

For comparison, here's a conventional BNF (Backus-Naur Form) grammar defining the same abstract syntax:

a ::= nat | a + a | a - a | a * a

b ::= true | false | a = a | a <= a | ˜ b | b && b

Compared to the Coq version above...

- The BNF is more informal – for example, it gives some suggestions about the surface syntax of expressions (like the fact that the addition operation is written with an infix +) while leaving other aspects of lexical analysis and parsing (like the relative precedence of +, -, and ×, the use of parens to group subexpressions, etc.) unspecified. Some additional information – and human intelligence – would be required to turn this description into a formal definition, e.g., for implementing a compiler.

  The Coq version consistently omits all this information and concentrates on the abstract syntax only.

- Conversely, the BNF version is lighter and easier to read. Its informality makes it flexible, a big advantage in situations like discussions at the blackboard, where conveying general ideas is more important than getting every detail nailed down precisely.

  Indeed, there are dozens of BNF-like notations and people switch freely among them, usually without bothering to say which kind of BNF they're using because there is no need to: a rough-and-ready informal understanding is all that's important.

It's good to be comfortable with both sorts of notations: informal ones for communicating between humans and formal ones for carrying out implementations and proofs.

2.1.2 Evaluation

Evaluating an arithmetic expression produces a number.

```
Fixpoint aeval (a : aexp) : nat :=
  match a with
  | ANum n ⇒ n
  | APlus a1 a2 ⇒ (aeval a1) + (aeval a2)
  | AMinus a1 a2 ⇒ (aeval a1) - (aeval a2)
  | AMult a1 a2 ⇒ (aeval a1) × (aeval a2)
  end.

Example test_aeval1:
  aeval (APlus (ANum 2) (ANum 2)) = 4.
Proof. reflexivity. Qed.
```

Similarly, evaluating a boolean expression yields a boolean.

```
Fixpoint beval (b : bexp) : bool :=
  match b with
  | BTrue ⇒ true
  | BFalse ⇒ false
  | BEq a1 a2 ⇒ (aeval a1) =? (aeval a2)
  | BLe a1 a2 ⇒ (aeval a1) <=? (aeval a2)
  | BNot b1 ⇒ negb (beval b1)
  | BAnd b1 b2 ⇒ andb (beval b1) (beval b2)
  end.
```

2.1.3 Optimization

We haven't defined very much yet, but we can already get some mileage out of the definitions. Suppose we define a function that takes an arithmetic expression and slightly simplifies it, changing every occurrence of $0 + e$ (i.e., (APlus (ANum 0) e) into just e.

```
Fixpoint optimize_0plus (a:aexp) : aexp :=
  match a with
  | ANum n ⇒ ANum n
  | APlus (ANum 0) e2 ⇒ optimize_0plus e2
  | APlus e1 e2 ⇒ APlus (optimize_0plus e1) (optimize_0plus e2)
  | AMinus e1 e2 ⇒ AMinus (optimize_0plus e1) (optimize_0plus e2)
  | AMult e1 e2 ⇒ AMult (optimize_0plus e1) (optimize_0plus e2)
  end.
```

To make sure our optimization is doing the right thing we can test it on some examples and see if the output looks OK.

```
Example test_optimize_0plus:
  optimize_0plus (APlus (ANum 2)
                        (APlus (ANum 0)
                               (APlus (ANum 0) (ANum 1))))
  = APlus (ANum 2) (ANum 1).
Proof. reflexivity. Qed.
```

But if we want to be sure the optimization is correct – i.e., that evaluating an optimized expression gives the same result as the original – we should prove it.

```
Theorem optimize_0plus_sound: ∀ a,
  aeval (optimize_0plus a) = aeval a.
Proof.
  intros a. induction a.
  - reflexivity.
  - destruct a1 eqn:Ea1.
    + destruct n eqn:En.
      × simpl. apply IHa2.
      × simpl. rewrite IHa2. reflexivity.
    +
      simpl. simpl in IHa1. rewrite IHa1.
      rewrite IHa2. reflexivity.
    +
      simpl. simpl in IHa1. rewrite IHa1.
      rewrite IHa2. reflexivity.
    +
      simpl. simpl in IHa1. rewrite IHa1.
      rewrite IHa2. reflexivity.
  -
    simpl. rewrite IHa1. rewrite IHa2. reflexivity.
  -
    simpl. rewrite IHa1. rewrite IHa2. reflexivity. Qed.
```

2.2 Coq Automation

The amount of repetition in this last proof is a little annoying. And if either the language of arithmetic expressions or the optimization being proved sound were significantly more complex, it would start to be a real problem.

So far, we've been doing all our proofs using just a small handful of Coq's tactics and completely ignoring its powerful facilities for constructing parts of proofs automatically. This section introduces some of these facilities, and we will see more over the next several chapters.

Getting used to them will take some energy – Coq's automation is a power tool – but it will allow us to scale up our efforts to more complex definitions and more interesting properties without becoming overwhelmed by boring, repetitive, low-level details.

2.2.1 Tacticals

Tacticals is Coq's term for tactics that take other tactics as arguments – "higher-order tactics," if you will.

The `try` Tactical

If T is a tactic, then `try` T is a tactic that is just like T except that, if T fails, `try` T *successfully* does nothing at all (rather than failing).

```
Theorem silly1 : ∀ ae, aeval ae = aeval ae.
Proof. try reflexivity. Qed.

Theorem silly2 : ∀ (P : Prop), P → P.
Proof.
  intros P HP.
  try reflexivity.   apply HP. Qed.
```

There is no real reason to use `try` in completely manual proofs like these, but it is very useful for doing automated proofs in conjunction with the ; tactical, which we show next.

The ; Tactical (Simple Form)

In its most common form, the ; tactical takes two tactics as arguments. The compound tactic T;*T'* first performs T and then performs *T'* on *each subgoal* generated by T.

For example, consider the following trivial lemma:

```
Lemma foo : ∀ n, 0 <=? n = true.
Proof.
  intros.
  destruct n.
    - simpl. reflexivity.
    - simpl. reflexivity.
Qed.
```

We can simplify this proof using the ; tactical:

```
Lemma foo' : ∀ n, 0 <=? n = true.
Proof.
  intros.
  destruct n;

  simpl;
```

```
  reflexivity.
Qed.
```

Using try and ; together, we can get rid of the repetition in the proof that was bothering us a little while ago.

```
Theorem optimize_0plus_sound': ∀ a,
  aeval (optimize_0plus a) = aeval a.
Proof.
  intros a.
  induction a;

    try (simpl; rewrite IHa1; rewrite IHa2; reflexivity).
  - reflexivity.
  -
    destruct a1 eqn:Ea1;

      try (simpl; simpl in IHa1; rewrite IHa1;
           rewrite IHa2; reflexivity).
    + destruct n eqn:En;
      simpl; rewrite IHa2; reflexivity. Qed.
```

Coq experts often use this "...; try... " idiom after a tactic like induction to take care of many similar cases all at once. Naturally, this practice has an analog in informal proofs. For example, here is an informal proof of the optimization theorem that matches the structure of the formal one:

Theorem: For all arithmetic expressions a,

aeval (optimize_0plus a) = aeval a.

Proof: By induction on a. Most cases follow directly from the IH. The remaining cases are as follows:

- Suppose a = ANum n for some n. We must show

  aeval (optimize_0plus (ANum n)) = aeval (ANum n).

  This is immediate from the definition of optimize_0plus.

- Suppose a = APlus *a1* *a2* for some *a1* and *a2*. We must show

  aeval (optimize_0plus (APlus a1 a2)) = aeval (APlus a1 a2).

  Consider the possible forms of *a1*. For most of them, optimize_0plus simply calls itself recursively for the subexpressions and rebuilds a new expression of the same form as *a1*; in these cases, the result follows directly from the IH.

  The interesting case is when *a1* = ANum n for some n. If n = 0, then

  optimize_0plus (APlus a1 a2) = optimize_0plus a2 and the IH for *a2* is exactly what we need. On the other hand, if n = S *n'* for some *n'*, then again optimize_0plus simply calls itself recursively, and the result follows from the IH. ☐

However, this proof can still be improved: the first case (for a = ANum n) is very trivial – even more trivial than the cases that we said simply followed from the IH – yet we have chosen to write it out in full. It would be better and clearer to drop it and just say, at the top, "Most cases are either immediate or direct from the IH. The only interesting case is the one for APlus..." We can make the same improvement in our formal proof too. Here's how it looks:

```
Theorem optimize_0plus_sound'': ∀ a,
  aeval (optimize_0plus a) = aeval a.
Proof.
  intros a.
  induction a;

    try (simpl; rewrite IHa1; rewrite IHa2; reflexivity);

    try reflexivity.
  -
    destruct a1; try (simpl; simpl in IHa1; rewrite IHa1;
                      rewrite IHa2; reflexivity).
    + destruct n;
      simpl; rewrite IHa2; reflexivity. Qed.
```

The ; Tactical (General Form)

The ; tactical also has a more general form than the simple T;*T'* we've seen above. If T, *T1*, ..., *Tn* are tactics, then

T; *T1* | *T2* | ... | *Tn*

is a tactic that first performs T and then performs *T1* on the first subgoal generated by T, performs *T2* on the second subgoal, etc.

So T;*T'* is just special notation for the case when all of the *Ti*'s are the same tactic; i.e., T;*T'* is shorthand for:

T; *T'* | *T'* | ... | *T'*

The `repeat` Tactical

The `repeat` tactical takes another tactic and keeps applying this tactic until it fails. Here is an example showing that 10 is in a long list using repeat.

```
Theorem In10 : In 10 [1;2;3;4;5;6;7;8;9;10].
Proof.
  repeat (try (left; reflexivity); right).
```

```
Qed.
```

The tactic `repeat` T never fails: if the tactic T doesn't apply to the original goal, then repeat still succeeds without changing the original goal (i.e., it repeats zero times).

```
Theorem In10' : In 10 [1;2;3;4;5;6;7;8;9;10].
Proof.
  repeat (left; reflexivity).
  repeat (right; try (left; reflexivity)).
Qed.
```

The tactic `repeat` T also does not have any upper bound on the number of times it applies T. If T is a tactic that always succeeds, then repeat T will loop forever (e.g., `repeat simpl` loops, since `simpl` always succeeds). While evaluation in Coq's term language, Gallina, is guaranteed to terminate, tactic evaluation is not! This does not affect Coq's logical consistency, however, since the job of `repeat` and other tactics is to guide Coq in constructing proofs; if the construction process diverges (i.e., it does not terminate), this simply means that we have failed to construct a proof, not that we have constructed a wrong one.

Exercise: 3 stars, standard (optimize_0plus_b_sound) Since the optimize_0plus transformation doesn't change the value of **aexp**s, we should be able to apply it to all the **aexp**s that appear in a **bexp** without changing the **bexp**'s value. Write a function that performs this transformation on **bexp**s and prove it is sound. Use the tacticals we've just seen to make the proof as elegant as possible.

```
Fixpoint optimize_0plus_b (b : bexp) : bexp
  . Admitted.

Theorem optimize_0plus_b_sound : ∀ b,
  beval (optimize_0plus_b b) = beval b.
Proof.
   Admitted.
   □
```

Exercise: 4 stars, standard, optional (optimize) *Design exercise*: The optimization implemented by our optimize_0plus function is only one of many possible optimizations on arithmetic and boolean expressions. Write a more sophisticated optimizer and prove it correct. (You will probably find it easiest to start small – add just a single, simple optimization and its correctness proof – and build up to something more interesting incrementally.)

2.2.2 Defining New Tactic Notations

Coq also provides several ways of "programming" tactic scripts.

- The `Tactic Notation` idiom illustrated below gives a handy way to define "shorthand tactics" that bundle several tactics into a single command.

- For more sophisticated programming, Coq offers a built-in language called `Ltac` with primitives that can examine and modify the proof state. The details are a bit too complicated to get into here (and it is generally agreed that `Ltac` is not the most beautiful part of Coq's design!), but they can be found in the reference manual and other books on Coq, and there are many examples of `Ltac` definitions in the Coq standard library that you can use as examples.

- There is also an OCaml API, which can be used to build tactics that access Coq's internal structures at a lower level, but this is seldom worth the trouble for ordinary Coq users.

The `Tactic Notation` mechanism is the easiest to come to grips with, and it offers plenty of power for many purposes. Here's an example.

```
Tactic Notation "simpl_and_try" tactic(c) :=
  simpl;
  try c.
```

This defines a new tactical called *simpl_and_try* that takes one tactic c as an argument and is defined to be equivalent to the tactic `simpl; try` c. Now writing "*simpl_and_try* `reflexivity`." in a proof will be the same as writing "`simpl; try reflexivity`."

2.2.3 The omega Tactic

The `omega` tactic implements a decision procedure for a subset of first-order logic called *Presburger arithmetic*. It is based on the Omega algorithm invented by William Pugh *Pugh* 1991 (in Bib.v).

If the goal is a universally quantified formula made out of

- numeric constants, addition (+ and S), subtraction (- and pred), and multiplication by constants (this is what makes it Presburger arithmetic),
- equality (= and $\neq$) and ordering ($\leq$), and
- the logical connectives $\wedge$, $\vee$, $\neg$, and $\rightarrow$,

then invoking `omega` will either solve the goal or fail, meaning that the goal is actually false. (If the goal is *not* of this form, `omega` will also fail.)

```
Example silly_presburger_example : ∀ m n o p,
  m + n ≤ n + o ∧ o + 3 = p + 3 →
  m ≤ p.
Proof.
  intros. omega.
Qed.
```

(Note the `From` *Coq* `Require Import omega.Omega.` at the top of the file.)

2.2.4 A Few More Handy Tactics

Finally, here are some miscellaneous tactics that you may find convenient.

- `clear` H: Delete hypothesis H from the context.
- `subst` x: For a variable x, find an assumption x $=$ e or e $=$ x in the context, replace x with e throughout the context and current goal, and clear the assumption.
- `subst`: Substitute away *all* assumptions of the form x $=$ e or e $=$ x (where x is a variable).
- `rename`... *into*...: Change the name of a hypothesis in the proof context. For example, if the context includes a variable named x, then `rename` x *into* y will change all occurrences of x to y.
- `assumption`: Try to find a hypothesis H in the context that exactly matches the goal; if one is found, behave like `apply` H.
- *contradiction*: Try to find a hypothesis H in the current context that is logically equivalent to **False**. If one is found, solve the goal.
- `constructor`: Try to find a constructor c (from some `Inductive` definition in the current environment) that can be applied to solve the current goal. If one is found, behave like `apply` c.

We'll see examples of all of these as we go along.

2.3 Evaluation as a Relation

We have presented aeval and beval as functions defined by `Fixpoints`. Another way to think about evaluation – one that we will see is often more flexible – is as a *relation* between expressions and their values. This leads naturally to `Inductive` definitions like the following one for arithmetic expressions...

```
Module AEVALR_FIRST_TRY.

Inductive aevalR : aexp → nat → Prop :=
  | E_ANum n :
      aevalR (ANum n) n
  | E_APlus (e1 e2: aexp) (n1 n2: nat) :
      aevalR e1 n1 →
      aevalR e2 n2 →
      aevalR (APlus e1 e2) (n1 + n2)
  | E_AMinus (e1 e2: aexp) (n1 n2: nat) :
      aevalR e1 n1 →
```

```
    aevalR e2 n2 →
    aevalR (AMinus e1 e2) (n1 - n2)
| E_AMult (e1 e2: aexp) (n1 n2: nat) :
    aevalR e1 n1 →
    aevalR e2 n2 →
    aevalR (AMult e1 e2) (n1 × n2).
```

```
Module TooHardToRead.
```

```
Inductive aevalR : aexp → nat → Prop :=
  | E_ANum n :
      aevalR (ANum n) n
  | E_APlus (e1 e2: aexp) (n1 n2: nat)
      (H1 : aevalR e1 n1)
      (H2 : aevalR e2 n2) :
      aevalR (APlus e1 e2) (n1 + n2)
  | E_AMinus (e1 e2: aexp) (n1 n2: nat)
      (H1 : aevalR e1 n1)
      (H2 : aevalR e2 n2) :
      aevalR (AMinus e1 e2) (n1 - n2)
  | E_AMult (e1 e2: aexp) (n1 n2: nat)
      (H1 : aevalR e1 n1)
      (H2 : aevalR e2 n2) :
      aevalR (AMult e1 e2) (n1 × n2).
```

Instead, we've chosen to leave the hypotheses anonymous, just giving their types. This style gives us less control over the names that Coq chooses during proofs involving **aevalR**, but it makes the definition itself quite a bit lighter.

```
End TooHardToRead.
```

It will be convenient to have an infix notation for **aevalR**. We'll write e \\ n to mean that arithmetic expression e evaluates to value n.

```
Notation "e '\\' n"
          := (aevalR e n)
             (at level 50, left associativity)
          : type_scope.
```

```
End aevalR_first_try.
```

In fact, Coq provides a way to use this notation in the definition of **aevalR** itself. This reduces confusion by avoiding situations where we're working on a proof involving statements in the form e \\ n but we have to refer back to a definition written using the form **aevalR** e n.

We do this by first "reserving" the notation, then giving the definition together with a declaration of what the notation means.

```
Reserved Notation "e '\\' n" (at level 90, left associativity).
```

```
Inductive aevalR : aexp → nat → Prop :=
  | E_ANum (n : nat) :
      (ANum n) \\ n
  | E_APlus (e1 e2 : aexp) (n1 n2 : nat) :
      (e1 \\ n1) → (e2 \\ n2) → (APlus e1 e2) \\ (n1 + n2)
  | E_AMinus (e1 e2 : aexp) (n1 n2 : nat) :
      (e1 \\ n1) → (e2 \\ n2) → (AMinus e1 e2) \\ (n1 - n2)
  | E_AMult (e1 e2 : aexp) (n1 n2 : nat) :
      (e1 \\ n1) → (e2 \\ n2) → (AMult e1 e2) \\ (n1 × n2)

  where "e '\\' n" := (aevalR e n) : type_scope.
```

2.3.1 Inference Rule Notation

In informal discussions, it is convenient to write the rules for **aevalR** and similar relations in the more readable graphical form of *inference rules*, where the premises above the line justify the conclusion below the line (we have already seen them in the *IndProp* chapter).

For example, the constructor E_APlus...

| E_APlus : forall (e1 e2: aexp) (n1 n2: nat), aevalR e1 n1 -> aevalR e2 n2 -> aevalR (APlus e1 e2) (n1 + n2)

...would be written like this as an inference rule:

e1 \\ n1 e2 \\ n2

(E_APlus) APlus e1 e2 \\ n1+n2

Formally, there is nothing deep about inference rules: they are just implications. You can read the rule name on the right as the name of the constructor and read each of the linebreaks between the premises above the line (as well as the line itself) as →. All the variables mentioned in the rule (*e1*, *n1*, etc.) are implicitly bound by universal quantifiers at the beginning. (Such variables are often called *metavariables* to distinguish them from the variables of the language we are defining. At the moment, our arithmetic expressions don't include variables, but we'll soon be adding them.) The whole collection of rules is understood as being wrapped in an `Inductive` declaration. In informal prose, this is either elided or else indicated by saying something like "Let **aevalR** be the smallest relation closed under the following rules...".

For example, \\ is the smallest relation closed under these rules:

(E_ANum) ANum n \\ n

e1 \\ n1 e2 \\ n2

(E_APlus) APlus e1 e2 \\ n1+n2

e1 \\ n1 e2 \\ n2

(E_AMinus) AMinus e1 e2 \\ n1-n2

e1 \\ n1 e2 \\ n2

(E_AMult) AMult e1 e2 \\ n1*n2

Exercise: 1 star, standard, optional (beval_rules) Here, again, is the Coq definition of the beval function:

Fixpoint beval (e : bexp) : bool := match e with | BTrue => true | BFalse => false | BEq a1 a2 => (aeval a1) =? (aeval a2) | BLe a1 a2 => (aeval a1) <=? (aeval a2) | BNot b1 => negb (beval b1) | BAnd b1 b2 => andb (beval b1) (beval b2) end.

Write out a corresponding definition of boolean evaluation as a relation (in inference rule notation).

```
Definition manual_grade_for_beval_rules : option (nat×string) := None.
```

□

2.3.2 Equivalence of the Definitions

It is straightforward to prove that the relational and functional definitions of evaluation agree:

```
Theorem aeval_iff_aevalR : ∀ a n,
  (a \\ n) ↔ aeval a = n.
Proof.
 split.
 -
   intros H.
   induction H; simpl.
   +
     reflexivity.
   +
     rewrite IHaevalR1. rewrite IHaevalR2. reflexivity.
   +
     rewrite IHaevalR1. rewrite IHaevalR2. reflexivity.
   +
     rewrite IHaevalR1. rewrite IHaevalR2. reflexivity.
 -
   generalize dependent n.
   induction a;
      simpl; intros; subst.
   +
     apply E_ANum.
   +
     apply E_APlus.
      apply IHa1. reflexivity.
```

```
      apply IHa2. reflexivity.
    +
      apply E_AMinus.
       apply IHa1. reflexivity.
       apply IHa2. reflexivity.
    +
      apply E_AMult.
       apply IHa1. reflexivity.
       apply IHa2. reflexivity.
Qed.
```

We can make the proof quite a bit shorter by making more use of tacticals.

```
Theorem aeval_iff_aevalR' : ∀ a n,
  (a \\ n) ↔ aeval a = n.
Proof.
  split.
  -
    intros H; induction H; subst; reflexivity.
  -
    generalize dependent n.
    induction a; simpl; intros; subst; constructor;
       try apply IHa1; try apply IHa2; reflexivity.
Qed.
```

Exercise: 3 stars, standard (bevalR) Write a relation **bevalR** in the same style as **aevalR**, and prove that it is equivalent to beval.

```
Inductive bevalR: bexp → bool → Prop :=

.
Lemma beval_iff_bevalR : ∀ b bv,
  bevalR b bv ↔ beval b = bv.
Proof.
   Admitted.
   □
End AExp.
```

2.3.3 Computational vs. Relational Definitions

For the definitions of evaluation for arithmetic and boolean expressions, the choice of whether to use functional or relational definitions is mainly a matter of taste: either way works.

However, there are circumstances where relational definitions of evaluation work much better than functional ones.

Module AEVALR_DIVISION.

For example, suppose that we wanted to extend the arithmetic operations with division:

```
Inductive aexp : Type :=
  | ANum (n : nat)
  | APlus (a1 a2 : aexp)
  | AMinus (a1 a2 : aexp)
  | AMult (a1 a2 : aexp)
  | ADiv (a1 a2 : aexp).
```

Extending the definition of aeval to handle this new operation would not be straightforward (what should we return as the result of ADiv (ANum 5) (ANum 0)?). But extending **aevalR** is straightforward.

```
Reserved Notation "e '\\' n"
                  (at level 90, left associativity).

Inductive aevalR : aexp → nat → Prop :=
  | E_ANum (n : nat) :
      (ANum n) \\ n
  | E_APlus (a1 a2 : aexp) (n1 n2 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (APlus a1 a2) \\ (n1 + n2)
  | E_AMinus (a1 a2 : aexp) (n1 n2 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (AMinus a1 a2) \\ (n1 - n2)
  | E_AMult (a1 a2 : aexp) (n1 n2 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (AMult a1 a2) \\ (n1 × n2)
  | E_ADiv (a1 a2 : aexp) (n1 n2 n3 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (n2 > 0) →
      (mult n2 n3 = n1) → (ADiv a1 a2) \\ n3

where "a '\\' n" := (aevalR a n) : type_scope.
```

End AEVALR_DIVISION.

Module AEVALR_EXTENDED.

Or suppose that we want to extend the arithmetic operations by a nondeterministic number generator *any* that, when evaluated, may yield any number. (Note that this is not the same as making a *probabilistic* choice among all possible numbers – we're not specifying any particular probability distribution for the results, just saying what results are *possible*.)

```
Reserved Notation "e '\\' n" (at level 90, left associativity).

Inductive aexp : Type :=
  | AAny
  | ANum (n : nat)
  | APlus (a1 a2 : aexp)
  | AMinus (a1 a2 : aexp)
  | AMult (a1 a2 : aexp).
```

Again, extending aeval would be tricky, since now evaluation is *not* a deterministic function from expressions to numbers, but extending **aevalR** is no problem...

```
Inductive aevalR : aexp → nat → Prop :=
  | E_Any (n : nat) :
      AAny \\ n
  | E_ANum (n : nat) :
      (ANum n) \\ n
  | E_APlus (a1 a2 : aexp) (n1 n2 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (APlus a1 a2) \\ (n1 + n2)
  | E_AMinus (a1 a2 : aexp) (n1 n2 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (AMinus a1 a2) \\ (n1 - n2)
  | E_AMult (a1 a2 : aexp) (n1 n2 : nat) :
      (a1 \\ n1) → (a2 \\ n2) → (AMult a1 a2) \\ (n1 × n2)

where "a '\\' n" := (aevalR a n) : type_scope.

End AEVALR_EXTENDED.
```

At this point you maybe wondering: which style should I use by default? In the examples we've just seen, relational definitions turned out to be more useful than functional ones. For situations like these, where the thing being defined is not easy to express as a function, or indeed where it is *not* a function, there is no real choice. But what about when both styles are workable?

One point in favor of relational definitions is that they can be more elegant and easier to understand.

Another is that Coq automatically generates nice inversion and induction principles from `Inductive` definitions.

On the other hand, functional definitions can often be more convenient:

- Functions are by definition deterministic and defined on all arguments; for a relation we have to show these properties explicitly if we need them.

- With functions we can also take advantage of Coq's computation mechanism to simplify expressions during proofs.

Furthermore, functions can be directly "extracted" from Gallina to executable code in OCaml or Haskell.

Ultimately, the choice often comes down to either the specifics of a particular situation or simply a question of taste. Indeed, in large Coq developments it is common to see a definition given in *both* functional and relational styles, plus a lemma stating that the two coincide, allowing further proofs to switch from one point of view to the other at will.

2.4 Expressions With Variables

Back to defining Imp. The next thing we need to do is to enrich our arithmetic and boolean expressions with variables. To keep things simple, we'll assume that all variables are global and that they only hold numbers.

2.4.1 States

Since we'll want to look variables up to find out their current values, we'll reuse maps from the Maps chapter, and **strings** will be used to represent variables in Imp.

A *machine state* (or just *state*) represents the current values of *all* variables at some point in the execution of a program.

For simplicity, we assume that the state is defined for *all* variables, even though any given program is only going to mention a finite number of them. The state captures all of the information stored in memory. For Imp programs, because each variable stores a natural number, we can represent the state as a mapping from strings to **nat**, and will use 0 as default value in the store. For more complex programming languages, the state might have more structure.

```
Definition state := total_map nat.
```

2.4.2 Syntax

We can add variables to the arithmetic expressions we had before by simply adding one more constructor:

```
Inductive aexp : Type :=
  | ANum (n : nat)
  | AId (x : string)
  | APlus (a1 a2 : aexp)
  | AMinus (a1 a2 : aexp)
  | AMult (a1 a2 : aexp).
```

Defining a few variable names as notational shorthands will make examples easier to read:

```
Definition W : string := "W".
Definition X : string := "X".
Definition Y : string := "Y".
Definition Z : string := "Z".
```

(This convention for naming program variables (X, Y, Z) clashes a bit with our earlier use of uppercase letters for types. Since we're not using polymorphism heavily in the chapters developed to Imp, this overloading should not cause confusion.)

The definition of **bexp**s is unchanged (except that it now refers to the new **aexp**s):

```
Inductive bexp : Type :=
```

```
| BTrue
| BFalse
| BEq (a1 a2 : aexp)
| BLe (a1 a2 : aexp)
| BNot (b : bexp)
| BAnd (b1 b2 : bexp).
```

2.4.3 Notations

To make Imp programs easier to read and write, we introduce some notations and implicit coercions.

You do not need to understand exactly what these declarations do. Briefly, though, the `Coercion` declaration in Coq stipulates that a function (or constructor) can be implicitly used by the type system to coerce a value of the input type to a value of the output type. For instance, the coercion declaration for AId allows us to use plain strings when an **aexp** is expected; the string will implicitly be wrapped with AId.

The notations below are declared in specific *notation scopes*, in order to avoid conflicts with other interpretations of the same symbols. Again, it is not necessary to understand the details, but it is important to recognize that we are defining *new* interpretations for some familiar operators like +, -, ×, =, ≤, etc.

```
Coercion AId : string >-> aexp.
Coercion ANum : nat >-> aexp.

Definition bool_to_bexp (b : bool) : bexp :=
  if b then BTrue else BFalse.
Coercion bool_to_bexp : bool >-> bexp.

Bind Scope imp_scope with aexp.
Bind Scope imp_scope with bexp.
Delimit Scope imp_scope with imp.

Notation "x + y" := (APlus x y) (at level 50, left associativity) : imp_scope.
Notation "x - y" := (AMinus x y) (at level 50, left associativity) : imp_scope.
Notation "x * y" := (AMult x y) (at level 40, left associativity) : imp_scope.
Notation "x <= y" := (BLe x y) (at level 70, no associativity) : imp_scope.
Notation "x = y" := (BEq x y) (at level 70, no associativity) : imp_scope.
Notation "x && y" := (BAnd x y) (at level 40, left associativity) : imp_scope.
Notation "'~' b" := (BNot b) (at level 75, right associativity) : imp_scope.
```

We can now write 3 + (X × 2) instead of APlus 3 (AMult X 2), and true && ~(X ≤ 4) instead of BAnd true (BNot (BLe X 4)).

```
Definition example_aexp := (3 + (X × 2))%imp : aexp.
Definition example_bexp := (true && ~(X ≤ 4))%imp : bexp.
```

One downside of these coercions is that they can make it a little harder for humans to calculate the types of expressions. If you get confused, try doing `Set Printing` *Coercions* to see exactly what is going on.

```
Set Printing Coercions.
Print example_bexp.
Unset Printing Coercions.
```

2.4.4 Evaluation

The arith and boolean evaluators are extended to handle variables in the obvious way, taking a state as an extra argument:

```
Fixpoint aeval (st : state) (a : aexp) : nat :=
  match a with
  | ANum n ⇒ n
  | AId x ⇒ st x
  | APlus a1 a2 ⇒ (aeval st a1) + (aeval st a2)
  | AMinus a1 a2 ⇒ (aeval st a1) - (aeval st a2)
  | AMult a1 a2 ⇒ (aeval st a1) × (aeval st a2)
  end.

Fixpoint beval (st : state) (b : bexp) : bool :=
  match b with
  | BTrue ⇒ true
  | BFalse ⇒ false
  | BEq a1 a2 ⇒ (aeval st a1) =? (aeval st a2)
  | BLe a1 a2 ⇒ (aeval st a1) <=? (aeval st a2)
  | BNot b1 ⇒ negb (beval st b1)
  | BAnd b1 b2 ⇒ andb (beval st b1) (beval st b2)
  end.
```

We specialize our notation for total maps to the specific case of states, i.e. using (_ !-> 0) as empty state.

```
Definition empty_st := (_ !-> 0).
```

Now we can add a notation for a "singleton state" with just one variable bound to a value.

```
Notation "a '!->' x" := (t_update empty_st a x) (at level 100).

Example aexp1 :
    aeval (X !-> 5) (3 + (X × 2))%imp
  = 13.
Proof. reflexivity. Qed.

Example bexp1 :
    beval (X !-> 5) (true && ~(X ≤ 4))%imp
  = true.
Proof. reflexivity. Qed.
```

2.5 Commands

Now we are ready define the syntax and behavior of Imp *commands* (sometimes called *statements*).

2.5.1 Syntax

Informally, commands *c* are described by the following BNF grammar.

c ::= SKIP | x ::= a | c ;; c | TEST b THEN c ELSE c FI | WHILE b DO c END

(We choose this slightly awkward concrete syntax for the sake of being able to define Imp syntax using Coq's notation mechanism. In particular, we use *TEST* to avoid conflicting with the `if` and *IF* notations from the standard library.) For example, here's factorial in Imp:

Z ::= X;; Y ::= 1;; WHILE ~(Z = 0) DO Y ::= Y * Z;; Z ::= Z - 1 END

When this command terminates, the variable Y will contain the factorial of the initial value of X.

Here is the formal definition of the abstract syntax of commands:

```
Inductive com : Type :=
  | CSkip
  | CAss (x : string) (a : aexp)
  | CSeq (c1 c2 : com)
  | CIf (b : bexp) (c1 c2 : com)
  | CWhile (b : bexp) (c : com).
```

As for expressions, we can use a few `Notation` declarations to make reading and writing Imp programs more convenient.

```
Bind Scope imp_scope with com.
Notation "'SKIP'" :=
   CSkip : imp_scope.
Notation "x '::=' a" :=
  (CAss x a) (at level 60) : imp_scope.
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity) : imp_scope.
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity) : imp_scope.
Notation "'TEST' c1 'THEN' c2 'ELSE' c3 'FI'" :=
  (CIf c1 c2 c3) (at level 80, right associativity) : imp_scope.
```

For example, here is the factorial function again, written as a formal definition to Coq:

```
Definition fact_in_coq : com :=
  (Z ::= X;;
  Y ::= 1;;
  WHILE ~(Z = 0) DO
```

```
    Y ::= Y × Z;;
    Z ::= Z - 1
  END)%imp.
```

2.5.2 Desugaring notations

Coq offers a rich set of features to manage the increasing complexity of the objects we work with, such as coercions and notations. However, their heavy usage can make for quite overwhelming syntax. It is often instructive to "turn off" those features to get a more elementary picture of things, using the following commands:

- `Unset Printing` *Notations* (undo with `Set Printing` *Notations*)
- `Set Printing` *Coercions* (undo with `Unset Printing` *Coercions*)
- `Set Printing All` (undo with `Unset Printing All`)

These commands can also be used in the middle of a proof, to elaborate the current goal and context.

```
Unset Printing Notations.
Print fact_in_coq.
Set Printing Notations.
```

```
Set Printing Coercions.
Print fact_in_coq.
Unset Printing Coercions.
```

2.5.3 The `Locate` command

Finding notations

When faced with unknown notation, use `Locate` with a *string* containing one of its symbols to see its possible interpretations. `Locate "&&".`

```
Locate ";;".
```

```
Locate "WHILE".
```

Finding identifiers

When used with an identifier, the command `Locate` prints the full path to every value in scope with the same name. This is useful to troubleshoot problems due to variable shadowing. `Locate` *aexp*.

2.5.4 More Examples

Assignment:

```
Definition plus2 : com :=
  X ::= X + 2.
```

```
Definition XtimesYinZ : com :=
  Z ::= X × Y.
```

```
Definition subtract_slowly_body : com :=
  Z ::= Z - 1 ;;
  X ::= X - 1.
```

Loops

```
Definition subtract_slowly : com :=
  (WHILE ~(X = 0) DO
    subtract_slowly_body
  END)%imp.
```

```
Definition subtract_3_from_5_slowly : com :=
  X ::= 3 ;;
  Z ::= 5 ;;
  subtract_slowly.
```

An infinite loop:

```
Definition loop : com :=
  WHILE true DO
    SKIP
  END.
```

2.6 Evaluating Commands

Next we need to define what it means to evaluate an Imp command. The fact that *WHILE* loops don't necessarily terminate makes defining an evaluation function tricky...

2.6.1 Evaluation as a Function (Failed Attempt)

Here's an attempt at defining an evaluation function for commands, omitting the *WHILE* case.

The following declaration is needed to be able to use the notations in match patterns.

```
Open Scope imp_scope.
Fixpoint ceval_fun_no_while (st : state) (c : com)
```

```
                              : state :=
  match c with
    | SKIP ⇒
        st
    | x ::= a1 ⇒
        (x !-> (aeval st a1) ; st)
    | c1 ;; c2 ⇒
        let st' := ceval_fun_no_while st c1 in
        ceval_fun_no_while st' c2
    | TEST b THEN c1 ELSE c2 FI ⇒
        if (beval st b)
          then ceval_fun_no_while st c1
          else ceval_fun_no_while st c2
    | WHILE b DO c END ⇒
        st
  end.
Close Scope imp_scope.
```

In a traditional functional programming language like OCaml or Haskell we could add the *WHILE* case as follows:

Fixpoint ceval_fun (st : state) (c : com) : state := match c with ... | WHILE b DO c END => if (beval st b) then ceval_fun st (c ;; WHILE b DO c END) else st end.

Coq doesn't accept such a definition ("Error: Cannot guess decreasing argument of fix") because the function we want to define is not guaranteed to terminate. Indeed, it *doesn't* always terminate: for example, the full version of the *ceval_fun* function applied to the loop program above would never terminate. Since Coq is not just a functional programming language but also a consistent logic, any potentially non-terminating function needs to be rejected. Here is an (invalid!) program showing what would go wrong if Coq allowed non-terminating recursive functions:

Fixpoint loop_false (n : nat) : False := loop_false n.

That is, propositions like **False** would become provable (*loop_false* 0 would be a proof of **False**), which would be a disaster for Coq's logical consistency.

Thus, because it doesn't terminate on all inputs, of *ceval_fun* cannot be written in Coq – at least not without additional tricks and workarounds (see chapter *ImpCEvalFun* if you're curious about what those might be).

2.6.2 Evaluation as a Relation

Here's a better way: define **ceval** as a *relation* rather than a *function* – i.e., define it in `Prop` instead of `Type`, as we did for **aevalR** above.

This is an important change. Besides freeing us from awkward workarounds, it gives us a lot more flexibility in the definition. For example, if we add nondeterministic features like *any* to the language, we want the definition of evaluation to be nondeterministic – i.e., not only will it not be total, it will not even be a function!

We'll use the notation *st* =[*c*]=> *st'* for the **ceval** relation: *st* =[*c*]=> *st'* means that executing program *c* in a starting state *st* results in an ending state *st'*. This can be pronounced "*c* takes state *st* to *st'*".

Operational Semantics

Here is an informal definition of evaluation, presented as inference rules for readability:

(E_Skip) st = *SKIP* => st
aeval st a1 = n

(E_Ass) st = x := *a1* => (x !-> n ; st)
st = *c1* => st' st' = *c2* => st"

(E_Seq) st = *c1*;;*c2* => st"
beval st b1 = true st = *c1* => st'

(E_IfTrue) st = *TEST b1 THEN c1 ELSE c2 FI* => st'
beval st b1 = false st = *c2* => st'

(E_IfFalse) st = *TEST b1 THEN c1 ELSE c2 FI* => st'
beval st b = false

(E_WhileFalse) st = *WHILE b DO c END* => st
beval st b = true st = *c* => st' st' = *WHILE b DO c END* => st"

(E_WhileTrue) st = *WHILE b DO c END* => st"

Here is the formal definition. Make sure you understand how it corresponds to the inference rules.

```
Reserved Notation "st '=[' c ']=>' st'"
                  (at level 40).

Inductive ceval : com → state → state → Prop :=
  | E_Skip : ∀ st,
      st =[ SKIP ]=> st
  | E_Ass : ∀ st a1 n x,
      aeval st a1 = n →
      st =[ x ::= a1 ]=> (x !-> n ; st)
  | E_Seq : ∀ c1 c2 st st' st'',
      st =[ c1 ]=> st' →
      st' =[ c2 ]=> st'' →
      st =[ c1 ;; c2 ]=> st''
```

```
  | E_IfTrue : ∀ st st' b c1 c2,
      beval st b = true →
      st =[ c1 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_IfFalse : ∀ st st' b c1 c2,
      beval st b = false →
      st =[ c2 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_WhileFalse : ∀ b st c,
      beval st b = false →
      st =[ WHILE b DO c END ]=> st
  | E_WhileTrue : ∀ st st' st'' b c,
      beval st b = true →
      st =[ c ]=> st' →
      st' =[ WHILE b DO c END ]=> st'' →
      st =[ WHILE b DO c END ]=> st''

  where "st =[ c ]=> st'" := (ceval c st st').
```

The cost of defining evaluation as a relation instead of a function is that we now need to construct *proofs* that some program evaluates to some result state, rather than just letting Coq's computation mechanism do it for us.

```
Example ceval_example1:
  empty_st =[
     X ::= 2;;
     TEST X ≤ 1
       THEN Y ::= 3
       ELSE Z ::= 4
     FI
  ]=> (Z !-> 4 ; X !-> 2).
Proof.
  apply E_Seq with (X !-> 2).
  -
    apply E_Ass. reflexivity.
  -
    apply E_IfFalse.
    reflexivity.
    apply E_Ass. reflexivity.
Qed.
```

Exercise: 2 stars, standard (ceval_example2) Example ceval_example2:

```
  empty_st =[
    X ::= 0;; Y ::= 1;; Z ::= 2
```

```
  ]=> (Z !-> 2 ; Y !-> 1 ; X !-> 0).
Proof.
    Admitted.
```
□

Exercise: 3 stars, standard, optional (pup_to_n) Write an Imp program that sums the numbers from 1 to X (inclusive: 1 + 2 + ... + X) in the variable Y. Prove that this program executes as intended for X = 2 (this is trickier than you might expect).

```
Definition pup_to_n : com
  . Admitted.

Theorem pup_to_2_ceval :
  (X !-> 2) =[
    pup_to_n
  ]=> (X !-> 0 ; Y !-> 3 ; X !-> 1 ; Y !-> 2 ; Y !-> 0 ; X !-> 2).
Proof.
    Admitted.
```
□

2.6.3 Determinism of Evaluation

Changing from a computational to a relational definition of evaluation is a good move because it frees us from the artificial requirement that evaluation should be a total function. But it also raises a question: Is the second definition of evaluation really a partial function? Or is it possible that, beginning from the same state *st*, we could evaluate some command *c* in different ways to reach two different output states *st'* and *st''*?

In fact, this cannot happen: **ceval** *is* a partial function:

```
Theorem ceval_deterministic: ∀ c st st1 st2,
      st =[ c ]=> st1 →
      st =[ c ]=> st2 →
      st1 = st2.
Proof.
  intros c st st1 st2 E1 E2.
  generalize dependent st2.
  induction E1;
           intros st2 E2; inversion E2; subst.
  - reflexivity.
  - reflexivity.
  -
    assert (st' = st'0) as EQ1.
    { apply IHE1_1; assumption. }
    subst st'0.
    apply IHE1_2. assumption.
```

```
-
    apply IHE1. assumption.
-
    rewrite H in H5. discriminate H5.
-
    rewrite H in H5. discriminate H5.
-
    apply IHE1. assumption.
-
  reflexivity.
-
  rewrite H in H2. discriminate H2.
-
  rewrite H in H4. discriminate H4.
-
    assert (st' = st'0) as EQ1.
    { apply IHE1_1; assumption. }
    subst st'0.
    apply IHE1_2. assumption. Qed.
```

2.7 Reasoning About Imp Programs

We'll get deeper into more systematic and powerful techniques for reasoning about Imp programs in *Programming Language Foundations*, but we can get some distance just working with the bare definitions. This section explores some examples.

```
Theorem plus2_spec : ∀ st n st',
  st X = n →
  st =[ plus2 ]=> st' →
  st' X = n + 2.
Proof.
  intros st n st' HX Heval.
```

Inverting *Heval* essentially forces Coq to expand one step of the **ceval** computation – in this case revealing that *st'* must be *st* extended with the new value of X, since plus2 is an assignment.

```
  inversion Heval. subst. clear Heval. simpl.
  apply t_update_eq. Qed.
```

Exercise: 3 stars, standard, recommended (XtimesYinZ_spec) State and prove a specification of XtimesYinZ.

```
Definition manual_grade_for_XtimesYinZ_spec : option (nat×string) := None.
```

□

Exercise: 3 stars, standard, recommended (loop_never_stops) Theorem loop_never_stops : ∀ *st st'*,

```
  ~(st =[ loop ]=> st').
Proof.
  intros st st' contra. unfold loop in contra.
  remember (WHILE true DO SKIP END)%imp as loopdef
           eqn:Heqloopdef.
```

Proceed by induction on the assumed derivation showing that *loopdef* terminates. Most of the cases are immediately contradictory (and so can be solved in one step with `discriminate`).

Admitted.

□

Exercise: 3 stars, standard (no_whiles_eqv) Consider the following function:

```
Open Scope imp_scope.
Fixpoint no_whiles (c : com) : bool :=
  match c with
  | SKIP ⇒
      true
  | _ ::= _ ⇒
      true
  | c1 ;; c2 ⇒
      andb (no_whiles c1) (no_whiles c2)
  | TEST _ THEN ct ELSE cf FI ⇒
      andb (no_whiles ct) (no_whiles cf)
  | WHILE _ DO _ END ⇒
      false
  end.
Close Scope imp_scope.
```

This predicate yields true just on programs that have no while loops. Using `Inductive`, write a property **no_whilesR** such that **no_whilesR** *c* is provable exactly when *c* is a program with no while loops. Then prove its equivalence with no_whiles.

```
Inductive no_whilesR: com → Prop :=

.

Theorem no_whiles_eqv:
   ∀ c, no_whiles c = true ↔ no_whilesR c.
Proof.
   Admitted.
```

□

Exercise: 4 stars, standard (no_whiles_terminating) Imp programs that don't involve while loops always terminate. State and prove a theorem *no_whiles_terminating* that says this.

Use either no_whiles or **no_whilesR**, as you prefer.

```
Definition manual_grade_for_no_whiles_terminating : option (nat×string) := None.
```

□

2.8 Additional Exercises

Exercise: 3 stars, standard (stack_compiler) Old HP Calculators, programming languages like Forth and Postscript, and abstract machines like the Java Virtual Machine all evaluate arithmetic expressions using a *stack*. For instance, the expression

(2*3)+(3*(4-2))

would be written as

2 3 * 3 4 2 - * +

and evaluated like this (where we show the program being evaluated on the right and the contents of the stack on the left):

| 2 3 * 3 4 2 - * + 2 | 3 * 3 4 2 - * + 3, 2 | * 3 4 2 - * + 6 | 3 4 2 - * + 3, 6 | 4 2 - * + 4, 3, 6 | 2 - * + 2, 4, 3, 6 | - * + 2, 3, 6 | * + 6, 6 | + 12 |

The goal of this exercise is to write a small compiler that translates **aexp**s into stack machine instructions.

The instruction set for our stack language will consist of the following instructions:

- SPush n: Push the number n on the stack.
- SLoad x: Load the identifier x from the store and push it on the stack
- SPlus: Pop the two top numbers from the stack, add them, and push the result onto the stack.
- SMinus: Similar, but subtract.
- SMult: Similar, but multiply.

```
Inductive sinstr : Type :=
| SPush (n : nat)
| SLoad (x : string)
| SPlus
| SMinus
| SMult.
```

Write a function to evaluate programs in the stack language. It should take as input a state, a stack represented as a list of numbers (top stack item is the head of the list), and a program represented as a list of instructions, and it should return the stack after executing the program. Test your function on the examples below.

Note that the specification leaves unspecified what to do when encountering an SPlus, SMinus, or SMult instruction if the stack contains less than two elements. In a sense, it is immaterial what we do, since our compiler will never emit such a malformed program.

```
Fixpoint s_execute (st : state) (stack : list nat)
                   (prog : list sinstr)
                 : list nat
  . Admitted.
```

```
Example s_execute1 :
     s_execute empty_st []
       [SPush 5; SPush 3; SPush 1; SMinus]
   = [2; 5].
   Admitted.
```

```
Example s_execute2 :
     s_execute (X !-> 3) [3;4]
       [SPush 4; SLoad X; SMult; SPlus]
   = [15; 4].
   Admitted.
```

Next, write a function that compiles an **aexp** into a stack machine program. The effect of running the program should be the same as pushing the value of the expression on the stack.

```
Fixpoint s_compile (e : aexp) : list sinstr
  . Admitted.
```

After you've defined s_compile, prove the following to test that it works.

```
Example s_compile1 :
  s_compile (X - (2 × Y))%imp
  = [SLoad X; SPush 2; SLoad Y; SMult; SMinus].
   Admitted.
```

□

Exercise: 4 stars, advanced (stack_compiler_correct) Now we'll prove the correctness of the compiler implemented in the previous exercise. Remember that the specification left unspecified what to do when encountering an SPlus, SMinus, or SMult instruction if the stack contains less than two elements. (In order to make your correctness proof easier you might find it helpful to go back and change your implementation!)

Prove the following theorem. You will need to start by stating a more general lemma to get a usable induction hypothesis; the main theorem will then be a simple corollary of this lemma.

```
Theorem s_compile_correct : ∀ (st : state) (e : aexp),
```

```
s_execute st [] (s_compile e) = [ aeval st e ].
Proof.
  Admitted.
  □
```

Exercise: 3 stars, standard, optional (short_circuit) Most modern programming languages use a "short-circuit" evaluation rule for boolean *and*: to evaluate BAnd *b1 b2*, first evaluate *b1*. If it evaluates to false, then the entire BAnd expression evaluates to false immediately, without evaluating *b2*. Otherwise, *b2* is evaluated to determine the result of the BAnd expression.

Write an alternate version of beval that performs short-circuit evaluation of BAnd in this manner, and prove that it is equivalent to beval. (N.b. This is only true because expression evaluation in Imp is rather simple. In a bigger language where evaluating an expression might diverge, the short-circuiting BAnd would *not* be equivalent to the original, since it would make more programs terminate.)

```
Module BreakImp.
```

Exercise: 4 stars, advanced (break_imp) Imperative languages like C and Java often include a *break* or similar statement for interrupting the execution of loops. In this exercise we consider how to add *break* to Imp. First, we need to enrich the language of commands with an additional case.

```
Inductive com : Type :=
  | CSkip
  | CBreak
  | CAss (x : string) (a : aexp)
  | CSeq (c1 c2 : com)
  | CIf (b : bexp) (c1 c2 : com)
  | CWhile (b : bexp) (c : com).
Notation "'SKIP'" :=
  CSkip.
Notation "'BREAK'" :=
  CBreak.
Notation "x '::=' a" :=
  (CAss x a) (at level 60).
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity).
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity).
Notation "'TEST' c1 'THEN' c2 'ELSE' c3 'FI'" :=
  (CIf c1 c2 c3) (at level 80, right associativity).
```

Next, we need to define the behavior of *BREAK*. Informally, whenever *BREAK* is executed in a sequence of commands, it stops the execution of that sequence and signals that the innermost enclosing loop should terminate. (If there aren't any enclosing loops, then the whole program simply terminates.) The final state should be the same as the one in which the *BREAK* statement was executed.

One important point is what to do when there are multiple loops enclosing a given *BREAK*. In those cases, *BREAK* should only terminate the *innermost* loop. Thus, after executing the following...

X ::= 0;; Y ::= 1;; WHILE ~(0 = Y) DO WHILE true DO BREAK END;; X ::= 1;; Y ::= Y - 1 END

... the value of X should be 1, and not 0.

One way of expressing this behavior is to add another parameter to the evaluation relation that specifies whether evaluation of a command executes a *BREAK* statement:

```
Inductive result : Type :=
  | SContinue
  | SBreak.
```

```
Reserved Notation "st '=[' c ']=>' st' '/' s"
         (at level 40, st' at next level).
```

Intuitively, *st* =[*c*]=> *st'* / s means that, if *c* is started in state *st*, then it terminates in state *st'* and either signals that the innermost surrounding loop (or the whole program) should exit immediately (s = SBreak) or that execution should continue normally (s = SContinue).

The definition of the "*st* =[*c*]=> *st'* / s" relation is very similar to the one we gave above for the regular evaluation relation (*st* =[*c*]=> *st'*) – we just need to handle the termination signals appropriately:

- If the command is *SKIP*, then the state doesn't change and execution of any enclosing loop can continue normally.
- If the command is *BREAK*, the state stays unchanged but we signal a SBreak.
- If the command is an assignment, then we update the binding for that variable in the state accordingly and signal that execution can continue normally.
- If the command is of the form *TEST b THEN c1 ELSE c2 FI*, then the state is updated as in the original semantics of Imp, except that we also propagate the signal from the execution of whichever branch was taken.
- If the command is a sequence *c1* ;; *c2*, we first execute *c1*. If this yields a SBreak, we skip the execution of *c2* and propagate the SBreak signal to the surrounding context; the resulting state is the same as the one obtained by executing *c1* alone. Otherwise, we execute *c2* on the state obtained after executing *c1*, and propagate the signal generated there.

- Finally, for a loop of the form *WHILE b DO c END*, the semantics is almost the same as before. The only difference is that, when b evaluates to true, we execute c and check the signal that it raises. If that signal is SContinue, then the execution proceeds as in the original semantics. Otherwise, we stop the execution of the loop, and the resulting state is the same as the one resulting from the execution of the current iteration. In either case, since *BREAK* only terminates the innermost loop, *WHILE* signals SContinue.

Based on the above description, complete the definition of the **ceval** relation.

```
Inductive ceval : com → state → result → state → Prop :=
  | E_Skip : ∀ st,
      st =[ CSkip ]=> st / SContinue


  where "st '=[' c ']=>' st' '/' s" := (ceval c st s st').
```

Now prove the following properties of your definition of **ceval**:

```
Theorem break_ignore : ∀ c st st' s,
     st =[ BREAK;; c ]=> st' / s →
     st = st'.
Proof.
   Admitted.

Theorem while_continue : ∀ b c st st' s,
  st =[ WHILE b DO c END ]=> st' / s →
  s = SContinue.
Proof.
   Admitted.

Theorem while_stops_on_break : ∀ b c st st',
  beval st b = true →
  st =[ c ]=> st' / SBreak →
  st =[ WHILE b DO c END ]=> st' / SContinue.
Proof.
   Admitted.
```

□

Exercise: 3 stars, advanced, optional (while_break_true)

```
Theorem while_break_true
: ∀ b c st st',
  st =[ WHILE b DO c END ]=> st' / SContinue →
  beval st' b = true →
  ∃ st'', st'' =[ c ]=> st' / SBreak.
Proof.
   Admitted.
```

□

Exercise: 4 stars, advanced, optional (ceval_deterministic) Theorem ceval_deterministic:
$\forall$ (c:**com**) st $st1$ $st2$ $s1$ $s2$,

```
    st =[ c ]=> st1 / s1 →
    st =[ c ]=> st2 / s2 →
    st1 = st2 ∧ s1 = s2.
```

Proof.

Admitted.

☐ End BreakImp.

Exercise: 4 stars, standard, optional (add_for_loop) Add C-style `for` loops to the language of commands, update the **ceval** definition to define the semantics of `for` loops, and add cases for `for` loops as needed so that all the proofs in this file are accepted by Coq.

A `for` loop should be parameterized by (a) a statement executed initially, (b) a test that is run on each iteration of the loop to determine whether the loop should continue, (c) a statement executed at the end of each loop iteration, and (d) a statement that makes up the body of the loop. (You don't need to worry about making up a concrete Notation for `for` loops, but feel free to play with this too if you like.)

Chapter 3

Preface

3.1 Welcome

This electronic book is a survey of basic concepts in the mathematical study of programs and programming languages. Topics include advanced use of the Coq proof assistant, operational semantics, Hoare logic, and static type systems. The exposition is intended for a broad range of readers, from advanced undergraduates to PhD students and researchers. No specific background in logic or programming languages is assumed, though a degree of mathematical maturity will be helpful.

As with all of the books in the *Software Foundations* series, this one is one hundred percent formalized and machine-checked: the entire text is literally a script for Coq. It is intended to be read alongside (or inside) an interactive session with Coq. All the details in the text are fully formalized in Coq, and most of the exercises are designed to be worked using Coq.

The files are organized into a sequence of core chapters, covering about one half semester's worth of material and organized into a coherent linear narrative, plus a number of "offshoot" chapters covering additional topics. All the core chapters are suitable for both upper-level undergraduate and graduate students.

The book builds on the material from *Logical Foundations* (*Software Foundations*, volume 1). It can be used together with that book for a one-semester course on the theory of programming languages. Or, for classes where students who are already familiar with some or all of the material in *Logical Foundations*, there is plenty of additional material to fill most of a semester from this book alone.

3.2 Overview

The book develops two main conceptual threads:

(1) formal techniques for *reasoning about the properties of specific programs* (e.g., the fact that a sorting function or a compiler obeys some formal specification); and

(2) the use of *type systems* for establishing well-behavedness guarantees for *all* programs in a given programming language (e.g., the fact that well-typed Java programs cannot be subverted at runtime).

Each of these is easily rich enough to fill a whole course in its own right, and tackling both together naturally means that much will be left unsaid. Nevertheless, we hope readers will find that these themes illuminate and amplify each other and that bringing them together creates a good foundation for digging into any of them more deeply. Some suggestions for further reading can be found in the Postscript chapter. Bibliographic information for all cited works can be found in the file Bib.

3.2.1 Program Verification

In the first part of the book, we introduce two broad topics of critical importance in building reliable software (and hardware): techniques for proving specific properties of particular *programs* and for proving general properties of whole programming *languages*.

For both of these, the first thing we need is a way of representing programs as mathematical objects, so we can talk about them precisely, plus ways of describing their behavior in terms of mathematical functions or relations. Our main tools for these tasks are *abstract syntax* and *operational semantics*, a method of specifying programming languages by writing abstract interpreters. At the beginning, we work with operational semantics in the so-called "big-step" style, which leads to simple and readable definitions when it is applicable. Later on, we switch to a lower-level "small-step" style, which helps make some useful distinctions (e.g., between different sorts of nonterminating program behaviors) and which is applicable to a broader range of language features, including concurrency.

The first programming language we consider in detail is *Imp*, a tiny toy language capturing the core features of conventional imperative programming: variables, assignment, conditionals, and loops.

We study two different ways of reasoning about the properties of Imp programs. First, we consider what it means to say that two Imp programs are *equivalent* in the intuitive sense that they exhibit the same behavior when started in any initial memory state. This notion of equivalence then becomes a criterion for judging the correctness of *metaprograms* – programs that manipulate other programs, such as compilers and optimizers. We build a simple optimizer for Imp and prove that it is correct.

Second, we develop a methodology for proving that a given Imp program satisfies some formal specifications of its behavior. We introduce the notion of *Hoare triples* – Imp programs annotated with pre- and post-conditions describing what they expect to be true about the memory in which they are started and what they promise to make true about the memory in which they terminate – and the reasoning principles of *Hoare Logic*, a domain-specific logic specialized for convenient compositional reasoning about imperative programs, with concepts like "loop invariant" built in.

This part of the course is intended to give readers a taste of the key ideas and mathematical tools used in a wide variety of real-world software and hardware verification tasks.

3.2.2 Type Systems

Our other major topic, covering approximately the second half of the book, is *type systems* – powerful tools for establishing properties of *all* programs in a given language.

Type systems are the best established and most popular example of a highly successful class of formal verification techniques known as *lightweight formal methods*. These are reasoning techniques of modest power – modest enough that automatic checkers can be built into compilers, linkers, or program analyzers and thus be applied even by programmers unfamiliar with the underlying theories. Other examples of lightweight formal methods include hardware and software model checkers, contract checkers, and run-time monitoring techniques.

This also completes a full circle with the beginning of the book: the language whose properties we study in this part, the *simply typed lambda-calculus*, is essentially a simplified model of the core of Coq itself!

3.2.3 Further Reading

This text is intended to be self contained, but readers looking for a deeper treatment of particular topics will find some suggestions for further reading in the Postscript chapter.

3.3 Note for Instructors

If you plan to use these materials in your own course, you will undoubtedly find things you'd like to change, improve, or add. Your contributions are welcome! Please see the Preface to *Logical Foundations* for instructions.

3.4 Thanks

Development of the *Software Foundations* series has been supported, in part, by the National Science Foundation under the NSF Expeditions grant 1521523, *The Science of Deep Specification.*

Chapter 4

Equiv: Program Equivalence

```
Set Warnings "-notation-overridden,-parsing".
From PLF Require Import Maps.
From Coq Require Import Bool.Bool.
From Coq Require Import Arith.Arith.
From Coq Require Import Init.Nat.
From Coq Require Import Arith.PeanoNat. Import Nat.
From Coq Require Import Arith.EqNat.
From Coq Require Import omega.Omega.
From Coq Require Import Lists.List.
From Coq Require Import Logic.FunctionalExtensionality.
Import ListNotations.
From PLF Require Import Imp.
```

Some Advice for Working on Exercises:

- Most of the Coq proofs we ask you to do are similar to proofs that we've provided. Before starting to work on exercises problems, take the time to work through our proofs (both informally and in Coq) and make sure you understand them in detail. This will save you a lot of time.

- The Coq proofs we're doing now are sufficiently complicated that it is more or less impossible to complete them by random experimentation or "following your nose." You need to start with an idea about why the property is true and how the proof is going to go. The best way to do this is to write out at least a sketch of an informal proof on paper – one that intuitively convinces you of the truth of the theorem – before starting to work on the formal one. Alternately, grab a friend and try to convince them that the theorem is true; then try to formalize your explanation.

- Use automation to save work! The proofs in this chapter can get pretty long if you try to write out all the cases explicitly.

4.1 Behavioral Equivalence

In an earlier chapter, we investigated the correctness of a very simple program transformation: the optimize_0plus function. The programming language we were considering was the first version of the language of arithmetic expressions – with no variables – so in that setting it was very easy to define what it means for a program transformation to be correct: it should always yield a program that evaluates to the same number as the original.

To talk about the correctness of program transformations for the full Imp language, in particular assignment, we need to consider the role of variables and state.

4.1.1 Definitions

For **aexp**s and **bexp**s with variables, the definition we want is clear: Two **aexp**s or **bexp**s are *behaviorally equivalent* if they evaluate to the same result in every state.

```
Definition aequiv (a1 a2 : aexp) : Prop :=
  ∀ (st : state),
    aeval st a1 = aeval st a2.
```

```
Definition bequiv (b1 b2 : bexp) : Prop :=
  ∀ (st : state),
    beval st b1 = beval st b2.
```

Here are some simple examples of equivalences of arithmetic and boolean expressions.

```
Theorem aequiv_example: aequiv (X - X) 0.
Proof.
  intros st. simpl. omega.
Qed.
```

```
Theorem bequiv_example: bequiv (X - X = 0)%imp true.
Proof.
  intros st. unfold beval.
  rewrite aequiv_example. reflexivity.
Qed.
```

For commands, the situation is a little more subtle. We can't simply say "two commands are behaviorally equivalent if they evaluate to the same ending state whenever they are started in the same initial state," because some commands, when run in some starting states, don't terminate in any final state at all! What we need instead is this: two commands are behaviorally equivalent if, for any given starting state, they either (1) both diverge or (2) both terminate in the same final state. A compact way to express this is "if the first one terminates in a particular state then so does the second, and vice versa."

```
Definition cequiv (c1 c2 : com) : Prop :=
  ∀ (st st' : state),
    (st =[ c1 ]=> st') ↔ (st =[ c2 ]=> st').
```

4.1.2 Simple Examples

For examples of command equivalence, let's start by looking at some trivial program transformations involving *SKIP*:

```
Theorem skip_left : ∀ c,
  cequiv
    (SKIP;; c)
    c.
Proof.
  intros c st st'.
  split; intros H.
  -
    inversion H; subst.
    inversion H2. subst.
    assumption.
  -
    apply E_Seq with st.
    apply E_Skip.
    assumption.
Qed.
```

Exercise: 2 stars, standard (skip_right) Prove that adding a *SKIP* after a command results in an equivalent program

```
Theorem skip_right : ∀ c,
  cequiv
    (c ;; SKIP)
    c.
Proof.
   Admitted.
```

□

Similarly, here is a simple transformation that optimizes *TEST* commands:

```
Theorem TEST_true_simple : ∀ c1 c2,
  cequiv
    (TEST true THEN c1 ELSE c2 FI)
    c1.
Proof.
  intros c1 c2.
  split; intros H.
  -
    inversion H; subst. assumption. inversion H5.
  -
    apply E_IfTrue. reflexivity. assumption. Qed.
```

Of course, no (human) programmer would write a conditional whose guard is literally true. But they might write one whose guard is *equivalent* to true:

Theorem: If b is equivalent to BTrue, then *TEST b THEN c1 ELSE c2 FI* is equivalent to *c1*. *Proof*:

- ($\rightarrow$) We must show, for all *st* and *st'*, that if *st* =[*TEST b THEN c1 ELSE c2 FI*]=> *st'* then *st* =[*c1*]=> *st'*.

  Proceed by cases on the rules that could possibly have been used to show *st* =[*TEST b THEN c1 ELSE c2 FI*]=> *st'*, namely E_IfTrue and E_IfFalse.

  - Suppose the final rule in the derivation of *st* =[*TEST b THEN c1 ELSE c2 FI*]=> *st'* was E_IfTrue. We then have, by the premises of E_IfTrue, that *st* =[*c1*]=> *st'*. This is exactly what we set out to prove.
  - On the other hand, suppose the final rule in the derivation of *st* =[*TEST b THEN c1 ELSE c2 FI*]=> *st'* was E_IfFalse. We then know that beval *st b* = false and *st* =[*c2*]=> *st'*.

    Recall that *b* is equivalent to BTrue, i.e., forall *st*, beval *st b* = beval *st* BTrue. In particular, this means that beval *st b* = true, since beval *st* BTrue = true. But this is a contradiction, since E_IfFalse requires that beval *st b* = false. Thus, the final rule could not have been E_IfFalse.

- ($\leftarrow$) We must show, for all *st* and *st'*, that if *st* =[*c1*]=> *st'* then *st* =[*TEST b THEN c1 ELSE c2 FI*]=> *st'*.

  Since *b* is equivalent to BTrue, we know that beval *st b* = beval *st* BTrue = true. Together with the assumption that *st* =[*c1*]=> *st'*, we can apply E_IfTrue to derive *st* =[*TEST b THEN c1 ELSE c2 FI*]=> *st'*. □

Here is the formal version of this proof:

```
Theorem TEST_true: ∀ b c1 c2,
  bequiv b BTrue →
  cequiv
    (TEST b THEN c1 ELSE c2 FI)
    c1.
Proof.
  intros b c1 c2 Hb.
  split; intros H.
  -
    inversion H; subst.
    +
      assumption.
    +
      unfold bequiv in Hb. simpl in Hb.
```

```
      rewrite Hb in H5.
      inversion H5.
  -
    apply E_IfTrue; try assumption.
    unfold bequiv in Hb. simpl in Hb.
    rewrite Hb. reflexivity. Qed.
```

Exercise: 2 stars, standard, recommended (TEST_false) Theorem TEST_false : ∀ *b c1 c2*,

```
  bequiv b BFalse →
  cequiv
    (TEST b THEN c1 ELSE c2 FI)
    c2.
Proof.
   Admitted.
```

☐

Exercise: 3 stars, standard (swap_if_branches) Show that we can swap the branches of an IF if we also negate its guard.

```
Theorem swap_if_branches : ∀ b e1 e2,
  cequiv
    (TEST b THEN e1 ELSE e2 FI)
    (TEST BNot b THEN e2 ELSE e1 FI).
Proof.
   Admitted.
```

☐

For *WHILE* loops, we can give a similar pair of theorems. A loop whose guard is equivalent to BFalse is equivalent to *SKIP*, while a loop whose guard is equivalent to BTrue is equivalent to *WHILE* BTrue *DO SKIP END* (or any other non-terminating program). The first of these facts is easy.

```
Theorem WHILE_false : ∀ b c,
  bequiv b BFalse →
  cequiv
    (WHILE b DO c END)
    SKIP.
Proof.
  intros b c Hb. split; intros H.
  -
    inversion H; subst.
    +
      apply E_Skip.
    +
```

```
      rewrite Hb in H2. inversion H2.
  -
    inversion H; subst.
    apply E_WhileFalse.
    rewrite Hb.
    reflexivity. Qed.
```

Exercise: 2 stars, advanced, optional (WHILE_false_informal) Write an informal proof of WHILE_false.

☐

To prove the second fact, we need an auxiliary lemma stating that *WHILE* loops whose guards are equivalent to BTrue never terminate.

Lemma: If *b* is equivalent to BTrue, then it cannot be the case that *st* =[*WHILE b DO c END*]=> *st'*.

Proof: Suppose that *st* =[*WHILE b DO c END*]=> *st'*. We show, by induction on a derivation of *st* =[*WHILE b DO c END*]=> *st'*, that this assumption leads to a contradiction. The only two cases to consider are E_WhileFalse and E_WhileTrue, the others are contradictory.

- Suppose *st* =[*WHILE b DO c END*]=> *st'* is proved using rule E_WhileFalse. Then by assumption beval *st b* = false. But this contradicts the assumption that *b* is equivalent to BTrue.

- Suppose *st* =[*WHILE b DO c END*]=> *st'* is proved using rule E_WhileTrue. We must have that:

  1. beval *st b* = true, 2. there is some *st0* such that *st* =[*c*]=> *st0* and *st0* =[*WHILE b DO c END*]=> *st'*, 3. and we are given the induction hypothesis that *st0* =[*WHILE b DO c END*]=> *st'* leads to a contradiction,

  We obtain a contradiction by 2 and 3. ☐

```
Lemma WHILE_true_nonterm : ∀ b c st st',
  bequiv b BTrue →
  ~( st =[ WHILE b DO c END ]=> st' ).
Proof.
  intros b c st st' Hb.
  intros H.
  remember (WHILE b DO c END)%imp as cw eqn:Heqcw.
  induction H;

  inversion Heqcw; subst; clear Heqcw.
  -
    unfold bequiv in Hb.
```

```
    rewrite Hb in H. inversion H.
  -
    apply IHceval2. reflexivity. Qed.
```

Exercise: 2 stars, standard, optional (WHILE_true_nonterm_informal) Explain what the lemma WHILE_true_nonterm means in English.

□

Exercise: 2 stars, standard, recommended (WHILE_true) Prove the following theorem. *Hint*: You'll want to use WHILE_true_nonterm here.

```
Theorem WHILE_true : ∀ b c,
  bequiv b true →
  cequiv
    (WHILE b DO c END)
    (WHILE true DO SKIP END).
Proof.
   Admitted.
```

□

A more interesting fact about *WHILE* commands is that any number of copies of the body can be "unrolled" without changing meaning. Loop unrolling is a common transformation in real compilers.

```
Theorem loop_unrolling : ∀ b c,
  cequiv
    (WHILE b DO c END)
    (TEST b THEN (c ;; WHILE b DO c END) ELSE SKIP FI).
Proof.
  intros b c st st'.
  split; intros Hce.
  -
    inversion Hce; subst.
    +
      apply E_IfFalse. assumption. apply E_Skip.
    +
      apply E_IfTrue. assumption.
      apply E_Seq with (st' := st'0). assumption. assumption.
  -
    inversion Hce; subst.
    +
      inversion H5; subst.
      apply E_WhileTrue with (st' := st'0).
      assumption. assumption. assumption.
    +
```

```
    inversion H5; subst. apply E_WhileFalse. assumption. Qed.
```

Exercise: 2 stars, standard, optional (seq_assoc) Theorem seq_assoc : ∀ c1 c2 c3,

```
  cequiv ((c1;;c2);;c3) (c1;;(c2;;c3)).
Proof.
  Admitted.
```

□

Proving program properties involving assignments is one place where the fact that program states are treated extensionally (e.g., x !-> m x ; m and m are equal maps) comes in handy.

```
Theorem identity_assignment : ∀ x,
  cequiv
    (x ::= x)
    SKIP.
Proof.
  intros.
  split; intro H; inversion H; subst.
  -
    rewrite t_update_same.
    apply E_Skip.
  -
    assert (Hx : st' =[ x ::= x ]=> (x !-> st' x ; st')).
    { apply E_Ass. reflexivity. }
    rewrite t_update_same in Hx.
    apply Hx.
Qed.
```

Exercise: 2 stars, standard, recommended (assign_aequiv) Theorem assign_aequiv : ∀ (x : string) e,

```
  aequiv x e →
  cequiv SKIP (x ::= e).
Proof.
  Admitted.
```

□

Exercise: 2 stars, standard (equiv_classes) Given the following programs, group together those that are equivalent in Imp. Your answer should be given as a list of lists, where each sub-list represents a group of equivalent programs. For example, if you think programs (a) through (h) are all equivalent to each other, but not to (i), your answer should look like this:

[prog_a;prog_b;prog_c;prog_d;prog_e;prog_f;prog_g;prog_h] ; [prog_i]

Write down your answer below in the definition of equiv_classes.

```
Definition prog_a : com :=
  (WHILE ~(X ≤ 0) DO
    X ::= X + 1
  END)%imp.

Definition prog_b : com :=
  (TEST X = 0 THEN
    X ::= X + 1;;
    Y ::= 1
  ELSE
    Y ::= 0
  FI;;
  X ::= X - Y;;
  Y ::= 0)%imp.

Definition prog_c : com :=
  SKIP%imp.

Definition prog_d : com :=
  (WHILE ~(X = 0) DO
    X ::= (X × Y) + 1
  END)%imp.

Definition prog_e : com :=
  (Y ::= 0)%imp.

Definition prog_f : com :=
  (Y ::= X + 1;;
  WHILE ~(X = Y) DO
    Y ::= X + 1
  END)%imp.

Definition prog_g : com :=
  (WHILE true DO
    SKIP
  END)%imp.

Definition prog_h : com :=
  (WHILE ~(X = X) DO
    X ::= X + 1
  END)%imp.

Definition prog_i : com :=
  (WHILE ~(X = Y) DO
    X ::= Y + 1
  END)%imp.

Definition equiv_classes : list (list com)
  . Admitted.
```

```
Definition manual_grade_for_equiv_classes : option (nat×string) := None.
```
□

4.2 Properties of Behavioral Equivalence

We next consider some fundamental properties of program equivalence.

4.2.1 Behavioral Equivalence Is an Equivalence

First, we verify that the equivalences on *aexps*, *bexps*, and **com**s really are *equivalences* – i.e., that they are reflexive, symmetric, and transitive. The proofs are all easy.

```
Lemma refl_aequiv : ∀ (a : aexp), aequiv a a.
Proof.
  intros a st. reflexivity. Qed.

Lemma sym_aequiv : ∀ (a1 a2 : aexp),
  aequiv a1 a2 → aequiv a2 a1.
Proof.
  intros a1 a2 H. intros st. symmetry. apply H. Qed.

Lemma trans_aequiv : ∀ (a1 a2 a3 : aexp),
  aequiv a1 a2 → aequiv a2 a3 → aequiv a1 a3.
Proof.
  unfold aequiv. intros a1 a2 a3 H12 H23 st.
  rewrite (H12 st). rewrite (H23 st). reflexivity. Qed.

Lemma refl_bequiv : ∀ (b : bexp), bequiv b b.
Proof.
  unfold bequiv. intros b st. reflexivity. Qed.

Lemma sym_bequiv : ∀ (b1 b2 : bexp),
  bequiv b1 b2 → bequiv b2 b1.
Proof.
  unfold bequiv. intros b1 b2 H. intros st. symmetry. apply H. Qed.

Lemma trans_bequiv : ∀ (b1 b2 b3 : bexp),
  bequiv b1 b2 → bequiv b2 b3 → bequiv b1 b3.
Proof.
  unfold bequiv. intros b1 b2 b3 H12 H23 st.
  rewrite (H12 st). rewrite (H23 st). reflexivity. Qed.

Lemma refl_cequiv : ∀ (c : com), cequiv c c.
Proof.
  unfold cequiv. intros c st st'. apply iff_refl. Qed.

Lemma sym_cequiv : ∀ (c1 c2 : com),
  cequiv c1 c2 → cequiv c2 c1.
```

```
Proof.
  unfold cequiv. intros c1 c2 H st st'.
  assert (st =[ c1 ]=> st' ↔ st =[ c2 ]=> st') as H'.
  { apply H. }
  apply iff_sym. assumption.
Qed.

Lemma iff_trans : ∀ (P1 P2 P3 : Prop),
  (P1 ↔ P2) → (P2 ↔ P3) → (P1 ↔ P3).
Proof.
  intros P1 P2 P3 H12 H23.
  inversion H12. inversion H23.
  split; intros A.
    apply H1. apply H. apply A.
    apply H0. apply H2. apply A. Qed.

Lemma trans_cequiv : ∀ (c1 c2 c3 : com),
  cequiv c1 c2 → cequiv c2 c3 → cequiv c1 c3.
Proof.
  unfold cequiv. intros c1 c2 c3 H12 H23 st st'.
  apply iff_trans with (st =[ c2 ]=> st'). apply H12. apply H23. Qed.
```

4.2.2 Behavioral Equivalence Is a Congruence

Less obviously, behavioral equivalence is also a *congruence*. That is, the equivalence of two subprograms implies the equivalence of the larger programs in which they are embedded:

aequiv a1 a1'

---

cequiv (x ::= a1) (x ::= a1')

cequiv c1 c1' cequiv c2 c2'

---

cequiv (c1;;c2) (c1';;c2')

... and so on for the other forms of commands.

(Note that we are using the inference rule notation here not as part of a definition, but simply to write down some valid implications in a readable format. We prove these implications below.)

We will see a concrete example of why these congruence properties are important in the following section (in the proof of fold_constants_com_sound), but the main idea is that they allow us to replace a small part of a large program with an equivalent small part and know that the whole large programs are equivalent *without* doing an explicit proof about the non-varying parts – i.e., the "proof burden" of a small change to a large program is proportional to the size of the change, not the program.

```
Theorem CAss_congruence : ∀ x a1 a1',
  aequiv a1 a1' →
```

```
  cequiv (CAss x a1) (CAss x a1').
Proof.
  intros x a1 a2 Heqv st st'.
  split; intros Hceval.
  -
    inversion Hceval. subst. apply E_Ass.
    rewrite Heqv. reflexivity.
  -
    inversion Hceval. subst. apply E_Ass.
    rewrite Heqv. reflexivity. Qed.
```

The congruence property for loops is a little more interesting, since it requires induction.

Theorem: Equivalence is a congruence for *WHILE* – that is, if *b1* is equivalent to *b1'* and *c1* is equivalent to *c1'*, then *WHILE b1 DO c1 END* is equivalent to *WHILE b1' DO c1' END*.

Proof: Suppose *b1* is equivalent to *b1'* and *c1* is equivalent to *c1'*. We must show, for every *st* and *st'*, that *st* =[*WHILE b1 DO c1 END*]=> *st'* iff *st* =[*WHILE b1' DO c1' END*]=> *st'*. We consider the two directions separately.

- (→) We show that *st* =[*WHILE b1 DO c1 END*]=> *st'* implies *st* =[*WHILE b1' DO c1' END*]=> *st'*, by induction on a derivation of *st* =[*WHILE b1 DO c1 END*]=> *st'*. The only nontrivial cases are when the final rule in the derivation is E_WhileFalse or E_WhileTrue.

  - E_WhileFalse: In this case, the form of the rule gives us beval *st b1* = false and *st* = *st'*. But then, since *b1* and *b1'* are equivalent, we have beval *st b1'* = false, and E_WhileFalse applies, giving us *st* =[*WHILE b1' DO c1' END*]=> *st'*, as required.

  - E_WhileTrue: The form of the rule now gives us beval *st b1* = true, with *st* =[*c1*]=> *st'0* and *st'0* =[*WHILE b1 DO c1 END*]=> *st'* for some state *st'0*, with the induction hypothesis *st'0* =[*WHILE b1' DO c1' END*]=> *st'*.

    Since *c1* and *c1'* are equivalent, we know that *st* =[*c1'*]=> *st'0*. And since *b1* and *b1'* are equivalent, we have beval *st b1'* = true. Now E_WhileTrue applies, giving us *st* =[*WHILE b1' DO c1' END*]=> *st'*, as required.

- (←) Similar. □

```
Theorem CWhile_congruence : ∀ b1 b1' c1 c1',
  bequiv b1 b1' → cequiv c1 c1' →
  cequiv (WHILE b1 DO c1 END) (WHILE b1' DO c1' END).
Proof.
  unfold bequiv,cequiv.
  intros b1 b1' c1 c1' Hb1e Hc1e st st'.
```

```
split; intros Hce.
-
  remember (WHILE b1 DO c1 END)%imp as cwhile
    eqn:Heqcwhile.
  induction Hce; inversion Heqcwhile; subst.
  +
    apply E_WhileFalse. rewrite ← Hb1e. apply H.
  +
    apply E_WhileTrue with (st' := st').
    × rewrite ← Hb1e. apply H.
    ×
      apply (Hc1e st st'). apply Hce1.
    ×
      apply IHHce2. reflexivity.
-
  remember (WHILE b1' DO c1' END)%imp as c'while
    eqn:Heqc'while.
  induction Hce; inversion Heqc'while; subst.
  +
    apply E_WhileFalse. rewrite → Hb1e. apply H.
  +
    apply E_WhileTrue with (st' := st').
    × rewrite → Hb1e. apply H.
    ×
      apply (Hc1e st st'). apply Hce1.
    ×
      apply IHHce2. reflexivity. Qed.
```

Exercise: 3 stars, standard, optional (CSeq_congruence) Theorem CSeq_congruence : ∀ c1 c1' c2 c2',

```
  cequiv c1 c1' → cequiv c2 c2' →
  cequiv (c1;;c2) (c1';;c2').
Proof.
   Admitted.
```

□

Exercise: 3 stars, standard (CIf_congruence) Theorem CIf_congruence : ∀ b b' c1 c1' c2 c2',

```
  bequiv b b' → cequiv c1 c1' → cequiv c2 c2' →
  cequiv (TEST b THEN c1 ELSE c2 FI)
         (TEST b' THEN c1' ELSE c2' FI).
Proof.
```

Admitted.

☐

For example, here are two equivalent programs and a proof of their equivalence...

```
Example congruence_example:
  cequiv

    (X ::= 0;;
     TEST X = 0
     THEN
       Y ::= 0
     ELSE
       Y ::= 42
     FI)

    (X ::= 0;;
     TEST X = 0
     THEN
       Y ::= X - X
     ELSE
       Y ::= 42
     FI).
Proof.
  apply CSeq_congruence.
  - apply refl_cequiv.
  - apply CIf_congruence.
    + apply refl_bequiv.
    + apply CAss_congruence. unfold aequiv. simpl.
      × symmetry. apply minus_diag.
    + apply refl_cequiv.
Qed.
```

Exercise: 3 stars, advanced, optional (not_congr) We've shown that the cequiv relation is both an equivalence and a congruence on commands. Can you think of a relation on commands that is an equivalence but *not* a congruence?

4.3 Program Transformations

A *program transformation* is a function that takes a program as input and produces some variant of the program as output. Compiler optimizations such as constant folding are a canonical example, but there are many others.

A program transformation is *sound* if it preserves the behavior of the original program.

```
Definition atrans_sound (atrans : aexp → aexp) : Prop :=
  ∀ (a : aexp),
    aequiv a (atrans a).
Definition btrans_sound (btrans : bexp → bexp) : Prop :=
  ∀ (b : bexp),
    bequiv b (btrans b).
Definition ctrans_sound (ctrans : com → com) : Prop :=
  ∀ (c : com),
    cequiv c (ctrans c).
```

4.3.1 The Constant-Folding Transformation

An expression is *constant* when it contains no variable references.

Constant folding is an optimization that finds constant expressions and replaces them by their values.

```
Fixpoint fold_constants_aexp (a : aexp) : aexp :=
  match a with
  | ANum n ⇒ ANum n
  | AId x ⇒ AId x
  | APlus a1 a2 ⇒
    match (fold_constants_aexp a1,
           fold_constants_aexp a2)
    with
    | (ANum n1, ANum n2) ⇒ ANum (n1 + n2)
    | (a1', a2') ⇒ APlus a1' a2'
    end
  | AMinus a1 a2 ⇒
    match (fold_constants_aexp a1,
           fold_constants_aexp a2)
    with
    | (ANum n1, ANum n2) ⇒ ANum (n1 - n2)
    | (a1', a2') ⇒ AMinus a1' a2'
    end
  | AMult a1 a2 ⇒
    match (fold_constants_aexp a1,
           fold_constants_aexp a2)
    with
    | (ANum n1, ANum n2) ⇒ ANum (n1 × n2)
    | (a1', a2') ⇒ AMult a1' a2'
    end
  end.
Example fold_aexp_ex1 :
```

```
    fold_constants_aexp ((1 + 2) × X)
  = (3 × X)%imp.
Proof. reflexivity. Qed.
```

Note that this version of constant folding doesn't eliminate trivial additions, etc. – we are focusing attention on a single optimization for the sake of simplicity. It is not hard to incorporate other ways of simplifying expressions; the definitions and proofs just get longer.

```
Example fold_aexp_ex2 :
  fold_constants_aexp (X - ((0 × 6) + Y))%imp = (X - (0 + Y))%imp.
Proof. reflexivity. Qed.
```

Not only can we lift fold_constants_aexp to **bexp**s (in the BEq and BLe cases); we can also look for constant *boolean* expressions and evaluate them in-place.

```
Fixpoint fold_constants_bexp (b : bexp) : bexp :=
  match b with
  | BTrue ⇒ BTrue
  | BFalse ⇒ BFalse
  | BEq a1 a2 ⇒
      match (fold_constants_aexp a1,
             fold_constants_aexp a2) with
      | (ANum n1, ANum n2) ⇒
          if n1 =? n2 then BTrue else BFalse
      | (a1', a2') ⇒
          BEq a1' a2'
      end
  | BLe a1 a2 ⇒
      match (fold_constants_aexp a1,
             fold_constants_aexp a2) with
      | (ANum n1, ANum n2) ⇒
          if n1 <=? n2 then BTrue else BFalse
      | (a1', a2') ⇒
          BLe a1' a2'
      end
  | BNot b1 ⇒
      match (fold_constants_bexp b1) with
      | BTrue ⇒ BFalse
      | BFalse ⇒ BTrue
      | b1' ⇒ BNot b1'
      end
  | BAnd b1 b2 ⇒
      match (fold_constants_bexp b1,
             fold_constants_bexp b2) with
      | (BTrue, BTrue) ⇒ BTrue
```

```
      | (BTrue, BFalse) ⇒ BFalse
      | (BFalse, BTrue) ⇒ BFalse
      | (BFalse, BFalse) ⇒ BFalse
      | (b1', b2') ⇒ BAnd b1' b2'
      end
  end.
Example fold_bexp_ex1 :
  fold_constants_bexp (true && ~(false && true))%imp
  = true.
Proof. reflexivity. Qed.
Example fold_bexp_ex2 :
  fold_constants_bexp ((X = Y) && (0 = (2 - (1 + 1))))%imp
  = ((X = Y) && true)%imp.
Proof. reflexivity. Qed.
```

To fold constants in a command, we apply the appropriate folding functions on all embedded expressions.

```
Open Scope imp.
Fixpoint fold_constants_com (c : com) : com :=
  match c with
  | SKIP ⇒
      SKIP
  | x ::= a ⇒
      x ::= (fold_constants_aexp a)
  | c1 ;; c2 ⇒
      (fold_constants_com c1) ;; (fold_constants_com c2)
  | TEST b THEN c1 ELSE c2 FI ⇒
      match fold_constants_bexp b with
      | BTrue ⇒ fold_constants_com c1
      | BFalse ⇒ fold_constants_com c2
      | b' ⇒ TEST b' THEN fold_constants_com c1
                     ELSE fold_constants_com c2 FI
      end
  | WHILE b DO c END ⇒
      match fold_constants_bexp b with
      | BTrue ⇒ WHILE BTrue DO SKIP END
      | BFalse ⇒ SKIP
      | b' ⇒ WHILE b' DO (fold_constants_com c) END
      end
  end.
Close Scope imp.
Example fold_com_ex1 :
```

```
  fold_constants_com
    (X ::= 4 + 5;;
     Y ::= X - 3;;
     TEST (X - Y) = (2 + 4) THEN SKIP
     ELSE Y ::= 0 FI;;
     TEST 0 ≤ (4 - (2 + 1)) THEN Y ::= 0
     ELSE SKIP FI;;
     WHILE Y = 0 DO
       X ::= X + 1
     END)%imp
  =
    (X ::= 9;;
     Y ::= X - 3;;
     TEST (X - Y) = 6 THEN SKIP
     ELSE Y ::= 0 FI;;
     Y ::= 0;;
     WHILE Y = 0 DO
       X ::= X + 1
     END)%imp.
Proof. reflexivity. Qed.
```

4.3.2 Soundness of Constant Folding

Now we need to show that what we've done is correct.

Here's the proof for arithmetic expressions:

```
Theorem fold_constants_aexp_sound :
  atrans_sound fold_constants_aexp.
Proof.
  unfold atrans_sound. intros a. unfold aequiv. intros st.
  induction a; simpl;

    try reflexivity;

    try (destruct (fold_constants_aexp a1);
         destruct (fold_constants_aexp a2);
         rewrite IHa1; rewrite IHa2; reflexivity). Qed.
```

Exercise: 3 stars, standard, optional (fold_bexp_Eq_informal) Here is an informal proof of the BEq case of the soundness argument for boolean expression constant folding. Read it carefully and compare it to the formal proof that follows. Then fill in the BLe case of the formal proof (without looking at the BEq case, if possible).

Theorem: The constant folding function for booleans, fold_constants_bexp, is sound.

Proof: We must show that b is equivalent to fold_constants_bexp b, for all boolean expressions b. Proceed by induction on b. We show just the case where b has the form BEq $a1$ $a2$.

In this case, we must show

beval st (BEq a1 a2) = beval st (fold_constants_bexp (BEq a1 a2)).

There are two cases to consider:

- First, suppose fold_constants_aexp $a1$ = ANum $n1$ and fold_constants_aexp $a2$ = ANum $n2$ for some $n1$ and $n2$.

  In this case, we have

  fold_constants_bexp (BEq a1 a2) = if n1 =? n2 then BTrue else BFalse

  and

  beval st (BEq a1 a2) = (aeval st a1) =? (aeval st a2).

  By the soundness of constant folding for arithmetic expressions (Lemma fold_constants_aexp_sound), we know

  aeval st a1 = aeval st (fold_constants_aexp a1) = aeval st (ANum n1) = n1

  and

  aeval st a2 = aeval st (fold_constants_aexp a2) = aeval st (ANum n2) = n2,

  so

  beval st (BEq a1 a2) = (aeval a1) =? (aeval a2) = n1 =? n2.

  Also, it is easy to see (by considering the cases $n1 = n2$ and $n1 \neq n2$ separately) that

  beval st (if n1 =? n2 then BTrue else BFalse) = if n1 =? n2 then beval st BTrue else beval st BFalse = if n1 =? n2 then true else false = n1 =? n2.

  So

  beval st (BEq a1 a2) = n1 =? n2. = beval st (if n1 =? n2 then BTrue else BFalse),

  as required.

- Otherwise, one of fold_constants_aexp $a1$ and fold_constants_aexp $a2$ is not a constant. In this case, we must show

  beval st (BEq a1 a2) = beval st (BEq (fold_constants_aexp a1) (fold_constants_aexp a2)),

  which, by the definition of beval, is the same as showing

  (aeval st a1) =? (aeval st a2) = (aeval st (fold_constants_aexp a1)) =? (aeval st (fold_constants_aexp a2)).

  But the soundness of constant folding for arithmetic expressions (fold_constants_aexp_sound) gives us aeval st a1 = aeval st (fold_constants_aexp a1) aeval st a2 = aeval st (fold_constants_aexp a2),

completing the case. □

```
Theorem fold_constants_bexp_sound:
  btrans_sound fold_constants_bexp.
Proof.
  unfold btrans_sound. intros b. unfold bequiv. intros st.
  induction b;

    try reflexivity.
  -
    simpl.
```

(Doing induction when there are a lot of constructors makes specifying variable names a chore, but Coq doesn't always choose nice variable names. We can rename entries in the context with the `rename` tactic: `rename a` *into a1* will change `a` to *a1* in the current goal and context.)

```
    remember (fold_constants_aexp a1) as a1' eqn:Heqa1'.
    remember (fold_constants_aexp a2) as a2' eqn:Heqa2'.
    replace (aeval st a1) with (aeval st a1') by
       (subst a1'; rewrite ← fold_constants_aexp_sound; reflexivity).
    replace (aeval st a2) with (aeval st a2') by
       (subst a2'; rewrite ← fold_constants_aexp_sound; reflexivity).
    destruct a1'; destruct a2'; try reflexivity.
      simpl. destruct (n =? n0); reflexivity.
  -
    admit.
  -
    simpl. remember (fold_constants_bexp b) as b' eqn:Heqb'.
    rewrite IHb.
    destruct b'; reflexivity.
  -
    simpl.
    remember (fold_constants_bexp b1) as b1' eqn:Heqb1'.
    remember (fold_constants_bexp b2) as b2' eqn:Heqb2'.
    rewrite IHb1. rewrite IHb2.
    destruct b1'; destruct b2'; reflexivity.
   Admitted.
```
□

Exercise: 3 stars, standard (fold_constants_com_sound) Complete the *WHILE* case of the following proof.

```
Theorem fold_constants_com_sound :
  ctrans_sound fold_constants_com.
Proof.
  unfold ctrans_sound. intros c.
  induction c; simpl.
  - apply refl_cequiv.
  - apply CAss_congruence.
              apply fold_constants_aexp_sound.
  - apply CSeq_congruence; assumption.
  -
    assert (bequiv b (fold_constants_bexp b)). {
      apply fold_constants_bexp_sound. }
    destruct (fold_constants_bexp b) eqn:Heqb;
      try (apply CIf_congruence; assumption).
    +
      apply trans_cequiv with c1; try assumption.
      apply TEST_true; assumption.
    +
      apply trans_cequiv with c2; try assumption.
      apply TEST_false; assumption.
  -
    Admitted.
```
□

Soundness of (0 + n) Elimination, Redux

Exercise: 4 stars, advanced, optional (optimize_0plus) Recall the definition optimize_0plus from the Imp chapter of *Logical Foundations*:

Fixpoint optimize_0plus (e:aexp) : aexp := match e with | ANum n => ANum n | APlus (ANum 0) e2 => optimize_0plus e2 | APlus e1 e2 => APlus (optimize_0plus e1) (optimize_0plus e2) | AMinus e1 e2 => AMinus (optimize_0plus e1) (optimize_0plus e2) | AMult e1 e2 => AMult (optimize_0plus e1) (optimize_0plus e2) end.

Note that this function is defined over the old **aexp**s, without states.

Write a new version of this function that accounts for variables, plus analogous ones for **bexp**s and commands:

optimize_0plus_aexp optimize_0plus_bexp optimize_0plus_com

Prove that these three functions are sound, as we did for *fold_constants_×*. Make sure you use the congruence lemmas in the proof of *optimize_0plus_com* – otherwise it will be *long*!

Then define an optimizer on commands that first folds constants (using fold_constants_com) and then eliminates 0 + n terms (using *optimize_0plus_com*).

- Give a meaningful example of this optimizer's output.

- Prove that the optimizer is sound. (This part should be *very* easy.)

4.4 Proving Inequivalence

Suppose that *c1* is a command of the form X ::= *a1*;; Y ::= *a2* and *c2* is the command X ::= *a1*;; Y ::= *a2'*, where *a2'* is formed by substituting *a1* for all occurrences of X in *a2*. For example, *c1* and *c2* might be:

c1 = (X ::= 42 + 53;; Y ::= Y + X) c2 = (X ::= 42 + 53;; Y ::= Y + (42 + 53))

Clearly, this *particular c1* and *c2* are equivalent. Is this true in general?

We will see in a moment that it is not, but it is worthwhile to pause, now, and see if you can find a counter-example on your own.

More formally, here is the function that substitutes an arithmetic expression *u* for each occurrence of a given variable x in another expression a:

```
Fixpoint subst_aexp (x : string) (u : aexp) (a : aexp) : aexp :=
  match a with
  | ANum n ⇒
      ANum n
  | AId x' ⇒
      if eqb_string x x' then u else AId x'
  | APlus a1 a2 ⇒
      APlus (subst_aexp x u a1) (subst_aexp x u a2)
  | AMinus a1 a2 ⇒
      AMinus (subst_aexp x u a1) (subst_aexp x u a2)
  | AMult a1 a2 ⇒
      AMult (subst_aexp x u a1) (subst_aexp x u a2)
  end.
```

```
Example subst_aexp_ex :
  subst_aexp X (42 + 53) (Y + X)%imp
  = (Y + (42 + 53))%imp.
Proof. reflexivity. Qed.
```

And here is the property we are interested in, expressing the claim that commands *c1* and *c2* as described above are always equivalent.

```
Definition subst_equiv_property := ∀ x1 x2 a1 a2,
  cequiv (x1 ::= a1;; x2 ::= a2)
         (x1 ::= a1;; x2 ::= subst_aexp x1 a1 a2).
```

Sadly, the property does *not* always hold – i.e., it is not the case that, for all *x1*, *x2*, *a1*, and *a2*,

cequiv (x1 ::= a1;; x2 ::= a2) (x1 ::= a1;; x2 ::= subst_aexp x1 a1 a2).

To see this, suppose (for a contradiction) that for all *x1*, *x2*, *a1*, and *a2*, we have

cequiv (x1 ::= a1;; x2 ::= a2) (x1 ::= a1;; x2 ::= subst_aexp x1 a1 a2).
Consider the following program:
X ::= X + 1;; Y ::= X
Note that
empty_st = X ::= X + 1;; Y ::= X => st1,
where *st1* = (Y !-> 1 ; X !-> 1).
By assumption, we know that
cequiv (X ::= X + 1;; Y ::= X) (X ::= X + 1;; Y ::= X + 1)
so, by the definition of cequiv, we have
empty_st = X ::= X + 1;; Y ::= X + 1 => st1.
But we can also derive
empty_st = X ::= X + 1;; Y ::= X + 1 => st2,
where *st2* = (Y !-> 2 ; X !-> 1). But *st1* ≠ *st2*, which is a contradiction, since **ceval** is deterministic! □

```
Theorem subst_inequiv :
  ¬ subst_equiv_property.
Proof.
  unfold subst_equiv_property.
  intros Contra.
  remember (X ::= X + 1;;
            Y ::= X)%imp
      as c1.
  remember (X ::= X + 1;;
            Y ::= X + 1)%imp
      as c2.
  assert (cequiv c1 c2) by (subst; apply Contra).
  remember (Y !-> 1 ; X !-> 1) as st1.
  remember (Y !-> 2 ; X !-> 1) as st2.
  assert (H1 : empty_st =[ c1 ]=> st1);
  assert (H2 : empty_st =[ c2 ]=> st2);
  try (subst;
       apply E_Seq with (st' := (X !-> 1));
       apply E_Ass; reflexivity).
  apply H in H1.
  assert (Hcontra : st1 = st2)
    by (apply (ceval_deterministic c2 empty_st); assumption).
  assert (Hcontra' : st1 Y = st2 Y)
    by (rewrite Hcontra; reflexivity).
  subst. inversion Hcontra'. Qed.
```

Exercise: 4 stars, standard, optional (better_subst_equiv) The equivalence we had in mind above was not complete nonsense – it was actually almost right. To make it correct, we just need to exclude the case where the variable X occurs in the right-hand-side of the first assignment statement.

```
Inductive var_not_used_in_aexp (x : string) : aexp → Prop :=
  | VNUNum : ∀ n, var_not_used_in_aexp x (ANum n)
  | VNUId : ∀ y, x ≠ y → var_not_used_in_aexp x (AId y)
  | VNUPlus : ∀ a1 a2,
      var_not_used_in_aexp x a1 →
      var_not_used_in_aexp x a2 →
      var_not_used_in_aexp x (APlus a1 a2)
  | VNUMinus : ∀ a1 a2,
      var_not_used_in_aexp x a1 →
      var_not_used_in_aexp x a2 →
      var_not_used_in_aexp x (AMinus a1 a2)
  | VNUMult : ∀ a1 a2,
      var_not_used_in_aexp x a1 →
      var_not_used_in_aexp x a2 →
      var_not_used_in_aexp x (AMult a1 a2).

Lemma aeval_weakening : ∀ x st a ni,
  var_not_used_in_aexp x a →
  aeval (x !-> ni ; st) a = aeval st a.
Proof.
   Admitted.
```

Using **var_not_used_in_aexp**, formalize and prove a correct version of subst_equiv_property.

Exercise: 3 stars, standard (inequiv_exercise) Prove that an infinite loop is not equivalent to *SKIP*

```
Theorem inequiv_exercise:
  ¬ cequiv (WHILE true DO SKIP END) SKIP.
Proof.
   Admitted.
```

□

4.5 Extended Exercise: Nondeterministic Imp

As we have seen (in theorem ceval_deterministic in the Imp chapter), Imp's evaluation relation is deterministic. However, *non*-determinism is an important part of the definition of many real programming languages. For example, in many imperative languages (such as C and its relatives), the order in which function arguments are evaluated is unspecified. The program fragment

x = 0;; f(++x, x)

might call f with arguments (1, 0) or (1, 1), depending how the compiler chooses to order things. This can be a little confusing for programmers, but it gives the compiler writer useful freedom.

In this exercise, we will extend Imp with a simple nondeterministic command and study how this change affects program equivalence. The new command has the syntax *HAVOC* X, where X is an identifier. The effect of executing *HAVOC* X is to assign an *arbitrary* number to the variable X, nondeterministically. For example, after executing the program:

HAVOC Y;; Z ::= Y * 2

the value of Y can be any number, while the value of Z is twice that of Y (so Z is always even). Note that we are not saying anything about the *probabilities* of the outcomes – just that there are (infinitely) many different outcomes that can possibly happen after executing this nondeterministic code.

In a sense, a variable on which we do *HAVOC* roughly corresponds to an uninitialized variable in a low-level language like C. After the *HAVOC*, the variable holds a fixed but arbitrary number. Most sources of nondeterminism in language definitions are there precisely because programmers don't care which choice is made (and so it is good to leave it open to the compiler to choose whichever will run faster).

We call this new language *Himp* ("Imp extended with *HAVOC*").

```
Module HIMP.
```

To formalize Himp, we first add a clause to the definition of commands.

```
Inductive com : Type :=
  | CSkip : com
  | CAss : string → aexp → com
  | CSeq : com → com → com
  | CIf : bexp → com → com → com
  | CWhile : bexp → com → com
  | CHavoc : string → com.

Notation "'SKIP'" :=
  CSkip : imp_scope.
Notation "X '::=' a" :=
  (CAss X a) (at level 60) : imp_scope.
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity) : imp_scope.
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity) : imp_scope.
Notation "'TEST' e1 'THEN' e2 'ELSE' e3 'FI'" :=
  (CIf e1 e2 e3) (at level 80, right associativity) : imp_scope.
Notation "'HAVOC' l" :=
```

(CHavoc l) (at level 60) : imp_scope.

Exercise: 2 stars, standard (himp_ceval) Now, we must extend the operational semantics. We have provided a template for the **ceval** relation below, specifying the big-step semantics. What rule(s) must be added to the definition of **ceval** to formalize the behavior of the *HAVOC* command?

```
Reserved Notation "st '=[' c ']=>' st'" (at level 40).

Open Scope imp_scope.
Inductive ceval : com → state → state → Prop :=
  | E_Skip : ∀ st,
      st =[ SKIP ]=> st
  | E_Ass : ∀ st a1 n x,
      aeval st a1 = n →
      st =[ x ::= a1 ]=> (x !-> n ; st)
  | E_Seq : ∀ c1 c2 st st' st'',
      st =[ c1 ]=> st' →
      st' =[ c2 ]=> st'' →
      st =[ c1 ;; c2 ]=> st''
  | E_IfTrue : ∀ st st' b c1 c2,
      beval st b = true →
      st =[ c1 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_IfFalse : ∀ st st' b c1 c2,
      beval st b = false →
      st =[ c2 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_WhileFalse : ∀ b st c,
      beval st b = false →
      st =[ WHILE b DO c END ]=> st
  | E_WhileTrue : ∀ st st' st'' b c,
      beval st b = true →
      st =[ c ]=> st' →
      st' =[ WHILE b DO c END ]=> st'' →
      st =[ WHILE b DO c END ]=> st''


  where "st =[ c ]=> st'" := (ceval c st st').
Close Scope imp_scope.
```

As a sanity check, the following claims should be provable for your definition:

```
Example havoc_example1 : empty_st =[ (HAVOC X)%imp ]=> (X !-> 0).
Proof.
```

Admitted.

```
Example havoc_example2 :
  empty_st =[ (SKIP;; HAVOC Z)%imp ]=> (Z !-> 42).
Proof.
  Admitted.
Definition manual_grade_for_Check_rule_for_HAVOC : option (nat×string) := None.
```

☐

Finally, we repeat the definition of command equivalence from above:

```
Definition cequiv (c1 c2 : com) : Prop := ∀ st st' : state,
  st =[ c1 ]=> st' ↔ st =[ c2 ]=> st'.
```

Let's apply this definition to prove some nondeterministic programs equivalent / inequivalent.

Exercise: 3 stars, standard (havoc_swap) Are the following two programs equivalent?

```
Definition pXY :=
  (HAVOC X;; HAVOC Y)%imp.
Definition pYX :=
  (HAVOC Y;; HAVOC X)%imp.
```

If you think they are equivalent, prove it. If you think they are not, prove that.

```
Theorem pXY_cequiv_pYX :
  cequiv pXY pYX ∨ ¬cequiv pXY pYX.
Proof. Admitted.
```

☐

Exercise: 4 stars, standard, optional (havoc_copy) Are the following two programs equivalent?

```
Definition ptwice :=
  (HAVOC X;; HAVOC Y)%imp.
Definition pcopy :=
  (HAVOC X;; Y ::= X)%imp.
```

If you think they are equivalent, then prove it. If you think they are not, then prove that. (Hint: You may find the `assert` tactic useful.)

```
Theorem ptwice_cequiv_pcopy :
  cequiv ptwice pcopy ∨ ¬cequiv ptwice pcopy.
Proof. Admitted.
```

☐

The definition of program equivalence we are using here has some subtle consequences on programs that may loop forever. What cequiv says is that the set of possible *terminating* outcomes of two equivalent programs is the same. However, in a language with nondeterminism, like Himp, some programs always terminate, some programs always diverge, and some programs can nondeterministically terminate in some runs and diverge in others. The final part of the following exercise illustrates this phenomenon.

Exercise: 4 stars, advanced (p1_p2_term) Consider the following commands:

```
Definition p1 : com :=
  (WHILE ¬ (X = 0) DO
    HAVOC Y;;
    X ::= X + 1
  END)%imp.

Definition p2 : com :=
  (WHILE ¬ (X = 0) DO
    SKIP
  END)%imp.
```

Intuitively, p1 and p2 have the same termination behavior: either they loop forever, or they terminate in the same state they started in. We can capture the termination behavior of p1 and p2 individually with these lemmas:

```
Lemma p1_may_diverge : ∀ st st', st X ≠ 0 →
  ¬ st =[ p1 ]=> st'.
Proof. Admitted.

Lemma p2_may_diverge : ∀ st st', st X ≠ 0 →
  ¬ st =[ p2 ]=> st'.
Proof.
  Admitted.
```

□

Exercise: 4 stars, advanced (p1_p2_equiv) Use these two lemmas to prove that p1 and p2 are actually equivalent.

```
Theorem p1_p2_equiv : cequiv p1 p2.
Proof. Admitted.
```

□

Exercise: 4 stars, advanced (p3_p4_inequiv) Prove that the following programs are *not* equivalent. (Hint: What should the value of Z be when p3 terminates? What about p4?)

```
Definition p3 : com :=
  (Z ::= 1;;
  WHILE ~(X = 0) DO
    HAVOC X;;
```

```
    HAVOC Z
  END)%imp.
```

```
Definition p4 : com :=
  (X ::= 0;;
  Z ::= 1)%imp.
```

```
Theorem p3_p4_inequiv : ¬ cequiv p3 p4.
Proof. Admitted.
```

□

Exercise: 5 stars, advanced, optional (p5_p6_equiv) Prove that the following commands are equivalent. (Hint: As mentioned above, our definition of cequiv for Himp only takes into account the sets of possible terminating configurations: two programs are equivalent if and only if the set of possible terminating states is the same for both programs when given a same starting state *st*. If p5 terminates, what should the final state be? Conversely, is it always possible to make p5 terminate?)

```
Definition p5 : com :=
  (WHILE ~(X = 1) DO
    HAVOC X
  END)%imp.
```

```
Definition p6 : com :=
  (X ::= 1)%imp.
```

```
Theorem p5_p6_equiv : cequiv p5 p6.
Proof. Admitted.
```

□

```
End Himp.
```

4.6 Additional Exercises

Exercise: 4 stars, standard, optional (for_while_equiv) This exercise extends the optional *add_for_loop* exercise from the Imp chapter, where you were asked to extend the language of commands with C-style `for` loops. Prove that the command:

for (c1 ; b ; c2) { c3 }

is equivalent to:

c1 ; WHILE b DO c3 ; c2 END

Exercise: 3 stars, standard, optional (swap_noninterfering_assignments) (Hint: You'll need functional_extensionality for this one.)

```
Theorem swap_noninterfering_assignments: ∀ l1 l2 a1 a2,
  l1 ≠ l2 →
  var_not_used_in_aexp l1 a2 →
```

```
  var_not_used_in_aexp l2 a1 →
  cequiv
    (l1 ::= a1;; l2 ::= a2)
    (l2 ::= a2;; l1 ::= a1).
Proof.
    Admitted.
```
□

Exercise: 4 stars, advanced, optional (capprox) In this exercise we define an asymmetric variant of program equivalence we call *program approximation*. We say that a program *c1 approximates* a program *c2* when, for each of the initial states for which *c1* terminates, *c2* also terminates and produces the same final state. Formally, program approximation is defined as follows:

```
Definition capprox (c1 c2 : com) : Prop := ∀ (st st' : state),
  st =[ c1 ]=> st' → st =[ c2 ]=> st'.
```

For example, the program

c1 = WHILE ~(X = 1) DO X ::= X - 1 END

approximates *c2* = X ::= 1, but *c2* does not approximate *c1* since *c1* does not terminate when X = 0 but *c2* does. If two programs approximate each other in both directions, then they are equivalent.

Find two programs c3 and c4 such that neither approximates the other.

```
Definition c3 : com
  . Admitted.
Definition c4 : com
  . Admitted.
Theorem c3_c4_different : ¬ capprox c3 c4 ∧ ¬ capprox c4 c3.
Proof. Admitted.
```

Find a program cmin that approximates every other program.

```
Definition cmin : com
  . Admitted.
Theorem cmin_minimal : ∀ c, capprox cmin c.
Proof. Admitted.
```

Finally, find a non-trivial property which is preserved by program approximation (when going from left to right).

```
Definition zprop (c : com) : Prop
  . Admitted.
Theorem zprop_preserving : ∀ c c',
  zprop c → capprox c c' → zprop c'.
Proof. Admitted.
```
□

Chapter 5

Hoare: Hoare Logic, Part I

```
Set Warnings "-notation-overridden,-parsing".
From PLF Require Import Maps.
From Coq Require Import Bool.Bool.
From Coq Require Import Arith.Arith.
From Coq Require Import Arith.EqNat.
From Coq Require Import Arith.PeanoNat. Import Nat.
From Coq Require Import omega.Omega.
From PLF Require Import Imp.
```

In the final chaper of *Logical Foundations* (*Software Foundations*, volume 1), we began applying the mathematical tools developed in the first part of the course to studying the theory of a small programming language, Imp.

- We defined a type of *abstract syntax trees* for Imp, together with an *evaluation relation* (a partial function on states) that specifies the *operational semantics* of programs.

  The language we defined, though small, captures some of the key features of full-blown languages like C, C++, and Java, including the fundamental notion of mutable state and some common control structures.

- We proved a number of *metatheoretic properties* – "meta" in the sense that they are properties of the language as a whole, rather than of particular programs in the language. These included:

  - determinism of evaluation
  - equivalence of some different ways of writing down the definitions (e.g., functional and relational definitions of arithmetic expression evaluation)
  - guaranteed termination of certain classes of programs
  - correctness (in the sense of preserving meaning) of a number of useful program transformations

- behavioral equivalence of programs (in the Equiv chapter).

If we stopped here, we would already have something useful: a set of tools for defining and discussing programming languages and language features that are mathematically precise, flexible, and easy to work with, applied to a set of key properties. All of these properties are things that language designers, compiler writers, and users might care about knowing. Indeed, many of them are so fundamental to our understanding of the programming languages we deal with that we might not consciously recognize them as "theorems." But properties that seem intuitively obvious can sometimes be quite subtle (sometimes also subtly wrong!).

We'll return to the theme of metatheoretic properties of whole languages later in this volume when we discuss *types* and *type soundness*. In this chapter, though, we turn to a different set of issues.

Our goal is to carry out some simple examples of *program verification* – i.e., to use the precise definition of Imp to prove formally that particular programs satisfy particular specifications of their behavior. We'll develop a reasoning system called *Floyd-Hoare Logic* – often shortened to just *Hoare Logic* – in which each of the syntactic constructs of Imp is equipped with a generic "proof rule" that can be used to reason compositionally about the correctness of programs involving this construct.

Hoare Logic originated in the 1960s, and it continues to be the subject of intensive research right up to the present day. It lies at the core of a multitude of tools that are being used in academia and industry to specify and verify real software systems.

Hoare Logic combines two beautiful ideas: a natural way of writing down *specifications* of programs, and a *compositional proof technique* for proving that programs are correct with respect to such specifications – where by "compositional" we mean that the structure of proofs directly mirrors the structure of the programs that they are about.

Overview of this chapter...

Topic:

- A systematic method for reasoning about the *functional correctness* of programs in Imp

Goals:

- a natural notation for *program specifications* and
- a *compositional* proof technique for program correctness

Plan:

- specifications (assertions / Hoare triples)
- proof rules
- loop invariants
- decorated programs
- examples

5.1 Assertions

To talk about specifications of programs, the first thing we need is a way of making *assertions* about properties that hold at particular points during a program's execution – i.e., claims about the current state of the memory when execution reaches that point. Formally, an assertion is just a family of propositions indexed by a state.

```
Definition Assertion := state → Prop.
```

Exercise: 1 star, standard, optional (assertions) Paraphrase the following assertions in English (or your favorite natural language).

```
Module ExAssertions.
Definition as1 : Assertion := fun st ⇒ st X = 3.
Definition as2 : Assertion := fun st ⇒ st X ≤ st Y.
Definition as3 : Assertion :=
  fun st ⇒ st X = 3 ∨ st X ≤ st Y.
Definition as4 : Assertion :=
  fun st ⇒ st Z × st Z ≤ st X ∧
              ¬ (((S (st Z)) × (S (st Z))) ≤ st X).
Definition as5 : Assertion := fun st ⇒ True.
Definition as6 : Assertion := fun st ⇒ False.
End ExAssertions.
```

□

This way of writing assertions can be a little bit heavy, for two reasons: (1) every single assertion that we ever write is going to begin with fun *st* ⇒ ; and (2) this state *st* is the only one that we ever use to look up variables in assertions (we will never need to talk about two different memory states at the same time). For discussing examples informally, we'll adopt some simplifying conventions: we'll drop the initial fun *st* ⇒, and we'll write just X to mean *st* X. Thus, instead of writing

fun st => (st Z) * (st Z) <= m /\ ~ ((S (st Z)) * (S (st Z)) <= m)

we'll write just

Z * Z <= m /\ ~((S Z) * (S Z) <= m).

This example also illustrates a convention that we'll use throughout the Hoare Logic chapters: in informal assertions, capital letters like X, Y, and Z are Imp variables, while lowercase letters like x, y, m, and n are ordinary Coq variables (of type **nat**). This is why, when translating from informal to formal, we replace X with *st* X but leave m alone.

Given two assertions *P* and *Q*, we say that *P implies Q*, written *P* ->> *Q*, if, whenever *P* holds in some state *st*, *Q* also holds.

```
Definition assert_implies (P Q : Assertion) : Prop :=
  ∀ st, P st → Q st.
```

```
Notation "P ->> Q" := (assert_implies P Q)
                        (at level 80) : hoare_spec_scope.
```

```
Open Scope hoare_spec_scope.
```

(The *hoare_spec_scope* annotation here tells Coq that this notation is not global but is intended to be used in particular contexts. The `Open Scope` tells Coq that this file is one such context.)

We'll also want the "iff" variant of implication between assertions:

```
Notation "P «-» Q" :=
  (P ->> Q ∧ Q ->> P) (at level 80) : hoare_spec_scope.
```

5.2 Hoare Triples

Next, we need a way of making formal claims about the behavior of commands.

In general, the behavior of a command is to transform one state to another, so it is natural to express claims about commands in terms of assertions that are true before and after the command executes:

- "If command c is started in a state satisfying assertion P, and if c eventually terminates in some final state, then this final state will satisfy the assertion Q."

Such a claim is called a *Hoare Triple*. The assertion P is called the *precondition* of c, while Q is the *postcondition*.

Formally:

```
Definition hoare_triple
           (P : Assertion) (c : com) (Q : Assertion) : Prop :=
  ∀ st st',
     st =[ c ]=> st' →
     P st →
     Q st'.
```

Since we'll be working a lot with Hoare triples, it's useful to have a compact notation:
[1] c [2].

(The traditional notation is $\{P\}\ c\ \{Q\}$, but single braces are already used for other things in Coq.)

```
Notation "{{ P }} c {{ Q }}" :=
  (hoare_triple P c Q) (at level 90, c at next level)
  : hoare_spec_scope.
```

Exercise: 1 star, standard, optional (triples) Paraphrase the following Hoare triples in English.

[1] P
[2] Q

1) [3] c [4]
2) [5] c [6]
3) [7] c [8]
4) [9] c [10]
5) [11] c [12]
6) [13] c [14]

Exercise: 1 star, standard, optional (valid_triples) Which of the following Hoare triples are *valid* – i.e., the claimed relation between P, c, and Q is true?

1) [15] X ::= 5 [16]
2) [17] X ::= X + 1 [18]
3) [19] X ::= 5;; Y ::= 0 [20]
4) [21] X ::= 5 [22]
5) [23] SKIP [24]
6) [25] SKIP [26]
7) [27] WHILE true DO SKIP END [28]
8) [29] WHILE X = 0 DO X ::= X + 1 END [30]

---

[3] `True`
[4] `X=5`
[5] `X=m`
[6] `X=m+5)`
[7] `X<=Y`
[8] `Y<=X`
[9] `True`
[10] `False`
[11] `X=m`
[12] `Y=real_factm`
[13] `X=m`
[14] `(Z*Z)<=m/\~(((SZ)*(SZ))<=m)`
[15] `True`
[16] `X=5`
[17] `X=2`
[18] `X=3`
[19] `True`
[20] `X=5`
[21] `X=2/\X=3`
[22] `X=0`
[23] `True`
[24] `False`
[25] `False`
[26] `True`
[27] `True`
[28] `False`
[29] `X=0`
[30] `X=1`

9) [31] WHILE ~(X = 0) DO X ::= X + 1 END [32]

To get us warmed up for what's coming, here are two simple facts about Hoare triples. (Make sure you understand what they mean.)

```
Theorem hoare_post_true : ∀ (P Q : Assertion) c,
  (∀ st, Q st) →
  {{P}} c {{Q}}.
Proof.
  intros P Q c H. unfold hoare_triple.
  intros st st' Heval HP.
  apply H. Qed.
```

```
Theorem hoare_pre_false : ∀ (P Q : Assertion) c,
  (∀ st, ¬ (P st)) →
  {{P}} c {{Q}}.
Proof.
  intros P Q c H. unfold hoare_triple.
  intros st st' Heval HP.
  unfold not in H. apply H in HP.
  inversion HP. Qed.
```

5.3 Proof Rules

The goal of Hoare logic is to provide a *compositional* method for proving the validity of specific Hoare triples. That is, we want the structure of a program's correctness proof to mirror the structure of the program itself. To this end, in the sections below, we'll introduce a rule for reasoning about each of the different syntactic forms of commands in Imp – one for assignment, one for sequencing, one for conditionals, etc. – plus a couple of "structural" rules for gluing things together. We will then be able to prove programs correct using these proof rules, without ever unfolding the definition of hoare_triple.

5.3.1 Assignment

The rule for assignment is the most fundamental of the Hoare logic proof rules. Here's how it works.

Consider this valid Hoare triple:

[33] X ::= Y [34]

In English: if we start out in a state where the value of Y is 1 and we assign Y to X, then we'll finish in a state where X is 1. That is, the property of being equal to 1 gets transferred

[31] X=1
[32] X=100
[33] Y=1
[34] X=1

from Y to X.

Similarly, in

[35] X ::= Y + Z [36]

the same property (being equal to one) gets transferred to X from the expression Y + Z on the right-hand side of the assignment.

More generally, if a is *any* arithmetic expression, then

[37] X ::= a [38]

is a valid Hoare triple.

Even more generally, to conclude that an arbitrary assertion *Q* holds after X ::= a, we need to assume that *Q* holds before X ::= a, but *with all occurrences of* X replaced by a in *Q*. This leads to the Hoare rule for assignment

[39] X ::= a [40]

where "*Q* [X |-> a]" is pronounced "*Q* where a is substituted for X".

For example, these are valid applications of the assignment rule:

[41] X ::= X + 1 [42]

[43] X ::= 3 [44]

[45] X ::= 3 [46]

To formalize the rule, we must first formalize the idea of "substituting an expression for an Imp variable in an assertion", which we refer to as assertion substitution, or assn_sub. That is, given a proposition *P*, a variable X, and an arithmetic expression a, we want to derive another proposition *P'* that is just the same as *P* except that *P'* should mention a wherever *P* mentions X.

Since *P* is an arbitrary Coq assertion, we can't directly "edit" its text. However, we can achieve the same effect by evaluating *P* in an updated state:

```
Definition assn_sub X a P : Assertion :=
  fun (st : state) ⇒
    P (X !-> aeval st a ; st).

Notation "P [ X |-> a ]" := (assn_sub X a P)
  (at level 10, X at next level).
```

That is, *P* [X |-> a] stands for an assertion – let's call it *P'* – that is just like *P* except that, wherever *P* looks up the variable X in the current state, *P'* instead uses the value of

[35] Y+Z=1
[36] X=1
[37] a=1
[38] X=1
[39] Q[X|->a]
[40] Q
[41] (X<=5)[X|->X+1]i.e.,X+1<=5
[42] X<=5
[43] (X=3)[X|->3]i.e.,3=3
[44] X=3
[45] (0<=X/\X<=5)[X|->3]i.e.,(0<=3/\3<=5)
[46] 0<=X/\X<=5

the expression a.

To see how this works, let's calculate what happens with a couple of examples. First, suppose *P'* is (X ≤ 5) [X |-> 3] – that is, more formally, *P'* is the Coq expression

fun st => (fun st' => st' X <= 5) (X !-> aeval st 3 ; st),

which simplifies to

fun st => (fun st' => st' X <= 5) (X !-> 3 ; st)

and further simplifies to

fun st => ((X !-> 3 ; st) X) <= 5

and finally to

fun st => 3 <= 5.

That is, *P'* is the assertion that 3 is less than or equal to 5 (as expected).

For a more interesting example, suppose *P'* is (X ≤ 5) [X |-> X + 1]. Formally, *P'* is the Coq expression

fun st => (fun st' => st' X <= 5) (X !-> aeval st (X + 1) ; st),

which simplifies to

fun st => (X !-> aeval st (X + 1) ; st) X <= 5

and further simplifies to

fun st => (aeval st (X + 1)) <= 5.

That is, *P'* is the assertion that X + 1 is at most 5.

Now, using the concept of substitution, we can give the precise proof rule for assignment:

(hoare_asgn) [47] X ::= a [48]

We can prove formally that this rule is indeed valid.

```
Theorem hoare_asgn : ∀ Q X a,
  {{Q [X |-> a]}} X ::= a {{Q}}.
Proof.
  unfold hoare_triple.
  intros Q X a st st' HE HQ.
  inversion HE. subst.
  unfold assn_sub in HQ. assumption. Qed.
```

Here's a first formal proof using this rule.

```
Example assn_sub_example :
  {{(fun st ⇒ st X < 5) [X |-> X + 1]}}
  X ::= X + 1
  {{fun st ⇒ st X < 5}}.
Proof.
  apply hoare_asgn. Qed.
```

Of course, what would be even more helpful is to prove this simpler triple:

---

[47] Q[X|->a]

[48] Q

[49] X ::= X + 1 [50]
We will see how to do so in the next section.

Exercise: 2 stars, standard (hoare_asgn_examples) Translate these informal Hoare triples...
1) [51] X ::= 2 * X [52]
2) [53] X ::= 3 [54]
...into formal statements (use the names *assn_sub_ex1* and *assn_sub_ex2*) and use hoare_asgn to prove them.

```
Definition manual_grade_for_hoare_asgn_examples : option (nat×string) := None.
```
□

Exercise: 2 stars, standard, recommended (hoare_asgn_wrong) The assignment rule looks backward to almost everyone the first time they see it. If it still seems puzzling, it may help to think a little about alternative "forward" rules. Here is a seemingly natural one:

(hoare_asgn_wrong) [55] X ::= a [56]

Give a counterexample showing that this rule is incorrect and argue informally that it is really a counterexample. (Hint: The rule universally quantifies over the arithmetic expression a, and your counterexample needs to exhibit an a for which the rule doesn't work.)

```
Definition manual_grade_for_hoare_asgn_wrong : option (nat×string) := None.
```
□

Exercise: 3 stars, advanced (hoare_asgn_fwd) However, by using a *parameter* m (a Coq number) to remember the original value of X we can define a Hoare rule for assignment that does, intuitively, "work forwards" rather than backwards.

(hoare_asgn_fwd) [57] X ::= a [58] (where st' = (X !-> m ; st))

Note that we use the original value of X to reconstruct the state *st'* before the assignment took place. Prove that this rule is correct. (Also note that this rule is more complicated than hoare_asgn.)

[49] `X<4`
[50] `X<5`
[51] `(X<=10)[X|->2*X]`
[52] `X<=10`
[53] `(0<=X/\X<=5)[X|->3]`
[54] `0<=X/\X<=5`
[55] `True`
[56] `X=a`
[57] `funst=>Pst/\stX=m`
[58] `funst=>Pst'/\stX=aevalst'a`

```
Theorem hoare_asgn_fwd :
  ∀ m a P,
  {{fun st ⇒ P st ∧ st X = m}}
    X ::= a
  {{fun st ⇒ P (X !-> m ; st)
             ∧ st X = aeval (X !-> m ; st) a }}.
Proof.
   Admitted.
   □
```

Exercise: 2 stars, advanced, optional (hoare_asgn_fwd_exists) Another way to define a forward rule for assignment is to existentially quantify over the previous value of the assigned variable. Prove that it is correct.

(hoare_asgn_fwd_exists) [59] X ::= a [60]

```
Theorem hoare_asgn_fwd_exists :
  ∀ a P,
  {{fun st ⇒ P st}}
    X ::= a
  {{fun st ⇒ ∃ m, P (X !-> m ; st) ∧
                  st X = aeval (X !-> m ; st) a }}.
Proof.
  intros a P.
   Admitted.
   □
```

5.3.2 Consequence

Sometimes the preconditions and postconditions we get from the Hoare rules won't quite be the ones we want in the particular situation at hand – they may be logically equivalent but have a different syntactic form that fails to unify with the goal we are trying to prove, or they actually may be logically weaker (for preconditions) or stronger (for postconditions) than what we need.

For instance, while

[61] X ::= 3 [62],

follows directly from the assignment rule,

[63] X ::= 3 [64]

---

[59] `funst=>Pst`

[60] `funst=>existsm,P(X!->m;st)/\stX=aeval(X!->m;st)a`

[61] `(X=3)[X|->3]`

[62] `X=3`

[63] `True`

[64] `X=3`

does not. This triple is valid, but it is not an instance of hoare_asgn because **True** and (X = 3) [X |-> 3] are not syntactically equal assertions. However, they are logically *equivalent*, so if one triple is valid, then the other must certainly be as well. We can capture this observation with the following rule:

[65] c [66] P - P'

(hoare_consequence_pre_equiv) [67] c [68]

Taking this line of thought a bit further, we can see that strengthening the precondition or weakening the postcondition of a valid triple always produces another valid triple. This observation is captured by two *Rules of Consequence*.

[69] c [70] P -» P'

(hoare_consequence_pre) [71] c [72]

[73] c [74] Q' -» Q

(hoare_consequence_post) [75] c [76]

Here are the formal versions:

```
Theorem hoare_consequence_pre : ∀ (P P' Q : Assertion) c,
  {{P'}} c {{Q}} →
  P ->> P' →
  {{P}} c {{Q}}.
Proof.
  intros P P' Q c Hhoare Himp.
  intros st st' Hc HP. apply (Hhoare st st').
  assumption. apply Himp. assumption. Qed.

Theorem hoare_consequence_post : ∀ (P Q Q' : Assertion) c,
  {{P}} c {{Q'}} →
  Q' ->> Q →
  {{P}} c {{Q}}.
Proof.
  intros P Q Q' c Hhoare Himp.
```

---

[65] P'
[66] Q
[67] P
[68] Q
[69] P'
[70] Q
[71] P
[72] Q
[73] P
[74] Q'
[75] P
[76] Q

```
  intros st st' Hc HP.
  apply Himp.
  apply (Hhoare st st').
  assumption. assumption. Qed.
```

For example, we can use the first consequence rule like this:
[77] -» [78] X ::= 1 [79]
Or, formally...

```
Example hoare_asgn_example1 :
  {{fun st ⇒ True}} X ::= 1 {{fun st ⇒ st X = 1}}.
Proof.
  apply hoare_consequence_pre
    with (P' := (fun st ⇒ st X = 1) [X |-> 1]).
  apply hoare_asgn.
  intros st H. unfold assn_sub, t_update. simpl. reflexivity.
Qed.
```

We can also use it to prove the example mentioned earlier.
[80] -» [81] X ::= X + 1 [82]
Or, formally ...

```
Example assn_sub_example2 :
  {{(fun st ⇒ st X < 4)}}
  X ::= X + 1
  {{fun st ⇒ st X < 5}}.
Proof.
  apply hoare_consequence_pre
    with (P' := (fun st ⇒ st X < 5) [X |-> X + 1]).
  apply hoare_asgn.
  intros st H. unfold assn_sub, t_update. simpl. omega.
Qed.
```

Finally, for convenience in proofs, here is a combined rule of consequence that allows us to vary both the precondition and the postcondition in one go.
[83] c [84] P -» P' Q' -» Q

---

[77] True
[78] 1=1
[79] X=1
[80] X<4
[81] (X<5)[X|->X+1]
[82] X<5
[83] P'
[84] Q'

(hoare_consequence) [85] c [86]

```
Theorem hoare_consequence : ∀ (P P' Q Q' : Assertion) c,
  {{P'}} c {{Q'}} →
  P ->> P' →
  Q' ->> Q →
  {{P}} c {{Q}}.
Proof.
  intros P P' Q Q' c Hht HPP' HQ'Q.
  apply hoare_consequence_pre with (P' := P').
  apply hoare_consequence_post with (Q' := Q').
  assumption. assumption. assumption. Qed.
```

5.3.3 Digression: The `eapply` Tactic

This is a good moment to take another look at the `eapply` tactic, which we introduced briefly in the *Auto* chapter of *Logical Foundations*.

We had to write "`with` (*P'* := ...)" explicitly in the proof of hoare_asgn_example1 and hoare_consequence above, to make sure that all of the metavariables in the premises to the hoare_consequence_pre rule would be set to specific values. (Since *P'* doesn't appear in the conclusion of hoare_consequence_pre, the process of unifying the conclusion with the current goal doesn't constrain *P'* to a specific assertion.)

This is annoying, both because the assertion is a bit long and also because, in hoare_asgn_example1, the very next thing we are going to do – applying the hoare_asgn rule – will tell us exactly what it should be! We can use `eapply` instead of `apply` to tell Coq, essentially, "Be patient: The missing part is going to be filled in later in the proof."

```
Example hoare_asgn_example1' :
  {{fun st ⇒ True}}
  X ::= 1
  {{fun st ⇒ st X = 1}}.
Proof.
  eapply hoare_consequence_pre.
  apply hoare_asgn.
  intros st H. reflexivity. Qed.
```

In general, the `eapply` *H* tactic works just like `apply` *H* except that, instead of failing if unifying the goal with the conclusion of *H* does not determine how to instantiate all of the variables appearing in the premises of *H*, `eapply` *H* will replace these variables with *existential variables* (written ?*nnn*), which function as placeholders for expressions that will be determined (by further unification) later in the proof.

In order for `Qed` to succeed, all existential variables need to be determined by the end of the proof. Otherwise Coq will (rightly) refuse to accept the proof. Remember that the

[85] P

[86] Q

Coq tactics build proof objects, and proof objects containing existential variables are not complete.

```
Lemma silly1 : ∀ (P : nat → nat → Prop) (Q : nat → Prop),
  (∀ x y : nat, P x y) →
  (∀ x y : nat, P x y → Q x) →
  Q 42.
Proof.
  intros P Q HP HQ. eapply HQ. apply HP.
```

Coq gives a warning after `apply HP`. ("All the remaining goals are on the shelf," means that we've finished all our top-level proof obligations but along the way we've put some aside to be done later, and we have not finished those.) Trying to close the proof with `Qed` gives an error. `Abort.`

An additional constraint is that existential variables cannot be instantiated with terms containing ordinary variables that did not exist at the time the existential variable was created. (The reason for this technical restriction is that allowing such instantiation would lead to inconsistency of Coq's logic.)

```
Lemma silly2 :
  ∀ (P : nat → nat → Prop) (Q : nat → Prop),
  (∃ y, P 42 y) →
  (∀ x y : nat, P x y → Q x) →
  Q 42.
Proof.
  intros P Q HP HQ. eapply HQ. destruct HP as [y HP'].
```

Doing `apply HP'` above fails with the following error:

Error: Impossible to unify "?175" with "y".

In this case there is an easy fix: doing `destruct HP` *before* doing `eapply HQ`. `Abort.`

```
Lemma silly2_fixed :
  ∀ (P : nat → nat → Prop) (Q : nat → Prop),
  (∃ y, P 42 y) →
  (∀ x y : nat, P x y → Q x) →
  Q 42.
Proof.
  intros P Q HP HQ. destruct HP as [y HP'].
  eapply HQ. apply HP'.
Qed.
```

The `apply HP'` in the last step unifies the existential variable in the goal with the variable `y`.

Note that the `assumption` tactic doesn't work in this case, since it cannot handle existential variables. However, Coq also provides an *eassumption* tactic that solves the goal if one of the premises matches the goal up to instantiations of existential variables. We can use it instead of `apply HP'` if we like.

```
Lemma silly2_eassumption : ∀ (P : nat → nat → Prop) (Q : nat → Prop),
  (∃ y, P 42 y) →
  (∀ x y : nat, P x y → Q x) →
  Q 42.
Proof.
  intros P Q HP HQ. destruct HP as [y HP']. eapply HQ. eassumption.
Qed.
```

Exercise: 2 stars, standard (hoare_asgn_examples_2) Translate these informal Hoare triples...

[87] X ::= X + 1 [88] [89] X ::= 3 [90]

...into formal statements (name them *assn_sub_ex1'* and *assn_sub_ex2'*) and use hoare_asgn and hoare_consequence_pre to prove them.

```
Definition manual_grade_for_hoare_asgn_examples_2 : option (nat×string) := None.
```

□

5.3.4 Skip

Since *SKIP* doesn't change the state, it preserves any assertion *P*:

(hoare_skip) [91] SKIP [92]

```
Theorem hoare_skip : ∀ P,
     {{P}} SKIP {{P}}.
Proof.
  intros P st st' H HP. inversion H. subst.
  assumption. Qed.
```

5.3.5 Sequencing

More interestingly, if the command *c1* takes any state where P holds to a state where *Q* holds, and if *c2* takes any state where *Q* holds to one where R holds, then doing *c1* followed by *c2* will take any state where P holds to one where R holds:

[93] c1 [94] [95] c2 [96]

[87] X+1<=5
[88] X<=5
[89] 0<=3/\3<=5
[90] 0<=X/\X<=5
[91] P
[92] P
[93] P
[94] Q
[95] Q
[96] R

(hoare_seq) [97] c1;;c2 [98]

```
Theorem hoare_seq : ∀ P Q R c1 c2,
     {{Q}} c2 {{R}} →
     {{P}} c1 {{Q}} →
     {{P}} c1;;c2 {{R}}.
Proof.
  intros P Q R c1 c2 H1 H2 st st' H12 Pre.
  inversion H12; subst.
  apply (H1 st'0 st'); try assumption.
  apply (H2 st st'0); assumption. Qed.
```

Note that, in the formal rule hoare_seq, the premises are given in backwards order (*c2* before *c1*). This matches the natural flow of information in many of the situations where we'll use the rule, since the natural way to construct a Hoare-logic proof is to begin at the end of the program (with the final postcondition) and push postconditions backwards through commands until we reach the beginning.

Informally, a nice way of displaying a proof using the sequencing rule is as a "decorated program" where the intermediate assertion Q is written between *c1* and *c2*:

[99] X ::= a;; [100] <— decoration for Q SKIP [101]

Here's an example of a program involving both assignment and sequencing.

```
Example hoare_asgn_example3 : ∀ a n,
  {{fun st ⇒ aeval st a = n}}
  X ::= a;; SKIP
  {{fun st ⇒ st X = n}}.
Proof.
  intros a n. eapply hoare_seq.
  -
    apply hoare_skip.
  -
    eapply hoare_consequence_pre. apply hoare_asgn.
    intros st H. subst. reflexivity.
Qed.
```

We typically use hoare_seq in conjunction with hoare_consequence_pre and the eapply tactic, as in this example.

Exercise: 2 stars, standard, recommended (hoare_asgn_example4) Translate this "decorated program" into a formal proof:

[97] P

[98] R

[99] a=n

[100] X=n

[101] X=n

[102] -» [103] X ::= 1;; [104] -» [105] Y ::= 2 [106]

(Note the use of "-»" decorations, each marking a use of hoare_consequence_pre.)

```
Example hoare_asgn_example4 :
  {{fun st ⇒ True}}
  X ::= 1;; Y ::= 2
  {{fun st ⇒ st X = 1 ∧ st Y = 2}}.
Proof.
  Admitted.
```

□

Exercise: 3 stars, standard (swap_exercise) Write an Imp program c that swaps the values of X and Y and show that it satisfies the following specification:

[107] c [108]

Your proof should not need to use unfold hoare_triple. (Hint: Remember that the assignment rule works best when it's applied "back to front," from the postcondition to the precondition. So your proof will want to start at the end and work back to the beginning of your program.)

```
Definition swap_program : com
  . Admitted.

Theorem swap_exercise :
  {{fun st ⇒ st X ≤ st Y}}
  swap_program
  {{fun st ⇒ st Y ≤ st X}}.
Proof.
  Admitted.
```

□

Exercise: 3 stars, standard (hoarestate1) Explain why the following proposition can't be proven:

forall (a : aexp) (n : nat), [109] X ::= 3;; Y ::= a [110].

```
Definition manual_grade_for_hoarestate1 : option (nat×string) := None.
```

□

---

[102] True
[103] 1=1
[104] X=1
[105] X=1/\2=2
[106] X=1/\Y=2
[107] X<=Y
[108] Y<=X
[109] funst=>aevalsta=n
[110] funst=>stY=n

5.3.6 Conditionals

What sort of rule do we want for reasoning about conditional commands?

Certainly, if the same assertion Q holds after executing either of the branches, then it holds after the whole conditional. So we might be tempted to write:

[111] c1 [112] [113] c2 [114]

---

[115] TEST b THEN c1 ELSE c2 [116]

However, this is rather weak. For example, using this rule, we cannot show

[117] TEST X = 0 THEN Y ::= 2 ELSE Y ::= X + 1 FI [118]

since the rule tells us nothing about the state in which the assignments take place in the "then" and "else" branches.

Fortunately, we can say something more precise. In the "then" branch, we know that the boolean expression b evaluates to true, and in the "else" branch, we know it evaluates to false. Making this information available in the premises of the rule gives us more information to work with when reasoning about the behavior of *c1* and *c2* (i.e., the reasons why they establish the postcondition Q).

[119] c1 [120] [121] c2 [122]

(hoare_if) [123] TEST b THEN c1 ELSE c2 FI [124]

To interpret this rule formally, we need to do a little work. Strictly speaking, the assertion we've written, $P \wedge b$, is the conjunction of an assertion and a boolean expression – i.e., it doesn't typecheck. To fix this, we need a way of formally "lifting" any bexp b to an assertion. We'll write bassn b for the assertion "the boolean expression b evaluates to true (in the given state)."

```
Definition bassn b : Assertion :=
  fun st ⇒ (beval st b = true).
```

A couple of useful facts about bassn:

```
Lemma bexp_eval_true : ∀ b st,
  beval st b = true → (bassn b) st.
```

---

[111] `P`
[112] `Q`
[113] `P`
[114] `Q`
[115] `P`
[116] `Q`
[117] `True`
[118] `X<=Y`
[119] `P/\b`
[120] `Q`
[121] `P/\~b`
[122] `Q`
[123] `P`
[124] `Q`

```
Proof.
  intros b st Hbe.
  unfold bassn. assumption. Qed.

Lemma bexp_eval_false : ∀ b st,
  beval st b = false → ¬ ((bassn b) st).
Proof.
  intros b st Hbe contra.
  unfold bassn in contra.
  rewrite → contra in Hbe. inversion Hbe. Qed.
```

Now we can formalize the Hoare proof rule for conditionals and prove it correct.

```
Theorem hoare_if : ∀ P Q b c1 c2,
  {{fun st ⇒ P st ∧ bassn b st}} c1 {{Q}} →
  {{fun st ⇒ P st ∧ ¬ (bassn b st)}} c2 {{Q}} →
  {{P}} TEST b THEN c1 ELSE c2 FI {{Q}}.
Proof.
  intros P Q b c1 c2 HTrue HFalse st st' HE HP.
  inversion HE; subst.
  -
    apply (HTrue st st').
      assumption.
      split. assumption.
      apply bexp_eval_true. assumption.
  -
    apply (HFalse st st').
      assumption.
      split. assumption.
      apply bexp_eval_false. assumption. Qed.
```

Example

Here is a formal proof that the program we used to motivate the rule satisfies the specification we gave.

```
Example if_example :
    {{fun st ⇒ True}}
  TEST X = 0
    THEN Y ::= 2
    ELSE Y ::= X + 1
  FI
    {{fun st ⇒ st X ≤ st Y}}.
Proof.
  apply hoare_if.
  -
```

```
    eapply hoare_consequence_pre. apply hoare_asgn.
    unfold bassn, assn_sub, t_update, assert_implies.
    simpl. intros st [_ H].
    apply eqb_eq in H.
    rewrite H. omega.
  -
    eapply hoare_consequence_pre. apply hoare_asgn.
    unfold assn_sub, t_update, assert_implies.
    simpl; intros st _. omega.
Qed.
```

Exercise: 2 stars, standard (if_minus_plus) Prove the following hoare triple using hoare_if. Do not use unfold hoare_triple.

```
Theorem if_minus_plus :
  {{fun st ⇒ True}}
  TEST X ≤ Y
    THEN Z ::= Y - X
    ELSE Y ::= X + Z
  FI
  {{fun st ⇒ st Y = st X + st Z}}.
Proof.
```

Admitted.

□

Exercise: One-sided conditionals

Exercise: 4 stars, standard (if1_hoare) In this exercise we consider extending Imp with "one-sided conditionals" of the form *IF1 b THEN c FI*. Here *b* is a boolean expression, and *c* is a command. If *b* evaluates to true, then command *c* is evaluated. If *b* evaluates to false, then *IF1 b THEN c FI* does nothing.

We recommend that you complete this exercise before attempting the ones that follow, as it should help solidify your understanding of the material.

The first step is to extend the syntax of commands and introduce the usual notations. (We've done this for you. We use a separate module to prevent polluting the global name space.)

```
Module IF1.

Inductive com : Type :=
  | CSkip : com
  | CAss : string → aexp → com
  | CSeq : com → com → com
  | CIf : bexp → com → com → com
  | CWhile : bexp → com → com
```

```
  | CIf1 : bexp → com → com.
Notation "'SKIP'" :=
  CSkip : imp_scope.
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity) : imp_scope.
Notation "X '::=' a" :=
  (CAss X a) (at level 60) : imp_scope.
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity) : imp_scope.
Notation "'TEST' e1 'THEN' e2 'ELSE' e3 'FI'" :=
  (CIf e1 e2 e3) (at level 80, right associativity) : imp_scope.
Notation "'IF1' b 'THEN' c 'FI'" :=
  (CIf1 b c) (at level 80, right associativity) : imp_scope.
```

Next we need to extend the evaluation relation to accommodate *IF1* branches. This is for you to do... What rule(s) need to be added to **ceval** to evaluate one-sided conditionals?

```
Reserved Notation "st '=[' c ']=>' st'" (at level 40).

Open Scope imp_scope.
Inductive ceval : com → state → state → Prop :=
  | E_Skip : ∀ st,
      st =[ SKIP ]=> st
  | E_Ass : ∀ st a1 n x,
      aeval st a1 = n →
      st =[ x ::= a1 ]=> (x !-> n ; st)
  | E_Seq : ∀ c1 c2 st st' st'',
      st =[ c1 ]=> st' →
      st' =[ c2 ]=> st'' →
      st =[ c1 ;; c2 ]=> st''
  | E_IfTrue : ∀ st st' b c1 c2,
      beval st b = true →
      st =[ c1 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_IfFalse : ∀ st st' b c1 c2,
      beval st b = false →
      st =[ c2 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_WhileFalse : ∀ b st c,
      beval st b = false →
      st =[ WHILE b DO c END ]=> st
  | E_WhileTrue : ∀ st st' st'' b c,
      beval st b = true →
      st =[ c ]=> st' →
```

```
    st' =[ WHILE b DO c END ]=> st'' →
    st =[ WHILE b DO c END ]=> st''

  where "st '=[' c ']=>' st'" := (ceval c st st').
Close Scope imp_scope.
```

Now we repeat (verbatim) the definition and notation of Hoare triples.

```
Definition hoare_triple
           (P : Assertion) (c : com) (Q : Assertion) : Prop :=
  ∀ st st',
     st =[ c ]=> st' →
     P st →
     Q st'.
Notation "{{ P }} c {{ Q }}" := (hoare_triple P c Q)
                                  (at level 90, c at next level)
                                  : hoare_spec_scope.
```

Finally, we (i.e., you) need to state and prove a theorem, *hoare_if1*, that expresses an appropriate Hoare logic proof rule for one-sided conditionals. Try to come up with a rule that is both sound and as precise as possible.

For full credit, prove formally hoare_if1_good that your rule is precise enough to show the following valid Hoare triple:

[125] IF1 ~(Y = 0) THEN X ::= X + Y FI [126]

Hint: Your proof of this triple may need to use the other proof rules also. Because we're working in a separate module, you'll need to copy here the rules you find necessary.

```
Lemma hoare_if1_good :
  {{ fun st ⇒ st X + st Y = st Z }}
  (IF1 ~(Y = 0) THEN
    X ::= X + Y
  FI)%imp
  {{ fun st ⇒ st X = st Z }}.
Proof. Admitted.
End If1.
Definition manual_grade_for_if1_hoare : option (nat×string) := None.
```

□

5.3.7 Loops

Finally, we need a rule for reasoning about while loops.

[125] X+Y=Z

[126] X=Z

Suppose we have a loop

WHILE b DO c END

and we want to find a precondition P and a postcondition Q such that

[127] WHILE b DO c END [128]

is a valid triple.

First of all, let's think about the case where b is false at the beginning – i.e., let's assume that the loop body never executes at all. In this case, the loop behaves like *SKIP*, so we might be tempted to write:

[129] WHILE b DO c END [130].

But, as we remarked above for the conditional, we know a little more at the end – not just P, but also the fact that b is false in the current state. So we can enrich the postcondition a little:

[131] WHILE b DO c END [132]

What about the case where the loop body *does* get executed? In order to ensure that P holds when the loop finally exits, we certainly need to make sure that the command c guarantees that P holds whenever c is finished. Moreover, since P holds at the beginning of the first execution of c, and since each execution of c re-establishes P when it finishes, we can always assume that P holds at the beginning of c. This leads us to the following rule:

[133] c [134]

---

[135] WHILE b DO c END [136]

This is almost the rule we want, but again it can be improved a little: at the beginning of the loop body, we know not only that P holds, but also that the guard b is true in the current state.

This gives us a little more information to use in reasoning about c (showing that it establishes the invariant by the time it finishes).

And this leads us to the final version of the rule:

[137] c [138]

(hoare_while) [139] WHILE b DO c END [140]

---

[127] `P`
[128] `Q`
[129] `P`
[130] `P`
[131] `P`
[132] `P/\~b`
[133] `P`
[134] `P`
[135] `P`
[136] `P/\~b`
[137] `P/\b`
[138] `P`
[139] `P`
[140] `P/\~b`

The proposition P is called an *invariant* of the loop.

```
Theorem hoare_while : ∀ P b c,
  {{fun st ⇒ P st ∧ bassn b st}} c {{P}} →
  {{P}} WHILE b DO c END {{fun st ⇒ P st ∧ ¬ (bassn b st)}}.
Proof.
  intros P b c Hhoare st st' He HP.
  remember (WHILE b DO c END)%imp as wcom eqn:Heqwcom.
  induction He;
    try (inversion Heqwcom); subst; clear Heqwcom.
  -
    split. assumption. apply bexp_eval_false. assumption.
  -
    apply IHHe2. reflexivity.
    apply (Hhoare st st'). assumption.
      split. assumption. apply bexp_eval_true. assumption.
Qed.
```

One subtlety in the terminology is that calling some assertion P a "loop invariant" doesn't just mean that it is preserved by the body of the loop in question (i.e., $\{\{P\}\}\ c\ \{\{P\}\}$, where c is the loop body), but rather that P *together with the fact that the loop's guard is true* is a sufficient precondition for c to ensure P as a postcondition.

This is a slightly (but importantly) weaker requirement. For example, if P is the assertion $X = 0$, then P *is* an invariant of the loop

WHILE X = 2 DO X := 1 END

although it is clearly *not* preserved by the body of the loop.

```
Example while_example :
    {{fun st ⇒ st X ≤ 3}}
  WHILE X ≤ 2
  DO X ::= X + 1 END
    {{fun st ⇒ st X = 3}}.
Proof.
  eapply hoare_consequence_post.
  apply hoare_while.
  eapply hoare_consequence_pre.
  apply hoare_asgn.
  unfold bassn, assn_sub, assert_implies, t_update. simpl.
    intros st [H1 H2]. apply leb_complete in H2. omega.
  unfold bassn, assert_implies. intros st [Hle Hb].
    simpl in Hb. destruct ((st X) <=? 2) eqn : Heqle.
    exfalso. apply Hb; reflexivity.
    apply leb_iff_conv in Heqle. omega.
Qed.
```

We can use the WHILE rule to prove the following Hoare triple...

```
Theorem always_loop_hoare : ∀ P Q,
  {{P}} WHILE true DO SKIP END {{Q}}.
Proof.
  intros P Q.
  apply hoare_consequence_pre with (P' := fun st : state ⇒ True).
  eapply hoare_consequence_post.
  apply hoare_while.
  -
    apply hoare_post_true. intros st. apply I.
  -
    simpl. intros st [Hinv Hguard].
    exfalso. apply Hguard. reflexivity.
  -
    intros st H. constructor. Qed.
```

Of course, this result is not surprising if we remember that the definition of hoare_triple asserts that the postcondition must hold *only* when the command terminates. If the command doesn't terminate, we can prove anything we like about the postcondition.

Hoare rules that only talk about what happens when commands terminate (without proving that they do) are often said to describe a logic of "partial" correctness. It is also possible to give Hoare rules for "total" correctness, which build in the fact that the commands terminate. However, in this course we will only talk about partial correctness.

Exercise: *REPEAT*

Exercise: 4 stars, advanced (hoare_repeat) In this exercise, we'll add a new command to our language of commands: *REPEAT* c *UNTIL* b *END*. You will write the evaluation rule for *REPEAT* and add a new Hoare rule to the language for programs involving it. (You may recall that the evaluation rule is given in an example in the *Auto* chapter. Try to figure it out yourself here rather than peeking.)

```
Module RepeatExercise.

Inductive com : Type :=
  | CSkip : com
  | CAsgn : string → aexp → com
  | CSeq : com → com → com
  | CIf : bexp → com → com → com
  | CWhile : bexp → com → com
  | CRepeat : com → bexp → com.
```

REPEAT behaves like *WHILE*, except that the loop guard is checked *after* each execution of the body, with the loop repeating as long as the guard stays *false*. Because of this, the body will always execute at least once.

```
Notation "'SKIP'" :=
```

```
  CSkip.
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity).
Notation "X '::=' a" :=
  (CAsgn X a) (at level 60).
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity).
Notation "'TEST' e1 'THEN' e2 'ELSE' e3 'FI'" :=
  (CIf e1 e2 e3) (at level 80, right associativity).
Notation "'REPEAT' e1 'UNTIL' b2 'END'" :=
  (CRepeat e1 b2) (at level 80, right associativity).
```

Add new rules for *REPEAT* to **ceval** below. You can use the rules for *WHILE* as a guide, but remember that the body of a *REPEAT* should always execute at least once, and that the loop ends when the guard becomes true.

```
Reserved Notation "st '=[' c ']=>' st'" (at level 40).

Inductive ceval : state → com → state → Prop :=
  | E_Skip : ∀ st,
      st =[ SKIP ]=> st
  | E_Ass : ∀ st a1 n x,
      aeval st a1 = n →
      st =[ x ::= a1 ]=> (x !-> n ; st)
  | E_Seq : ∀ c1 c2 st st' st'',
      st =[ c1 ]=> st' →
      st' =[ c2 ]=> st'' →
      st =[ c1 ;; c2 ]=> st''
  | E_IfTrue : ∀ st st' b c1 c2,
      beval st b = true →
      st =[ c1 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_IfFalse : ∀ st st' b c1 c2,
      beval st b = false →
      st =[ c2 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_WhileFalse : ∀ b st c,
      beval st b = false →
      st =[ WHILE b DO c END ]=> st
  | E_WhileTrue : ∀ st st' st'' b c,
      beval st b = true →
      st =[ c ]=> st' →
      st' =[ WHILE b DO c END ]=> st'' →
      st =[ WHILE b DO c END ]=> st''
```

```
where "st '=[' c ']=>' st'" := (ceval st c st').
```

A couple of definitions from above, copied here so they use the new **ceval**.

```
Definition hoare_triple (P : Assertion) (c : com) (Q : Assertion)
                        : Prop :=
  ∀ st st', st =[ c ]=> st' → P st → Q st'.

Notation "{{ P }} c {{ Q }}" :=
  (hoare_triple P c Q) (at level 90, c at next level).
```

To make sure you've got the evaluation rules for *REPEAT* right, prove that ex1_repeat evaluates correctly.

```
Definition ex1_repeat :=
  REPEAT
    X ::= 1;;
    Y ::= Y + 1
  UNTIL X = 1 END.

Theorem ex1_repeat_works :
  empty_st =[ ex1_repeat ]=> (Y !-> 1 ; X !-> 1).
Proof.
  Admitted.
```

Now state and prove a theorem, *hoare_repeat*, that expresses an appropriate proof rule for `repeat` commands. Use hoare_while as a model, and try to make your rule as precise as possible.

For full credit, make sure (informally) that your rule can be used to prove the following valid Hoare triple:

[141] REPEAT Y ::= X;; X ::= X - 1 UNTIL X = 0 END [142]

```
End RepeatExercise.

Definition manual_grade_for_hoare_repeat : option (nat×string) := None.
```

□

5.4 Summary

So far, we've introduced Hoare Logic as a tool for reasoning about Imp programs. The rules of Hoare Logic are:

(hoare_asgn) [143] X::=a [144]

[141] `X>0`
[142] `X=0/\Y>0`
[143] `Q[X|->a]`
[144] `Q`

(hoare_skip) [145] SKIP [146]

[147] c1 [148] [149] c2 [150]

(hoare_seq) [151] c1;;c2 [152]

[153] c1 [154] [155] c2 [156]

(hoare_if) [157] TEST b THEN c1 ELSE c2 FI [158]

[159] c [160]

(hoare_while) [161] WHILE b DO c END [162]

[163] c [164] P -» P' Q' -» Q

(hoare_consequence) [165] c [166]

In the next chapter, we'll see how these rules are used to prove that programs satisfy specifications of their behavior.

5.5 Additional Exercises

Exercise: 3 stars, standard (hoare_havoc) In this exercise, we will derive proof rules for a *HAVOC* command, which is similar to the nondeterministic *any* expression from the the Imp chapter.

---

[145] `P`
[146] `P`
[147] `P`
[148] `Q`
[149] `Q`
[150] `R`
[151] `P`
[152] `R`
[153] `P/\b`
[154] `Q`
[155] `P/\~b`
[156] `Q`
[157] `P`
[158] `Q`
[159] `P/\b`
[160] `P`
[161] `P`
[162] `P/\~b`
[163] `P'`
[164] `Q'`
[165] `P`
[166] `Q`

First, we enclose this work in a separate module, and recall the syntax and big-step semantics of Himp commands.

```
Module HIMP.

Inductive com : Type :=
  | CSkip : com
  | CAsgn : string → aexp → com
  | CSeq : com → com → com
  | CIf : bexp → com → com → com
  | CWhile : bexp → com → com
  | CHavoc : string → com.

Notation "'SKIP'" :=
  CSkip.
Notation "X '::=' a" :=
  (CAsgn X a) (at level 60).
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity).
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity).
Notation "'TEST' e1 'THEN' e2 'ELSE' e3 'FI'" :=
  (CIf e1 e2 e3) (at level 80, right associativity).
Notation "'HAVOC' X" := (CHavoc X) (at level 60).

Reserved Notation "st '=[' c ']=>' st'" (at level 40).

Inductive ceval : com → state → state → Prop :=
  | E_Skip : ∀ st,
      st =[ SKIP ]=> st
  | E_Ass : ∀ st a1 n x,
      aeval st a1 = n →
      st =[ x ::= a1 ]=> (x !-> n ; st)
  | E_Seq : ∀ c1 c2 st st' st'',
      st =[ c1 ]=> st' →
      st' =[ c2 ]=> st'' →
      st =[ c1 ;; c2 ]=> st''
  | E_IfTrue : ∀ st st' b c1 c2,
      beval st b = true →
      st =[ c1 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_IfFalse : ∀ st st' b c1 c2,
      beval st b = false →
      st =[ c2 ]=> st' →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> st'
  | E_WhileFalse : ∀ b st c,
```

```
      beval st b = false →
      st =[ WHILE b DO c END ]=> st
  | E_WhileTrue : ∀ st st' st'' b c,
      beval st b = true →
      st =[ c ]=> st' →
      st' =[ WHILE b DO c END ]=> st'' →
      st =[ WHILE b DO c END ]=> st''
  | E_Havoc : ∀ st X n,
      st =[ HAVOC X ]=> (X !-> n ; st)

where "st '=[' c ']=>' st'" := (ceval c st st').
```

The definition of Hoare triples is exactly as before.

```
Definition hoare_triple (P:Assertion) (c:com) (Q:Assertion) : Prop :=
  ∀ st st', st =[ c ]=> st' → P st → Q st'.

Notation "{{ P }} c {{ Q }}" := (hoare_triple P c Q)
                                  (at level 90, c at next level)
                                  : hoare_spec_scope.
```

Complete the Hoare rule for *HAVOC* commands below by defining havoc_pre and prove that the resulting rule is correct.

```
Definition havoc_pre (X : string) (Q : Assertion) : Assertion
  . Admitted.

Theorem hoare_havoc : ∀ (Q : Assertion) (X : string),
  {{ havoc_pre X Q }} HAVOC X {{ Q }}.
Proof.
  Admitted.

End Himp.
```

□

Exercise: 4 stars, standard, optional (assert_vs_assume) Module HoareAssertAssume.

In this exercise, we will extend IMP with two commands, *ASSERT* and *ASSUME*. Both commands are ways to indicate that a certain statement should hold any time this part of the program is reached. However they differ as follows:

- If an *ASSERT* statement fails, it causes the program to go into an error state and exit.
- If an *ASSUME* statement fails, the program fails to evaluate at all. In other words, the program gets stuck and has no final state.

The new set of commands is:

```
Inductive com : Type :=
```

```
| CSkip : com
| CAss : string → aexp → com
| CSeq : com → com → com
| CIf : bexp → com → com → com
| CWhile : bexp → com → com
| CAssert : bexp → com
| CAssume : bexp → com.
Notation "'SKIP'" :=
  CSkip.
Notation "x '::=' a" :=
  (CAss x a) (at level 60).
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity).
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity).
Notation "'TEST' c1 'THEN' c2 'ELSE' c3 'FI'" :=
  (CIf c1 c2 c3) (at level 80, right associativity).
Notation "'ASSERT' b" :=
  (CAssert b) (at level 60).
Notation "'ASSUME' b" :=
  (CAssume b) (at level 60).
```

To define the behavior of *ASSERT* and *ASSUME*, we need to add notation for an error, which indicates that an assertion has failed. We modify the **ceval** relation, therefore, so that it relates a start state to either an end state or to *error*. The **result** type indicates the end value of a program, either a state or an error:

```
Inductive result : Type :=
  | RNormal : state → result
  | RError : result.
```

Now we are ready to give you the ceval relation for the new language.

```
Inductive ceval : com → state → result → Prop :=

  | E_Skip : ∀ st,
      st =[ SKIP ]=> RNormal st
  | E_Ass : ∀ st a1 n x,
      aeval st a1 = n →
      st =[ x ::= a1 ]=> RNormal (x !-> n ; st)
  | E_SeqNormal : ∀ c1 c2 st st' r,
      st =[ c1 ]=> RNormal st' →
      st' =[ c2 ]=> r →
      st =[ c1 ;; c2 ]=> r
  | E_SeqError : ∀ c1 c2 st,
```

```
      st =[ c1 ]=> RError →
      st =[ c1 ;; c2 ]=> RError
  | E_IfTrue : ∀ st r b c1 c2,
      beval st b = true →
      st =[ c1 ]=> r →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> r
  | E_IfFalse : ∀ st r b c1 c2,
      beval st b = false →
      st =[ c2 ]=> r →
      st =[ TEST b THEN c1 ELSE c2 FI ]=> r
  | E_WhileFalse : ∀ b st c,
      beval st b = false →
      st =[ WHILE b DO c END ]=> RNormal st
  | E_WhileTrueNormal : ∀ st st' r b c,
      beval st b = true →
      st =[ c ]=> RNormal st' →
      st' =[ WHILE b DO c END ]=> r →
      st =[ WHILE b DO c END ]=> r
  | E_WhileTrueError : ∀ st b c,
      beval st b = true →
      st =[ c ]=> RError →
      st =[ WHILE b DO c END ]=> RError

  | E_AssertTrue : ∀ st b,
      beval st b = true →
      st =[ ASSERT b ]=> RNormal st
  | E_AssertFalse : ∀ st b,
      beval st b = false →
      st =[ ASSERT b ]=> RError
  | E_Assume : ∀ st b,
      beval st b = true →
      st =[ ASSUME b ]=> RNormal st

where "st '=[' c ']=>' r" := (ceval c st r).
```

We redefine hoare triples: Now, {{P}} c {{Q}} means that, whenever c is started in a state satisfying P, and terminates with result r, then r is not an error and the state of r satisfies Q.

```
Definition hoare_triple
           (P : Assertion) (c : com) (Q : Assertion) : Prop :=
  ∀ st r,
     st =[ c ]=> r → P st →
     (∃ st', r = RNormal st' ∧ Q st').
```

```
Notation "{{ P }} c {{ Q }}" :=
  (hoare_triple P c Q) (at level 90, c at next level)
  : hoare_spec_scope.
```

To test your understanding of this modification, give an example precondition and postcondition that are satisfied by the *ASSUME* statement but not by the *ASSERT* statement. Then prove that any triple for *ASSERT* also works for *ASSUME*.

```
Theorem assert_assume_differ : ∃ P b Q,
       ({{P}} ASSUME b {{Q}})
  ∧ ¬ ({{P}} ASSERT b {{Q}}).
Proof.
   Admitted.

Theorem assert_implies_assume : ∀ P b Q,
     ({{P}} ASSERT b {{Q}})
  → ({{P}} ASSUME b {{Q}}).
Proof.
   Admitted.
```

Your task is now to state Hoare rules for *ASSERT* and *ASSUME*, and use them to prove a simple program correct. Name your hoare rule theorems *hoare_assert* and *hoare_assume*.

For your benefit, we provide proofs for the old hoare rules adapted to the new semantics.

```
Theorem hoare_asgn : ∀ Q X a,
  {{Q [X |-> a]}} X ::= a {{Q}}.
Proof.
  unfold hoare_triple.
  intros Q X a st st' HE HQ.
  inversion HE. subst.
  ∃ (X !-> aeval st a ; st). split; try reflexivity.
  assumption. Qed.

Theorem hoare_consequence_pre : ∀ (P P' Q : Assertion) c,
  {{P'}} c {{Q}} →
  P ->> P' →
  {{P}} c {{Q}}.
Proof.
  intros P P' Q c Hhoare Himp.
  intros st st' Hc HP. apply (Hhoare st st').
  assumption. apply Himp. assumption. Qed.

Theorem hoare_consequence_post : ∀ (P Q Q' : Assertion) c,
  {{P}} c {{Q'}} →
  Q' ->> Q →
  {{P}} c {{Q}}.
Proof.
  intros P Q Q' c Hhoare Himp.
```

```
  intros st r Hc HP.
  unfold hoare_triple in Hhoare.
  assert (∃ st', r = RNormal st' ∧ Q' st').
  { apply (Hhoare st); assumption. }
  destruct H as [st' [Hr HQ']].
  ∃ st'. split; try assumption.
  apply Himp. assumption.
Qed.

Theorem hoare_seq : ∀ P Q R c1 c2,
  {{Q}} c2 {{R}} →
  {{P}} c1 {{Q}} →
  {{P}} c1;;c2 {{R}}.
Proof.
  intros P Q R c1 c2 H1 H2 st r H12 Pre.
  inversion H12; subst.
  - eapply H1.
    + apply H6.
    + apply H2 in H3. apply H3 in Pre.
        destruct Pre as [st'0 [Heq HQ]].
        inversion Heq; subst. assumption.
  -
      apply H2 in H5. apply H5 in Pre.
      destruct Pre as [st' [C _]].
      inversion C.
Qed.
```

State and prove your hoare rules, *hoare_assert* and *hoare_assume*, below.

Here are the other proof rules (sanity check)

```
Theorem hoare_skip : ∀ P,
    {{P}} SKIP {{P}}.
Proof.
  intros P st st' H HP. inversion H. subst.
  eexists. split. reflexivity. assumption.
Qed.

Theorem hoare_if : ∀ P Q b c1 c2,
  {{fun st ⇒ P st ∧ bassn b st}} c1 {{Q}} →
  {{fun st ⇒ P st ∧ ¬ (bassn b st)}} c2 {{Q}} →
  {{P}} TEST b THEN c1 ELSE c2 FI {{Q}}.
Proof.
  intros P Q b c1 c2 HTrue HFalse st st' HE HP.
  inversion HE; subst.
  -
    apply (HTrue st st').
```

```
      assumption.
      split. assumption.
      apply bexp_eval_true. assumption.
  -
    apply (HFalse st st').
      assumption.
      split. assumption.
      apply bexp_eval_false. assumption. Qed.
Theorem hoare_while : ∀ P b c,
  {{fun st ⇒ P st ∧ bassn b st}} c {{P}} →
  {{P}} WHILE b DO c END {{fun st ⇒ P st ∧ ¬ (bassn b st)}}.
Proof.
  intros P b c Hhoare st st' He HP.
  remember (WHILE b DO c END) as wcom eqn:Heqwcom.
  induction He;
    try (inversion Heqwcom); subst; clear Heqwcom.
  -
    eexists. split. reflexivity. split.
    assumption. apply bexp_eval_false. assumption.
  -
    clear IHHe1.
    apply IHHe2. reflexivity.
    clear IHHe2 He2 r.
    unfold hoare_triple in Hhoare.
    apply Hhoare in He1.
    + destruct He1 as [st1 [Heq Hst1]].
        inversion Heq; subst.
        assumption.
    + split; assumption.
  -
     exfalso. clear IHHe.
     unfold hoare_triple in Hhoare.
     apply Hhoare in He.
     + destruct He as [st' [C _]]. inversion C.
     + split; assumption.
Qed.
Example assert_assume_example:
  {{fun st ⇒ True}}
  ASSUME (X = 1);;
  X ::= X + 1;;
  ASSERT (X = 2)
  {{fun st ⇒ True}}.
```

`Proof`.

Admitted.

`End` HoareAssertAssume.

□

Chapter 6

Hoare2: Hoare Logic, Part II

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Strings.String.
From PLF Require Import Maps.
From Coq Require Import Bool.Bool.
From Coq Require Import Arith.Arith.
From Coq Require Import Arith.EqNat.
From Coq Require Import Arith.PeanoNat. Import Nat.
From Coq Require Import omega.Omega.
From PLF Require Import Hoare.
From PLF Require Import Imp.
```

6.1 Decorated Programs

The beauty of Hoare Logic is that it is *compositional*: the structure of proofs exactly follows the structure of programs.

This suggests that we can record the essential ideas of a proof (informally, and leaving out some low-level calculational details) by "decorating" a program with appropriate assertions on each of its commands.

Such a *decorated program* carries within it an argument for its own correctness.

For example, consider the program:

X ::= m;; Z ::= p; WHILE ~(X = 0) DO Z ::= Z - 1;; X ::= X - 1 END

(Note the *parameters* m and *p*, which stand for fixed-but-arbitrary numbers. Formally, they are simply Coq variables of type **nat**.)

Here is one possible specification for this program:

[1] X ::= m;; Z ::= p; WHILE ~(X = 0) DO Z ::= Z - 1;; X ::= X - 1 END [2]

Here is a decorated version of the program, embodying a proof of this specification:

[1] `True`

[2] `Z=p-m`

[3] -» [4] X ::= m;; [5] -» [6] Z ::= p; [7] -» [8] WHILE ~(X = 0) DO [9] -» [10] Z ::= Z - 1;; [11] X ::= X - 1 [12] END [13] -» [14]

Concretely, a decorated program consists of the program text interleaved with assertions (either a single assertion or possibly two assertions separated by an implication).

To check that a decorated program represents a valid proof, we check that each individual command is *locally consistent* with its nearby assertions in the following sense:

- *SKIP* is locally consistent if its precondition and postcondition are the same:

  [15] SKIP [16]

- The sequential composition of *c1* and *c2* is locally consistent (with respect to assertions P and R) if *c1* is locally consistent (with respect to P and Q) and *c2* is locally consistent (with respect to Q and R):

  [17] c1;; [18] c2 [19]

- An assertion is locally consistent if its precondition is the appropriate substitution of its postcondition:

  [20] X ::= a [21]

- A conditional is locally consistent (with respect to assertions P and Q) if the assertions at the top of its "then" and "else" branches are exactly $P \wedge b$ and $P \wedge \neg b$ and if its "then" branch is locally consistent (with respect to $P \wedge b$ and Q) and its "else" branch is locally consistent (with respect to $P \wedge \neg b$ and Q):

---

[3] `True`
[4] `m=m`
[5] `X=m`
[6] `X=m/\p=p`
[7] `X=m/\Z=p`
[8] `Z-X=p-m`
[9] `Z-X=p-m/\X<>0`
[10] `(Z-1)-(X-1)=p-m`
[11] `Z-(X-1)=p-m`
[12] `Z-X=p-m`
[13] `Z-X=p-m/\~(X<>0)`
[14] `Z=p-m`
[15] `P`
[16] `P`
[17] `P`
[18] `Q`
[19] `R`
[20] `P[X|->a]`
[21] `P`

[22] TEST b THEN [23] c1 [24] ELSE [25] c2 [26] FI [27]

- A while loop with precondition P is locally consistent if its postcondition is $P \wedge \neg b$, if the pre- and postconditions of its body are exactly $P \wedge b$ and P, and if its body is locally consistent:

  [28] WHILE b DO [29] c1 [30] END [31]

- A pair of assertions separated by -» is locally consistent if the first implies the second:

  [32] -» [33]

  This corresponds to the application of hoare_consequence, and it is the *only* place in a decorated program where checking whether decorations are correct is not fully mechanical and syntactic, but rather may involve logical and/or arithmetic reasoning.

These local consistency conditions essentially describe a procedure for *verifying* the correctness of a given proof. This verification involves checking that every single command is locally consistent with the accompanying assertions.

If we are instead interested in *finding* a proof for a given specification, we need to discover the right assertions. This can be done in an almost mechanical way, with the exception of finding loop invariants, which is the subject of the next section. In the remainder of this section we explain in detail how to construct decorations for several simple programs that don't involve non-trivial loop invariants.

6.1.1 Example: Swapping Using Addition and Subtraction

Here is a program that swaps the values of two variables using addition and subtraction (instead of by assigning to a temporary variable).

X ::= X + Y;; Y ::= X - Y;; X ::= X - Y

We can prove (informally) using decorations that this program is correct – i.e., it always swaps the values of variables X and Y.

---

[22] `P`
[23] `P/\b`
[24] `Q`
[25] `P/\~b`
[26] `Q`
[27] `Q`
[28] `P`
[29] `P/\b`
[30] `P`
[31] `P/\~b`
[32] `P`
[33] `P'`

(1) [34] -» (2) [35] X ::= X + Y;; (3) [36] Y ::= X - Y;; (4) [37] X ::= X - Y (5) [38]
These decorations can be constructed as follows:

- We begin with the undecorated program (the unnumbered lines).
- We add the specification – i.e., the outer precondition (1) and postcondition (5). In the precondition, we use parameters m and n to remember the initial values of variables X and Y so that we can refer to them in the postcondition (5).
- We work backwards, mechanically, starting from (5) and proceeding until we get to (2). At each step, we obtain the precondition of the assignment from its postcondition by substituting the assigned variable with the right-hand-side of the assignment. For instance, we obtain (4) by substituting X with X - Y in (5), and we obtain (3) by substituting Y with X - Y in (4).
- Finally, we verify that (1) logically implies (2) – i.e., that the step from (1) to (2) is a valid use of the law of consequence. For this we substitute X by m and Y by n and calculate as follows:

  (m + n) - ((m + n) - n) = n /\ (m + n) - n = m (m + n) - m = n /\ m = m n = n /\ m = m

Note that, since we are working with natural numbers rather than fixed-width machine integers, we don't need to worry about the possibility of arithmetic overflow anywhere in this argument. This makes life quite a bit simpler!

6.1.2 Example: Simple Conditionals

Here is a simple decorated program using conditionals:
(1) [39] TEST X <= Y THEN (2) [40] -» (3) [41] Z ::= Y - X (4) [42] ELSE (5) [43] -» (6) [44] Z ::= X - Y (7) [45] FI (8) [46]
These decorations were constructed as follows:

- We start with the outer precondition (1) and postcondition (8).

[34] `X=m/\Y=n`
[35] `(X+Y)-((X+Y)-Y)=n/\(X+Y)-Y=m`
[36] `X-(X-Y)=n/\X-Y=m`
[37] `X-Y=n/\Y=m`
[38] `X=n/\Y=m`
[39] `True`
[40] `True/\X<=Y`
[41] `(Y-X)+X=Y\/(Y-X)+Y=X`
[42] `Z+X=Y\/Z+Y=X`
[43] `True/\~(X<=Y)`
[44] `(X-Y)+X=Y\/(X-Y)+Y=X`
[45] `Z+X=Y\/Z+Y=X`
[46] `Z+X=Y\/Z+Y=X`

- We follow the format dictated by the hoare_if rule and copy the postcondition (8) to (4) and (7). We conjoin the precondition (1) with the guard of the conditional to obtain (2). We conjoin (1) with the negated guard of the conditional to obtain (5).

- In order to use the assignment rule and obtain (3), we substitute Z by Y - X in (4). To obtain (6) we substitute Z by X - Y in (7).

- Finally, we verify that (2) implies (3) and (5) implies (6). Both of these implications crucially depend on the ordering of X and Y obtained from the guard. For instance, knowing that X $\leq$ Y ensures that subtracting X from Y and then adding back X produces Y, as required by the first disjunct of (3). Similarly, knowing that $\neg$ (X $\leq$ Y) ensures that subtracting Y from X and then adding back Y produces X, as needed by the second disjunct of (6). Note that n - m + m = n does *not* hold for arbitrary natural numbers n and m (for example, 3 - 5 + 5 = 5).

Exercise: 2 stars, standard (if_minus_plus_reloaded) Fill in valid decorations for the following program:
[47] TEST X <= Y THEN [48] -» [49] Z ::= Y - X [50] ELSE [51] -» [52] Y ::= X + Z [53] FI [54]

Definition manual_grade_for_decorations_in_if_minus_plus_reloaded : option (nat×string) := None.
□

6.1.3 Example: Reduce to Zero

Here is a *WHILE* loop that is so simple it needs no invariant (i.e., the invariant True will do the job).
(1) [55] WHILE ~(X = 0) DO (2) [56] -» (3) [57] X ::= X - 1 (4) [58] END (5) [59] -» (6) [60]
The decorations can be constructed as follows:

- Start with the outer precondition (1) and postcondition (6).

---

47 True
48
49
50
51
52
53
54 Y=X+Z
55 True
56 True/\X<>0
57 True
58 True
59 True/\X=0
60 X=0

- Following the format dictated by the hoare_while rule, we copy (1) to (4). We conjoin (1) with the guard to obtain (2) and with the negation of the guard to obtain (5). Note that, because the outer postcondition (6) does not syntactically match (5), we need a trivial use of the consequence rule from (5) to (6).

- Assertion (3) is the same as (4), because X does not appear in 4, so the substitution in the assignment rule is trivial.

- Finally, the implication between (2) and (3) is also trivial.

From an informal proof in the form of a decorated program, it is easy to read off a formal proof using the Coq versions of the Hoare rules. Note that we do *not* unfold the definition of hoare_triple anywhere in this proof – the idea is to use the Hoare rules as a self-contained logic for reasoning about programs.

```
Definition reduce_to_zero' : com :=
  (WHILE ~(X = 0) DO
    X ::= X - 1
  END)%imp.

Theorem reduce_to_zero_correct' :
  {{fun st ⇒ True}}
  reduce_to_zero'
  {{fun st ⇒ st X = 0}}.
Proof.
  unfold reduce_to_zero'.
  eapply hoare_consequence_post.
  apply hoare_while.
  -

    eapply hoare_consequence_pre. apply hoare_asgn.
    intros st [HT Hbp]. unfold assn_sub. constructor.
  -
    intros st [Inv GuardFalse].
    unfold bassn in GuardFalse. simpl in GuardFalse.
    rewrite not_true_iff_false in GuardFalse.
    rewrite negb_false_iff in GuardFalse.
    apply eqb_eq in GuardFalse.
    apply GuardFalse. Qed.
```

6.1.4 Example: Division

The following Imp program calculates the integer quotient and remainder of two numbers m and n that are arbitrary constants in the program.

X ::= m;; Y ::= 0;; WHILE n <= X DO X ::= X - n;; Y ::= Y + 1 END;

In we replace m and n by concrete numbers and execute the program, it will terminate with the variable X set to the remainder when m is divided by n and Y set to the quotient.

In order to give a specification to this program we need to remember that dividing m by n produces a remainder X and a quotient Y such that $n \times Y + X = m \wedge X < n$.

It turns out that we get lucky with this program and don't have to think very hard about the loop invariant: the invariant is just the first conjunct $n \times Y + X = m$, and we can use this to decorate the program.

(1) [61] -» (2) [62] X ::= m;; (3) [63] Y ::= 0;; (4) [64] WHILE n <= X DO (5) [65] -» (6) [66] X ::= X - n;; (7) [67] Y ::= Y + 1 (8) [68] END (9) [69]

Assertions (4), (5), (8), and (9) are derived mechanically from the invariant and the loop's guard. Assertions (8), (7), and (6) are derived using the assignment rule going backwards from (8) to (6). Assertions (4), (3), and (2) are again backwards applications of the assignment rule.

Now that we've decorated the program it only remains to check that the two uses of the consequence rule are correct – i.e., that (1) implies (2) and that (5) implies (6). This is indeed the case, so we have a valid decorated program.

6.2 Finding Loop Invariants

Once the outermost precondition and postcondition are chosen, the only creative part in verifying programs using Hoare Logic is finding the right loop invariants. The reason this is difficult is the same as the reason that inductive mathematical proofs are: strengthening the loop invariant (or the induction hypothesis) means that you have a stronger assumption to work with when trying to establish the postcondition of the loop body (or complete the induction step of the proof), but it also means that the loop body's postcondition (or the statement being proved inductively) is stronger and thus harder to prove!

This section explains how to approach the challenge of finding loop invariants through a series of examples and exercises.

6.2.1 Example: Slow Subtraction

The following program subtracts the value of X from the value of Y by repeatedly decrementing both X and Y. We want to verify its correctness with respect to the pre- and postconditions shown:

[61] `True`
[62] `n*0+m=m`
[63] `n*0+X=m`
[64] `n*Y+X=m`
[65] `n*Y+X=m/\n<=X`
[66] `n*(Y+1)+(X-n)=m`
[67] `n*(Y+1)+X=m`
[68] `n*Y+X=m`
[69] `n*Y+X=m/\X<n`

[70] WHILE ~(X = 0) DO Y ::= Y - 1;; X ::= X - 1 END [71]

To verify this program, we need to find an invariant *Inv* for the loop. As a first step we can leave *Inv* as an unknown and build a *skeleton* for the proof by applying the rules for local consistency (working from the end of the program to the beginning, as usual, and without any thinking at all yet).

This leads to the following skeleton:

(1) [72] -» (a) (2) [73] WHILE ~(X = 0) DO (3) [74] -» (c) (4) [75] Y ::= Y - 1;; (5) [76] X ::= X - 1 (6) [77] END (7) [78] -» (b) (8) [79]

By examining this skeleton, we can see that any valid *Inv* will have to respect three conditions:

- (a) it must be *weak* enough to be implied by the loop's precondition, i.e., (1) must imply (2);
- (b) it must be *strong* enough to imply the program's postcondition, i.e., (7) must imply (8);
- (c) it must be *preserved* by each iteration of the loop (given that the loop guard evaluates to true), i.e., (3) must imply (4).

These conditions are actually independent of the particular program and specification we are considering. Indeed, every loop invariant has to satisfy them. One way to find an invariant that simultaneously satisfies these three conditions is by using an iterative process: start with a "candidate" invariant (e.g., a guess or a heuristic choice) and check the three conditions above; if any of the checks fails, try to use the information that we get from the failure to produce another – hopefully better – candidate invariant, and repeat.

For instance, in the reduce-to-zero example above, we saw that, for a very simple loop, choosing **True** as an invariant did the job. So let's try instantiating *Inv* with **True** in the skeleton above and see what we get...

(1) [80] -» (a - OK) (2) [81] WHILE ~(X = 0) DO (3) [82] -» (c - OK) (4) [83] Y ::= Y - 1;; (5)

---

[70] `X=m/\Y=n`
[71] `Y=n-m`
[72] `X=m/\Y=n`
[73] `Inv`
[74] `Inv/\X<>0`
[75] `Inv[X|->X-1][Y|->Y-1]`
[76] `Inv[X|->X-1]`
[77] `Inv`
[78] `Inv/\~(X<>0)`
[79] `Y=n-m`
[80] `X=m/\Y=n`
[81] `True`
[82] `True/\X<>0`
[83] `True`

[84] X ::= X - 1 (6) [85] END (7) [86] -» (b - WRONG!) (8) [87]

While conditions (a) and (c) are trivially satisfied, condition (b) is wrong, i.e., it is not the case that **True** ∧ X = 0 (7) implies Y = n - m (8). In fact, the two assertions are completely unrelated, so it is very easy to find a counterexample to the implication (say, Y = X = m = 0 and n = 1).

If we want (b) to hold, we need to strengthen the invariant so that it implies the postcondition (8). One simple way to do this is to let the invariant *be* the postcondition. So let's return to our skeleton, instantiate *Inv* with Y = n - m, and check conditions (a) to (c) again.

(1) [88] -» (a - WRONG!) (2) [89] WHILE ~(X = 0) DO (3) [90] -» (c - WRONG!) (4) [91] Y ::= Y - 1;; (5) [92] X ::= X - 1 (6) [93] END (7) [94] -» (b - OK) (8) [95]

This time, condition (b) holds trivially, but (a) and (c) are broken. Condition (a) requires that (1) X = m ∧ Y = n implies (2) Y = n - m. If we substitute Y by n we have to show that n = n - m for arbitrary m and n, which is not the case (for instance, when m = n = 1). Condition (c) requires that n - m - 1 = n - m, which fails, for instance, for n = 1 and m = 0. So, although Y = n - m holds at the end of the loop, it does not hold from the start, and it doesn't hold on each iteration; it is not a correct invariant.

This failure is not very surprising: the variable Y changes during the loop, while m and n are constant, so the assertion we chose didn't have much chance of being an invariant!

To do better, we need to generalize (8) to some statement that is equivalent to (8) when X is 0, since this will be the case when the loop terminates, and that "fills the gap" in some appropriate way when X is nonzero. Looking at how the loop works, we can observe that X and Y are decremented together until X reaches 0. So, if X = 2 and Y = 5 initially, after one iteration of the loop we obtain X = 1 and Y = 4; after two iterations X = 0 and Y = 3; and then the loop stops. Notice that the difference between Y and X stays constant between iterations: initially, Y = n and X = m, and the difference is always n - m. So let's try instantiating *Inv* in the skeleton above with Y - X = n - m.

(1) [96] -» (a - OK) (2) [97] WHILE ~(X = 0) DO (3) [98] -» (c - OK) (4) [99] Y ::= Y - 1;; (5)

---

[84] `True`
[85] `True`
[86] `True/\X=0`
[87] `Y=n-m`
[88] `X=m/\Y=n`
[89] `Y=n-m`
[90] `Y=n-m/\X<>0`
[91] `Y-1=n-m`
[92] `Y=n-m`
[93] `Y=n-m`
[94] `Y=n-m/\X=0`
[95] `Y=n-m`
[96] `X=m/\Y=n`
[97] `Y-X=n-m`
[98] `Y-X=n-m/\X<>0`
[99] `(Y-1)-(X-1)=n-m`

[100] X ::= X - 1 (6) [101] END (7) [102] -» (b - OK) (8) [103]

Success! Conditions (a), (b) and (c) all hold now. (To verify (c), we need to check that, under the assumption that $X \neq 0$, we have Y - X = (Y - 1) - (X - 1); this holds for all natural numbers X and Y.)

6.2.2 Exercise: Slow Assignment

Exercise: 2 stars, standard (slow_assignment) A roundabout way of assigning a number currently stored in X to the variable Y is to start Y at 0, then decrement X until it hits 0, incrementing Y at each step. Here is a program that implements this idea:

[104] Y ::= 0;; WHILE ~(X = 0) DO X ::= X - 1;; Y ::= Y + 1 END [105]

Write an informal decorated program showing that this procedure is correct.

```
Definition manual_grade_for_decorations_in_slow_assignment : option (nat×string) := None.
```

□

6.2.3 Exercise: Slow Addition

Exercise: 3 stars, standard, optional (add_slowly_decoration) The following program adds the variable X into the variable Z by repeatedly decrementing X and incrementing Z.

WHILE ~(X = 0) DO Z ::= Z + 1;; X ::= X - 1 END

Following the pattern of the subtract_slowly example above, pick a precondition and postcondition that give an appropriate specification of *add_slowly*; then (informally) decorate the program accordingly.

6.2.4 Example: Parity

Here is a cute little program for computing the parity of the value initially stored in X (due to Daniel Cristofani).

[106] WHILE 2 <= X DO X ::= X - 2 END [107]

The mathematical parity function used in the specification is defined in Coq as follows:

```
Fixpoint parity x :=
  match x with
  | 0 ⇒ 0
```

[100] Y-(X-1)=n-m
[101] Y-X=n-m
[102] Y-X=n-m/\X=0
[103] Y=n-m
[104] X=m
[105] Y=m
[106] X=m
[107] X=paritym

| 1 ⇒ 1
| S (S x') ⇒ parity x'
end.

The postcondition does not hold at the beginning of the loop, since m = parity m does not hold for an arbitrary m, so we cannot use that as an invariant. To find an invariant that works, let's think a bit about what this loop does. On each iteration it decrements X by 2, which preserves the parity of X. So the parity of X does not change, i.e., it is invariant. The initial value of X is m, so the parity of X is always equal to the parity of m. Using parity X = parity m as an invariant we obtain the following decorated program:

[108] -» (a - OK) [109] WHILE 2 <= X DO [110] -» (c - OK) [111] X ::= X - 2 [112] END [113] -» (b - OK) [114]

With this invariant, conditions (a), (b), and (c) are all satisfied. For verifying (b), we observe that, when X < 2, we have parity X = X (we can easily see this in the definition of parity). For verifying (c), we observe that, when 2 ≤ X, we have parity X = parity (X-2).

Exercise: 3 stars, standard, optional (parity_formal) Translate this proof to Coq. Refer to the *reduce_to_zero* example for ideas. You may find the following two lemmas useful:

```
Lemma parity_ge_2 : ∀ x,
  2 ≤ x →
  parity (x - 2) = parity x.
Proof.
  induction x; intro. reflexivity.
  destruct x. inversion H. inversion H1.
  simpl. rewrite ← minus_n_O. reflexivity.
Qed.

Lemma parity_lt_2 : ∀ x,
  ¬ 2 ≤ x →
  parity (x) = x.
Proof.
  intros. induction x. reflexivity. destruct x. reflexivity.
    exfalso. apply H. omega.
Qed.

Theorem parity_correct : ∀ m,
    {{ fun st ⇒ st X = m }}
```

[108] X=m
[109] parityX=paritym
[110] parityX=paritym/\2<=X
[111] parity(X-2)=paritym
[112] parityX=paritym
[113] parityX=paritym/\X<2
[114] X=paritym

```
  WHILE 2 ≤ X DO
    X ::= X - 2
  END
    {{ fun st ⇒ st X = parity m }}.
Proof.
```

Admitted.

□

6.2.5 Example: Finding Square Roots

The following program computes the (integer) square root of X by naive iteration:

[115] Z ::= 0;; WHILE (Z+1)*(Z+1) <= X DO Z ::= Z+1 END [116]

As above, we can try to use the postcondition as a candidate invariant, obtaining the following decorated program:

(1) [117] -» (a - second conjunct of (2) WRONG!) (2) [118] Z ::= 0;; (3) [119] WHILE (Z+1)*(Z+1) <= X DO (4) [120] -» (c - WRONG!) (5) [121] Z ::= Z+1 (6) [122] END (7) [123] -» (b - OK) (8) [124]

This didn't work very well: conditions (a) and (c) both failed. Looking at condition (c), we see that the second conjunct of (4) is almost the same as the first conjunct of (5), except that (4) mentions X while (5) mentions m. But note that X is never assigned in this program, so we should always have X=m; we didn't propagate this information from (1) into the loop invariant, but we could!

Also, we don't need the second conjunct of (8), since we can obtain it from the negation of the guard – the third conjunct in (7) – again under the assumption that X=m. This allows us to simplify a bit.

So we now try $X{=}m \land Z{\times}Z \leq m$ as the loop invariant:

[125] -» (a - OK) [126] Z ::= 0; [127] WHILE (Z+1)*(Z+1) <= X DO [128] -» (c - OK) [129] Z ::=

[115] `X=m`
[116] `Z*Z<=m/\m<(Z+1)*(Z+1)`
[117] `X=m`
[118] `0*0<=m/\m<1*1`
[119] `Z*Z<=m/\m<(Z+1)*(Z+1)`
[120] `Z*Z<=m/\(Z+1)*(Z+1)<=X`
[121] `(Z+1)*(Z+1)<=m/\m<(Z+2)*(Z+2)`
[122] `Z*Z<=m/\m<(Z+1)*(Z+1)`
[123] `Z*Z<=m/\m<(Z+1)*(Z+1)/\~((Z+1)*(Z+1)<=X)`
[124] `Z*Z<=m/\m<(Z+1)*(Z+1)`
[125] `X=m`
[126] `X=m/\0*0<=m`
[127] `X=m/\Z*Z<=m`
[128] `X=m/\Z*Z<=m/\(Z+1)*(Z+1)<=X`
[129] `X=m/\(Z+1)*(Z+1)<=m`

Z + 1 [130] END [131] -» (b - OK) [132]

This works, since conditions (a), (b), and (c) are now all trivially satisfied.

Very often, if a variable is used in a loop in a read-only fashion (i.e., it is referred to by the program or by the specification and it is not changed by the loop), it is necessary to add the fact that it doesn't change to the loop invariant.

6.2.6 Example: Squaring

Here is a program that squares X by repeated addition:

[133] Y ::= 0;; Z ::= 0;; WHILE ~(Y = X) DO Z ::= Z + X;; Y ::= Y + 1 END [134]

The first thing to note is that the loop reads X but doesn't change its value. As we saw in the previous example, it is a good idea in such cases to add X = m to the invariant. The other thing that we know is often useful in the invariant is the postcondition, so let's add that too, leading to the invariant candidate Z = m × m ∧ X = m.

[135] -» (a - WRONG) [136] Y ::= 0;; [137] Z ::= 0;; [138] WHILE ~(Y = X) DO [139] -» (c - WRONG) [140] Z ::= Z + X;; [141] Y ::= Y + 1 [142] END [143] -» (b - OK) [144]

Conditions (a) and (c) fail because of the Z = m×m part. While Z starts at 0 and works itself up to m×m, we can't expect Z to be m×m from the start. If we look at how Z progresses in the loop, after the 1st iteration Z = m, after the 2nd iteration Z = 2*m, and at the end Z = m×m. Since the variable Y tracks how many times we go through the loop, this leads us to derive a new invariant candidate: Z = Y×m ∧ X = m.

[145] -» (a - OK) [146] Y ::= 0;; [147] Z ::= 0;; [148] WHILE ~(Y = X) DO [149] -» (c - OK) [150] Z

---

[130] X=m/\Z*Z<=m

[131] X=m/\Z*Z<=m/\X<(Z+1)*(Z+1)

[132] Z*Z<=m/\m<(Z+1)*(Z+1)

[133] X=m

[134] Z=m*m

[135] X=m

[136] 0=m*m/\X=m

[137] 0=m*m/\X=m

[138] Z=m*m/\X=m

[139] Z=Y*m/\X=m/\Y<>X

[140] Z+X=m*m/\X=m

[141] Z=m*m/\X=m

[142] Z=m*m/\X=m

[143] Z=m*m/\X=m/\~(Y<>X)

[144] Z=m*m

[145] X=m

[146] 0=0*m/\X=m

[147] 0=Y*m/\X=m

[148] Z=Y*m/\X=m

[149] Z=Y*m/\X=m/\Y<>X

[150] Z+X=(Y+1)*m/\X=m

::= Z + X; [151] Y ::= Y + 1 [152] END [153] -» (b - OK) [154]

This new invariant makes the proof go through: all three conditions are easy to check.

It is worth comparing the postcondition Z = m×m and the Z = Y×m conjunct of the invariant. It is often the case that one has to replace parameters with variables – or with expressions involving both variables and parameters, like m - Y – when going from postconditions to invariants.

6.2.7 Exercise: Factorial

Exercise: 3 stars, standard (factorial) Recall that n! denotes the factorial of n (i.e., n! = 1*2*...*n). Here is an Imp program that calculates the factorial of the number initially stored in the variable X and puts it in the variable Y:

[155] Y ::= 1 ;; WHILE ~(X = 0) DO Y ::= Y * X ;; X ::= X - 1 END [156]

Fill in the blanks in following decorated program. For full credit, make sure all the arithmetic operations used in the assertions are well-defined on natural numbers.

[157] -» [158] Y ::= 1;; [159] WHILE ~(X = 0) DO [160] -» [161] Y ::= Y * X;; [162] X ::= X - 1 [163] END [164] -» [165]

```
Definition manual_grade_for_decorations_in_factorial : option (nat×string) := None.
```

□

6.2.8 Exercise: Min

Exercise: 3 stars, standard (Min_Hoare) Fill in valid decorations for the following program. For the ⇒ steps in your annotations, you may rely (silently) on the following facts about min

Lemma lemma1 : forall x y, (x=0 \/ y=0) -> min x y = 0. Lemma lemma2 : forall x y, min (x-1) (y-1) = (min x y) - 1.

plus standard high-school algebra, as always.

---

[151] Z=(Y+1)*m/\X=m

[152] Z=Y*m/\X=m

[153] Z=Y*m/\X=m/\~(Y<>X)

[154] Z=m*m

[155] X=m

[156] Y=m!

[157] X=m

[158]

[159]

[160]

[161]

[162]

[163]

[164]

[165] Y=m!

[166] -» [167] X ::= a;; [168] Y ::= b;; [169] Z ::= 0;; [170] WHILE ~(X = 0) && ~(Y = 0) DO [171] -» [172] X := X - 1;; [173] Y := Y - 1;; [174] Z := Z + 1 [175] END [176] -» [177]

`Definition` manual_grade_for_decorations_in_Min_Hoare : **option** (**nat**×**string**) := None.

□

Exercise: 3 stars, standard (two_loops) Here is a very inefficient way of adding 3 numbers:

X ::= 0;; Y ::= 0;; Z ::= c;; WHILE ~(X = a) DO X ::= X + 1;; Z ::= Z + 1 END;; WHILE ~(Y = b) DO Y ::= Y + 1;; Z ::= Z + 1 END

Show that it does what it should by filling in the blanks in the following decorated program.

[178] -» [179] X ::= 0;; [180] Y ::= 0;; [181] Z ::= c;; [182] WHILE ~(X = a) DO [183] -» [184] X ::= X + 1;; [185] Z ::= Z + 1 [186] END;; [187] -» [188] WHILE ~(Y = b) DO [189] -» [190] Y ::= Y + 1;; [191] Z ::= Z + 1 [192] END [193] -» [194]

`Definition` manual_grade_for_decorations_in_two_loops : **option** (**nat**×**string**) := None.

□

---

[166] `True`
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177] `Z=minab`
[178] `True`
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194] `Z=a+b+c`

6.2.9 Exercise: Power Series

Exercise: 4 stars, standard, optional (dpow2_down) Here is a program that computes the series: 1 + 2 + 2^2 + ... + 2^m = 2^(m+1) - 1

X ::= 0;; Y ::= 1;; Z ::= 1;; WHILE ~(X = m) DO Z ::= 2 * Z;; Y ::= Y + Z;; X ::= X + 1 END

Write a decorated program for this.

6.3 Weakest Preconditions (Optional)

Some Hoare triples are more interesting than others. For example,

[195] X ::= Y + 1 [196]

is *not* very interesting: although it is perfectly valid, it tells us nothing useful. Since the precondition isn't satisfied by any state, it doesn't describe any situations where we can use the command X ::= Y + 1 to achieve the postcondition X $\leq$ 5.

By contrast,

[197] X ::= Y + 1 [198]

is useful: it tells us that, if we can somehow create a situation in which we know that Y $\leq$ 4 $\wedge$ Z = 0, then running this command will produce a state satisfying the postcondition. However, this triple is still not as useful as it could be, because the Z = 0 clause in the precondition actually has nothing to do with the postcondition X $\leq$ 5. The *most* useful triple (for this command and postcondition) is this one:

[199] X ::= Y + 1 [200]

In other words, Y $\leq$ 4 is the *weakest* valid precondition of the command X ::= Y + 1 for the postcondition X $\leq$ 5.

In general, we say that "P is the weakest precondition of command c for postcondition Q" if {{P}} c {{Q}} and if, whenever P' is an assertion such that {{P'}} c {{Q}}, it is the case that P' *st* implies P *st* for all states *st*.

```
Definition is_wp P c Q :=
  {{P}} c {{Q}} ∧
  ∀ P', {{P'}} c {{Q}} → (P' ->> P).
```

That is, P is the weakest precondition of c for Q if (a) P *is* a precondition for Q and c, and (b) P is the *weakest* (easiest to satisfy) assertion that guarantees that Q will hold after executing c.

[195] False
[196] X<=5
[197] Y<=4/\Z=0
[198] X<=5
[199] Y<=4
[200] X<=5

Exercise: 1 star, standard, optional (wp) What are the weakest preconditions of the following commands for the following postconditions?

1) [201] SKIP [202]
2) [203] X ::= Y + Z [204]
3) [205] X ::= Y [206]
4) [207] TEST X = 0 THEN Y ::= Z + 1 ELSE Y ::= W + 2 FI [208]
5) [209] X ::= 5 [210]
6) [211] WHILE true DO X ::= 0 END [212]

Exercise: 3 stars, advanced, optional (is_wp_formal) Prove formally, using the definition of hoare_triple, that $Y \leq 4$ is indeed the weakest precondition of X ::= Y + 1 with respect to postcondition $X \leq 5$.

```
Theorem is_wp_example :
  is_wp (fun st ⇒ st Y ≤ 4)
    (X ::= Y + 1) (fun st ⇒ st X ≤ 5).
Proof.
   Admitted.
```

□

Exercise: 2 stars, advanced, optional (hoare_asgn_weakest) Show that the precondition in the rule hoare_asgn is in fact the weakest precondition.

```
Theorem hoare_asgn_weakest : ∀ Q X a,
  is_wp (Q [X |-> a]) (X ::= a) Q.
Proof.
   Admitted.
```

□

Exercise: 2 stars, advanced, optional (hoare_havoc_weakest) Show that your havoc_pre rule from the *himp_hoare* exercise in the Hoare chapter returns the weakest precondition. `Module` HIMP2.

`Import` *Himp*.

---

[201] ?

[202] X=5

[203] ?

[204] X=5

[205] ?

[206] X=Y

[207] ?

[208] Y=5

[209] ?

[210] X=0

[211] ?

[212] X=0

```
Lemma hoare_havoc_weakest : ∀ (P Q : Assertion) (X : string),
  {{ P }} HAVOC X {{ Q }} →
  P ->> havoc_pre X Q.
Proof.
   Admitted.
End HIMP2.
```
□

6.4 Formal Decorated Programs (Advanced)

Our informal conventions for decorated programs amount to a way of displaying Hoare triples, in which commands are annotated with enough embedded assertions that checking the validity of a triple is reduced to simple logical and algebraic calculations showing that some assertions imply others. In this section, we show that this informal presentation style can actually be made completely formal and indeed that checking the validity of decorated programs can mostly be automated.

6.4.1 Syntax

The first thing we need to do is to formalize a variant of the syntax of commands with embedded assertions. We call the new commands *decorated commands*, or **dcom**s.

We don't want both preconditions and postconditions on each command, because a sequence of two commands would contain redundant decorations–the postcondition of the first likely being the same as the precondition of the second. Instead, decorations are added corresponding to postconditions only. A separate type, **decorated**, is used to add just one precondition for the entire program.

```
Inductive dcom : Type :=
  | DCSkip : Assertion → dcom
  | DCSeq : dcom → dcom → dcom
  | DCAsgn : string → aexp → Assertion → dcom
  | DCIf : bexp → Assertion → dcom → Assertion → dcom
             → Assertion→ dcom
  | DCWhile : bexp → Assertion → dcom → Assertion → dcom
  | DCPre : Assertion → dcom → dcom
  | DCPost : dcom → Assertion → dcom.

Inductive decorated : Type :=
  | Decorated : Assertion → dcom → decorated.

Delimit Scope default with default.

Notation "'SKIP' {{ P }}"
      := (DCSkip P)
      (at level 10) : dcom_scope.
```

```
Notation "l '::=' a {{ P }}"
      := (DCAsgn l a P)
      (at level 60, a at next level) : dcom_scope.
Notation "'WHILE' b 'DO' {{ Pbody }} d 'END' {{ Ppost }}"
      := (DCWhile b Pbody d Ppost)
      (at level 80, right associativity) : dcom_scope.
Notation "'TEST' b 'THEN' {{ P }} d 'ELSE' {{ P' }} d' 'FI' {{ Q }}"
      := (DCIf b P d P' d' Q)
      (at level 80, right associativity) : dcom_scope.
Notation "'->>' {{ P }} d"
      := (DCPre P d)
      (at level 90, right associativity) : dcom_scope.
Notation "d '->>' {{ P }}"
      := (DCPost d P)
      (at level 80, right associativity) : dcom_scope.
Notation " d ;; d' "
      := (DCSeq d d')
      (at level 80, right associativity) : dcom_scope.
Notation "{{ P }} d"
      := (Decorated P d)
      (at level 90) : dcom_scope.

Delimit Scope dcom_scope with dcom.
Open Scope dcom_scope.

Example dec0 :=
  SKIP {{ fun st ⇒ True }}.
Example dec1 :=
  WHILE true DO {{ fun st ⇒ True }} SKIP {{ fun st ⇒ True }} END
  {{ fun st ⇒ True }}.
Set Printing All.
```

To avoid clashing with the existing `Notation` definitions for ordinary **com**mands, we introduce these notations in a special scope called *dcom_scope*, and we `Open` this scope for the remainder of the file.

Careful readers will note that we've defined two notations for specifying a precondition explicitly, one with and one without a ->>. The "without" version is intended to be used to supply the initial precondition at the very top of the program.

```
Example dec_while : decorated :=
  {{ fun st ⇒ True }}
  WHILE ~(X = 0)
  DO
    {{ fun st ⇒ True ∧ st X ≠ 0}}
    X ::= X - 1
```

```
    {{ fun _ ⇒ True }}
  END
  {{ fun st ⇒ True ∧ st X = 0}} ->>
  {{ fun st ⇒ st X = 0 }}.
```

It is easy to go from a **dcom** to a **com** by erasing all annotations.

```
Fixpoint extract (d : dcom) : com :=
  match d with
  | DCSkip _ ⇒ SKIP
  | DCSeq d1 d2 ⇒ (extract d1 ;; extract d2)
  | DCAsgn X a _ ⇒ X ::= a
  | DCIf b _ d1 _ d2 _ ⇒ TEST b THEN extract d1 ELSE extract d2 FI
  | DCWhile b _ d _ ⇒ WHILE b DO extract d END
  | DCPre _ d ⇒ extract d
  | DCPost d _ ⇒ extract d
  end.

Definition extract_dec (dec : decorated) : com :=
  match dec with
  | Decorated P d ⇒ extract d
  end.
```

The choice of exactly where to put assertions in the definition of **dcom** is a bit subtle. The simplest thing to do would be to annotate every **dcom** with a precondition and postcondition. But this would result in very verbose programs with a lot of repeated annotations: for example, a program like *SKIP*;*SKIP* would have to be annotated as

[213] ([214] SKIP [215]) ;; ([216] SKIP [217]) [218],

with pre- and post-conditions on each *SKIP*, plus identical pre- and post-conditions on the semicolon!

Instead, the rule we've followed is this:

- The *post*-condition expected by each **dcom** *d* is embedded in *d*.
- The *pre*-condition is supplied by the context.

In other words, the invariant of the representation is that a **dcom** *d* together with a precondition *P* determines a Hoare triple {{*P*}} (extract *d*) {{post *d*}}, where post is defined as follows:

```
Fixpoint post (d : dcom) : Assertion :=
```

[213] P
[214] P
[215] P
[216] P
[217] P
[218] P

```
  match d with
  | DCSkip P ⇒ P
  | DCSeq d1 d2 ⇒ post d2
  | DCAsgn X a Q ⇒ Q
  | DCIf _ _ d1 _ d2 Q ⇒ Q
  | DCWhile b Pbody c Ppost ⇒ Ppost
  | DCPre _ d ⇒ post d
  | DCPost c Q ⇒ Q
  end.
```

It is straightforward to extract the precondition and postcondition from a decorated program.

```
Definition pre_dec (dec : decorated) : Assertion :=
  match dec with
  | Decorated P d ⇒ P
  end.

Definition post_dec (dec : decorated) : Assertion :=
  match dec with
  | Decorated P d ⇒ post d
  end.
```

We can express what it means for a decorated program to be correct as follows:

```
Definition dec_correct (dec : decorated) :=
  {{pre_dec dec}} (extract_dec dec) {{post_dec dec}}.
```

To check whether this Hoare triple is *valid*, we need a way to extract the "proof obligations" from a decorated program. These obligations are often called *verification conditions*, because they are the facts that must be verified to see that the decorations are logically consistent and thus add up to a complete proof of correctness.

6.4.2 Extracting Verification Conditions

The function verification_conditions takes a **dcom** d together with a precondition P and returns a *proposition* that, if it can be proved, implies that the triple $\{\{P\}\}$ (extract d) $\{\{$post $d\}\}$ is valid.

It does this by walking over d and generating a big conjunction including all the "local checks" that we listed when we described the informal rules for decorated programs. (Strictly speaking, we need to massage the informal rules a little bit to add some uses of the rule of consequence, but the correspondence should be clear.)

```
Fixpoint verification_conditions (P : Assertion) (d : dcom) : Prop :=
  match d with
  | DCSkip Q ⇒
      (P ->> Q)
  | DCSeq d1 d2 ⇒
```

```
    verification_conditions P d1
    ∧ verification_conditions (post d1) d2
  | DCAsgn X a Q ⇒
    (P ->> Q [X |-> a])
  | DCIf b P1 d1 P2 d2 Q ⇒
    ((fun st ⇒ P st ∧ bassn b st) ->> P1)
    ∧ ((fun st ⇒ P st ∧ ¬ (bassn b st)) ->> P2)
    ∧ (post d1 ->> Q) ∧ (post d2 ->> Q)
    ∧ verification_conditions P1 d1
    ∧ verification_conditions P2 d2
  | DCWhile b Pbody d Ppost ⇒

    (P ->> post d)
    ∧ ((fun st ⇒ post d st ∧ bassn b st) ->> Pbody)
    ∧ ((fun st ⇒ post d st ∧ ~(bassn b st)) ->> Ppost)
    ∧ verification_conditions Pbody d
  | DCPre P' d ⇒
    (P ->> P') ∧ verification_conditions P' d
  | DCPost d Q ⇒
    verification_conditions P d ∧ (post d ->> Q)
  end.
```

And now the key theorem, stating that verification_conditions does its job correctly. Not surprisingly, we need to use each of the Hoare Logic rules at some point in the proof.

```
Theorem verification_correct : ∀ d P,
  verification_conditions P d → {{P}} (extract d) {{post d}}.
Proof.
  induction d; intros P H; simpl in *.
  -
    eapply hoare_consequence_pre.
      apply hoare_skip.
      assumption.
  -
    destruct H as [H1 H2].
    eapply hoare_seq.
      apply IHd2. apply H2.
      apply IHd1. apply H1.
  -
    eapply hoare_consequence_pre.
      apply hoare_asgn.
      assumption.
  -
    destruct H as [HPre1 [HPre2 [Hd1 [Hd2 [HThen HElse]]]]].
```

```
    apply IHd1 in HThen. clear IHd1.
    apply IHd2 in HElse. clear IHd2.
    apply hoare_if.
      + eapply hoare_consequence_post with (Q':=post d1); eauto.
         eapply hoare_consequence_pre; eauto.
      + eapply hoare_consequence_post with (Q':=post d2); eauto.
         eapply hoare_consequence_pre; eauto.
  -
    destruct H as [Hpre [Hbody1 [Hpost1 Hd]]].
    eapply hoare_consequence_pre; eauto.
    eapply hoare_consequence_post; eauto.
    apply hoare_while.
    eapply hoare_consequence_pre; eauto.
  -
    destruct H as [HP Hd].
    eapply hoare_consequence_pre. apply IHd. apply Hd. assumption.
  -
    destruct H as [Hd HQ].
    eapply hoare_consequence_post. apply IHd. apply Hd. assumption.
Qed.
```

6.4.3 Automation

Now that all the pieces are in place, we can verify an entire program.

```
Definition verification_conditions_dec (dec : decorated) : Prop :=
  match dec with
  | Decorated P d ⇒ verification_conditions P d
  end.

Lemma verification_correct_dec : ∀ dec,
  verification_conditions_dec dec → dec_correct dec.
Proof.
  intros [P d]. apply verification_correct.
Qed.
```

The propositions generated by verification_conditions are fairly big, and they contain many conjuncts that are essentially trivial.

```
Eval simpl in (verification_conditions_dec dec_while).
```

===> (((fun _: state => True) -» (fun _: state => True)) /\ ((fun st : state => True /\ bassn (~(X = 0)) st) -» (fun st : state => True /\ st X <> 0)) /\ ((fun st : state => True /\ ~ bassn (~(X = 0)) st) -» (fun st : state => True /\ st X = 0)) /\ (fun st : state => True /\ st X <> 0) -» (fun _: state => True) X |-> X - 1) /\ (fun st : state => True /\ st X = 0) -» (fun st : state => st X = 0)

In principle, we could work with such propositions using just the tactics we have so far, but we can make things much smoother with a bit of automation. We first define a custom *verify* tactic that uses `split` repeatedly to turn all the conjunctions into separate subgoals and then uses `omega` and `eauto` (described in chapter *Auto* in *Logical Foundations*) to deal with as many of them as possible.

```
Tactic Notation "verify" :=
  apply verification_correct;
  repeat split;
  simpl; unfold assert_implies;
  unfold bassn in *; unfold beval in *; unfold aeval in *;
  unfold assn_sub; intros;
  repeat rewrite t_update_eq;
  repeat (rewrite t_update_neq; [| (intro X; inversion X)]);
  simpl in *;
  repeat match goal with [H : _ ∧ _ ⊢ _] ⇒ destruct H end;
  repeat rewrite not_true_iff_false in *;
  repeat rewrite not_false_iff_true in *;
  repeat rewrite negb_true_iff in *;
  repeat rewrite negb_false_iff in *;
  repeat rewrite eqb_eq in *;
  repeat rewrite eqb_neq in *;
  repeat rewrite leb_iff in *;
  repeat rewrite leb_iff_conv in *;
  try subst;
  repeat
    match goal with
      [st : state ⊢ _] ⇒
        match goal with
          [H : st _ = _ ⊢ _] ⇒ rewrite → H in *; clear H
        | [H : _ = st _ ⊢ _] ⇒ rewrite ← H in *; clear H
        end
    end;
  try eauto; try omega.
```

What's left after *verify* does its thing is "just the interesting parts" of checking that the decorations are correct. For very simple examples, *verify* sometimes even immediately solves the goal (provided that the annotations are correct!).

```
Theorem dec_while_correct :
  dec_correct dec_while.
Proof. verify. Qed.
```

Another example (formalizing a decorated program we've seen before):

```
Example subtract_slowly_dec (m : nat) (p : nat) : decorated :=
```

```
    {{ fun st ⇒ st X = m ∧ st Z = p }} ->>
    {{ fun st ⇒ st Z - st X = p - m }}
  WHILE ~(X = 0)
  DO {{ fun st ⇒ st Z - st X = p - m ∧ st X ≠ 0 }} ->>
       {{ fun st ⇒ (st Z - 1) - (st X - 1) = p - m }}
     Z ::= Z - 1
       {{ fun st ⇒ st Z - (st X - 1) = p - m }} ;;
     X ::= X - 1
       {{ fun st ⇒ st Z - st X = p - m }}
  END
    {{ fun st ⇒ st Z - st X = p - m ∧ st X = 0 }} ->>
    {{ fun st ⇒ st Z = p - m }}.

Theorem subtract_slowly_dec_correct : ∀ m p,
  dec_correct (subtract_slowly_dec m p).
Proof. intros m p. verify. Qed.
```

6.4.4 Examples

In this section, we use the automation developed above to verify formal decorated programs corresponding to most of the informal ones we have seen.

Swapping Using Addition and Subtraction

```
Definition swap : com :=
  X ::= X + Y;;
  Y ::= X - Y;;
  X ::= X - Y.

Definition swap_dec m n : decorated :=
   {{ fun st ⇒ st X = m ∧ st Y = n}} ->>
   {{ fun st ⇒ (st X + st Y) - ((st X + st Y) - st Y) = n
                ∧ (st X + st Y) - st Y = m }}
  X ::= X + Y
   {{ fun st ⇒ st X - (st X - st Y) = n ∧ st X - st Y = m }};;
  Y ::= X - Y
   {{ fun st ⇒ st X - st Y = n ∧ st Y = m }};;
  X ::= X - Y
   {{ fun st ⇒ st X = n ∧ st Y = m}}.

Theorem swap_correct : ∀ m n,
  dec_correct (swap_dec m n).
Proof. intros; verify. Qed.
```

Simple Conditionals

```
Definition if_minus_plus_com :=
  (TEST X ≤ Y
    THEN Z ::= Y - X
    ELSE Y ::= X + Z
  FI)%imp.

Definition if_minus_plus_dec :=
  {{fun st ⇒ True}}
  TEST X ≤ Y THEN
      {{ fun st ⇒ True ∧ st X ≤ st Y }} ->>
      {{ fun st ⇒ st Y = st X + (st Y - st X) }}
    Z ::= Y - X
      {{ fun st ⇒ st Y = st X + st Z }}
  ELSE
      {{ fun st ⇒ True ∧ ~(st X ≤ st Y) }} ->>
      {{ fun st ⇒ st X + st Z = st X + st Z }}
    Y ::= X + Z
      {{ fun st ⇒ st Y = st X + st Z }}
  FI
  {{fun st ⇒ st Y = st X + st Z}}.

Theorem if_minus_plus_correct :
  dec_correct if_minus_plus_dec.
Proof. verify. Qed.

Definition if_minus_dec :=
  {{fun st ⇒ True}}
  TEST X ≤ Y THEN
      {{fun st ⇒ True ∧ st X ≤ st Y }} ->>
      {{fun st ⇒ (st Y - st X) + st X = st Y
                 ∨ (st Y - st X) + st Y = st X}}
    Z ::= Y - X
      {{fun st ⇒ st Z + st X = st Y ∨ st Z + st Y = st X}}
  ELSE
      {{fun st ⇒ True ∧ ~(st X ≤ st Y) }} ->>
      {{fun st ⇒ (st X - st Y) + st X = st Y
                 ∨ (st X - st Y) + st Y = st X}}
    Z ::= X - Y
      {{fun st ⇒ st Z + st X = st Y ∨ st Z + st Y = st X}}
  FI
    {{fun st ⇒ st Z + st X = st Y ∨ st Z + st Y = st X}}.

Theorem if_minus_correct :
  dec_correct if_minus_dec.
```

```
Proof. verify. Qed.
```

Division

```
Definition div_mod_dec (a b : nat) : decorated :=
  {{ fun st ⇒ True }} ->>
  {{ fun st ⇒ b × 0 + a = a }}
  X ::= a
  {{ fun st ⇒ b × 0 + st X = a }};;
  Y ::= 0
  {{ fun st ⇒ b × st Y + st X = a }};;
  WHILE b ≤ X DO
    {{ fun st ⇒ b × st Y + st X = a ∧ b ≤ st X }} ->>
    {{ fun st ⇒ b × (st Y + 1) + (st X - b) = a }}
    X ::= X - b
    {{ fun st ⇒ b × (st Y + 1) + st X = a }};;
    Y ::= Y + 1
    {{ fun st ⇒ b × st Y + st X = a }}
  END
  {{ fun st ⇒ b × st Y + st X = a ∧ ~(b ≤ st X) }} ->>
  {{ fun st ⇒ b × st Y + st X = a ∧ (st X < b) }}.

Theorem div_mod_dec_correct : ∀ a b,
  dec_correct (div_mod_dec a b).
Proof. intros a b. verify.
  rewrite mult_plus_distr_l. omega.
Qed.
```

Parity

```
Definition find_parity : com :=
  WHILE 2 ≤ X DO
     X ::= X - 2
  END.
```

There are actually several ways to phrase the loop invariant for this program. Here is one natural one, which leads to a rather long proof:

```
Inductive ev : nat → Prop :=
  | ev_0 : ev O
  | ev_SS : ∀ n : nat, ev n → ev (S (S n)).

Definition find_parity_dec m : decorated :=
   {{ fun st ⇒ st X = m}} ->>
   {{ fun st ⇒ st X ≤ m ∧ ev (m - st X) }}
```

```
  WHILE 2 ≤ X DO
     {{ fun st ⇒ (st X ≤ m ∧ ev (m - st X)) ∧ 2 ≤ st X }} ->>
     {{ fun st ⇒ st X - 2 ≤ m ∧ (ev (m - (st X - 2))) }}
     X ::= X - 2
     {{ fun st ⇒ st X ≤ m ∧ ev (m - st X) }}
  END
   {{ fun st ⇒ (st X ≤ m ∧ ev (m - st X)) ∧ st X < 2 }} ->>
   {{ fun st ⇒ st X=0 ↔ ev m }}.
Lemma l1 : ∀ m n p,
  p ≤ n →
  n ≤ m →
  m - (n - p) = m - n + p.
Proof. intros. omega. Qed.
Lemma l2 : ∀ m,
  ev m →
  ev (m + 2).
Proof. intros. rewrite plus_comm. simpl. constructor. assumption. Qed.
Lemma l3' : ∀ m,
  ev m →
  ¬ev (S m).
Proof. induction m; intros H1 H2. inversion H2. apply IHm.
       inversion H2; subst; assumption. assumption. Qed.
Lemma l3 : ∀ m,
  1 ≤ m →
  ev m →
  ev (m - 1) →
  False.
Proof. intros. apply l2 in H1.
       assert (G : m - 1 + 2 = S m). clear H0 H1. omega.
       rewrite G in H1. apply l3' in H0. apply H0. assumption. Qed.
Theorem find_parity_correct : ∀ m,
  dec_correct (find_parity_dec m).
Proof.
  intro m. verify;

    fold (2 <=? (st X)) in *;
    try rewrite leb_iff in *;
    try rewrite leb_iff_conv in *; eauto; try omega.
    -
      rewrite minus_diag. constructor.
    -
```

```
      rewrite l1; try assumption.
      apply l2; assumption.
    -
      rewrite ← minus_n_O in H2. assumption.
    -
      destruct (st X) as [| [| n]].
      +
        reflexivity.
      +
        apply l3 in H; try assumption. inversion H.
      +
        clear H0 H2.          omega.
Qed.
```

Here is a more intuitive way of writing the invariant:

```
Definition find_parity_dec' m : decorated :=
  {{ fun st ⇒ st X = m}} ->>
  {{ fun st ⇒ ev (st X) ↔ ev m }}
 WHILE 2 ≤ X DO
    {{ fun st ⇒ (ev (st X) ↔ ev m) ∧ 2 ≤ st X }} ->>
    {{ fun st ⇒ (ev (st X - 2) ↔ ev m) }}
    X ::= X - 2
    {{ fun st ⇒ (ev (st X) ↔ ev m) }}
 END
 {{ fun st ⇒ (ev (st X) ↔ ev m) ∧ ~(2 ≤ st X) }} ->>
 {{ fun st ⇒ st X=0 ↔ ev m }}.

Lemma l4 : ∀ m,
  2 ≤ m →
  (ev (m - 2) ↔ ev m).
Proof.
  induction m; intros. split; intro; constructor.
  destruct m. inversion H. inversion H1. simpl in *.
  rewrite ← minus_n_O in *. split; intro.
    constructor. assumption.
    inversion H0. assumption.
Qed.

Theorem find_parity_correct' : ∀ m,
  dec_correct (find_parity_dec' m).
Proof.
  intros m. verify;

    fold (2 <=? (st X)) in *;
```

```
    try rewrite leb_iff in *;
    try rewrite leb_iff_conv in *; intuition; eauto; try omega.
  -
    rewrite l4 in H0; eauto.
  -
    rewrite l4; eauto.
  -
    apply H0. constructor.
  -
      destruct (st X) as [| [| n]].          +
        reflexivity.
      +
        inversion H.
      +
        clear H0 H H3.            omega.
Qed.
```

Here is the simplest invariant we've found for this program:

```
Definition parity_dec m : decorated :=
  {{ fun st ⇒ st X = m}} ->>
  {{ fun st ⇒ parity (st X) = parity m }}
 WHILE 2 ≤ X DO
    {{ fun st ⇒ parity (st X) = parity m ∧ 2 ≤ st X }} ->>
    {{ fun st ⇒ parity (st X - 2) = parity m }}
    X ::= X - 2
    {{ fun st ⇒ parity (st X) = parity m }}
 END
 {{ fun st ⇒ parity (st X) = parity m ∧ ~(2 ≤ st X) }} ->>
 {{ fun st ⇒ st X = parity m }}.

Theorem parity_dec_correct : ∀ m,
  dec_correct (parity_dec m).
Proof.
  intros. verify;

    fold (2 <=? (st X)) in *;
    try rewrite leb_iff in *;
    try rewrite leb_iff_conv in *; eauto; try omega.
  -
    rewrite ← H. apply parity_ge_2. assumption.
  -
    rewrite ← H. symmetry. apply parity_lt_2. assumption.
Qed.
```

Square Roots

```
Definition sqrt_dec m : decorated :=
    {{ fun st ⇒ st X = m }} ->>
    {{ fun st ⇒ st X = m ∧ 0×0 ≤ m }}
  Z ::= 0
    {{ fun st ⇒ st X = m ∧ st Z×st Z ≤ m }};;
  WHILE (Z+1)*(Z+1) ≤ X DO
      {{ fun st ⇒ (st X = m ∧ st Z×st Z≤m)
                   ∧ (st Z + 1)*(st Z + 1) ≤ st X }} ->>
      {{ fun st ⇒ st X = m ∧ (st Z+1)*(st Z+1)≤m }}
    Z ::= Z + 1
      {{ fun st ⇒ st X = m ∧ st Z×st Z≤m }}
  END
    {{ fun st ⇒ (st X = m ∧ st Z×st Z≤m)
                 ∧ ~((st Z + 1)*(st Z + 1) ≤ st X) }} ->>
    {{ fun st ⇒ st Z×st Z≤m ∧ m<(st Z+1)*(st Z+1) }}.

Theorem sqrt_correct : ∀ m,
  dec_correct (sqrt_dec m).
Proof. intro m. verify. Qed.
```

Squaring

Again, there are several ways of annotating the squaring program. The simplest variant we've found, square_simpler_dec, is given last.

```
Definition square_dec (m : nat) : decorated :=
  {{ fun st ⇒ st X = m }}
  Y ::= X
  {{ fun st ⇒ st X = m ∧ st Y = m }};;
  Z ::= 0
  {{ fun st ⇒ st X = m ∧ st Y = m ∧ st Z = 0}} ->>
  {{ fun st ⇒ st Z + st X × st Y = m × m }};;
  WHILE ~(Y = 0) DO
    {{ fun st ⇒ st Z + st X × st Y = m × m ∧ st Y ≠ 0 }} ->>
    {{ fun st ⇒ (st Z + st X) + st X × (st Y - 1) = m × m }}
    Z ::= Z + X
    {{ fun st ⇒ st Z + st X × (st Y - 1) = m × m }};;
    Y ::= Y - 1
    {{ fun st ⇒ st Z + st X × st Y = m × m }}
  END
  {{ fun st ⇒ st Z + st X × st Y = m × m ∧ st Y = 0 }} ->>
  {{ fun st ⇒ st Z = m × m }}.
```

```
Theorem square_dec_correct : ∀ m,
  dec_correct (square_dec m).
Proof.
  intro n. verify.
  -
    destruct (st Y) as [| y']. apply False_ind. apply H0.
    reflexivity.
    simpl. rewrite ← minus_n_O.
    assert (G : ∀ n m, n × S m = n + n × m). {
      clear. intros. induction n. reflexivity. simpl.
      rewrite IHn. omega. }
    rewrite ← H. rewrite G. rewrite plus_assoc. reflexivity.
Qed.

Definition square_dec' (n : nat) : decorated :=
  {{ fun st ⇒ True }}
  X ::= n
  {{ fun st ⇒ st X = n }};;
  Y ::= X
  {{ fun st ⇒ st X = n ∧ st Y = n }};;
  Z ::= 0
  {{ fun st ⇒ st X = n ∧ st Y = n ∧ st Z = 0 }} ->>
  {{ fun st ⇒ st Z = st X × (st X - st Y)
                ∧ st X = n ∧ st Y ≤ st X }};;
  WHILE ~(Y = 0) DO
    {{ fun st ⇒ (st Z = st X × (st X - st Y)
                 ∧ st X = n ∧ st Y ≤ st X)
                  ∧ st Y ≠ 0 }}
    Z ::= Z + X
    {{ fun st ⇒ st Z = st X × (st X - (st Y - 1))
                  ∧ st X = n ∧ st Y ≤ st X }};;
    Y ::= Y - 1
    {{ fun st ⇒ st Z = st X × (st X - st Y)
                  ∧ st X = n ∧ st Y ≤ st X }}
  END
  {{ fun st ⇒ (st Z = st X × (st X - st Y)
               ∧ st X = n ∧ st Y ≤ st X)
                ∧ st Y = 0 }} ->>
  {{ fun st ⇒ st Z = n × n }}.

Theorem square_dec'_correct : ∀ n,
  dec_correct (square_dec' n).
Proof.
  intro n. verify.
```

```
  -
    rewrite minus_diag. omega.
  - subst.
    rewrite mult_minus_distr_l.
    repeat rewrite mult_minus_distr_l. rewrite mult_1_r.
    assert (G : ∀ n m p,
                    m ≤ n → p ≤ m → n - (m - p) = n - m + p).
      intros. omega.
    rewrite G. reflexivity. apply mult_le_compat_l. assumption.
    destruct (st Y). apply False_ind. apply H0. reflexivity.
      clear. rewrite mult_succ_r. rewrite plus_comm.
      apply le_plus_l.
  -
    rewrite ← minus_n_O. reflexivity.
Qed.

Definition square_simpler_dec (m : nat) : decorated :=
  {{ fun st ⇒ st X = m }} -»
  {{ fun st ⇒ 0 = 0×m ∧ st X = m }}
  Y ::= 0
  {{ fun st ⇒ 0 = (st Y)*m ∧ st X = m }};;
  Z ::= 0
  {{ fun st ⇒ st Z = (st Y)*m ∧ st X = m }}-»
  {{ fun st ⇒ st Z = (st Y)*m ∧ st X = m }};;
  WHILE ~(Y = X) DO
    {{ fun st ⇒ (st Z = (st Y)*m ∧ st X = m)
        ∧ st Y ≠ st X }} -»
    {{ fun st ⇒ st Z + st X = ((st Y) + 1)*m ∧ st X = m }}
    Z ::= Z + X
    {{ fun st ⇒ st Z = ((st Y) + 1)*m ∧ st X = m }};;
    Y ::= Y + 1
    {{ fun st ⇒ st Z = (st Y)*m ∧ st X = m }}
  END
  {{ fun st ⇒ (st Z = (st Y)*m ∧ st X = m) ∧ st Y = st X }} -»
  {{ fun st ⇒ st Z = m×m }}.

Theorem square_simpler_dec_correct : ∀ m,
  dec_correct (square_simpler_dec m).
Proof.
  intro m. verify.
  rewrite mult_plus_distr_r. simpl. rewrite ← plus_n_O.
  reflexivity.
Qed.
```

Two loops

```
Definition two_loops_dec (a b c : nat) : decorated :=
  {{ fun st ⇒ True }} ->>
  {{ fun st ⇒ c = 0 + c ∧ 0 = 0 }}
  X ::= 0
  {{ fun st ⇒ c = st X + c ∧ 0 = 0 }};;
  Y ::= 0
  {{ fun st ⇒ c = st X + c ∧ st Y = 0 }};;
  Z ::= c
  {{ fun st ⇒ st Z = st X + c ∧ st Y = 0 }};;
  WHILE ~(X = a) DO
    {{ fun st ⇒ (st Z = st X + c ∧ st Y = 0) ∧ st X ≠ a }} ->>
    {{ fun st ⇒ st Z + 1 = st X + 1 + c ∧ st Y = 0 }}
    X ::= X + 1
    {{ fun st ⇒ st Z + 1 = st X + c ∧ st Y = 0 }};;
    Z ::= Z + 1
    {{ fun st ⇒ st Z = st X + c ∧ st Y = 0 }}
  END
  {{ fun st ⇒ (st Z = st X + c ∧ st Y = 0) ∧ st X = a }} ->>
  {{ fun st ⇒ st Z = a + st Y + c }};;
  WHILE ~(Y = b) DO
    {{ fun st ⇒ st Z = a + st Y + c ∧ st Y ≠ b }} ->>
    {{ fun st ⇒ st Z + 1 = a + st Y + 1 + c }}
    Y ::= Y + 1
    {{ fun st ⇒ st Z + 1 = a + st Y + c }};;
    Z ::= Z + 1
    {{ fun st ⇒ st Z = a + st Y + c }}
  END
  {{ fun st ⇒ (st Z = a + st Y + c) ∧ st Y = b }} ->>
  {{ fun st ⇒ st Z = a + b + c }}.

Theorem two_loops_correct : ∀ a b c,
  dec_correct (two_loops_dec a b c).
Proof. intros a b c. verify. Qed.
```

Power Series

```
Fixpoint pow2 n :=
  match n with
  | 0 ⇒ 1
  | S n' ⇒ 2 × (pow2 n')
  end.
```

```
Definition dpow2_down (n : nat) :=
  {{ fun st ⇒ True }} ->>
  {{ fun st ⇒ 1 = (pow2 (0 + 1))-1 ∧ 1 = pow2 0 }}
  X ::= 0
  {{ fun st ⇒ 1 = (pow2 (0 + 1))-1 ∧ 1 = pow2 (st X) }};;
  Y ::= 1
  {{ fun st ⇒ st Y = (pow2 (st X + 1))-1 ∧ 1 = pow2 (st X) }};;
  Z ::= 1
  {{ fun st ⇒ st Y = (pow2 (st X + 1))-1 ∧ st Z = pow2 (st X) }};;
  WHILE ~(X = n) DO
    {{ fun st ⇒ (st Y = (pow2 (st X + 1))-1 ∧ st Z = pow2 (st X))
                ∧ st X ≠ n }} ->>
    {{ fun st ⇒ st Y + 2 × st Z = (pow2 (st X + 2))-1
                ∧ 2 × st Z = pow2 (st X + 1) }}
    Z ::= 2 × Z
    {{ fun st ⇒ st Y + st Z = (pow2 (st X + 2))-1
                ∧ st Z = pow2 (st X + 1) }};;
    Y ::= Y + Z
    {{ fun st ⇒ st Y = (pow2 (st X + 2))-1
                ∧ st Z = pow2 (st X + 1) }};;
    X ::= X + 1
    {{ fun st ⇒ st Y = (pow2 (st X + 1))-1
                ∧ st Z = pow2 (st X) }}
  END
  {{ fun st ⇒ (st Y = (pow2 (st X + 1))-1 ∧ st Z = pow2 (st X))
              ∧ st X = n }} ->>
  {{ fun st ⇒ st Y = pow2 (n+1) - 1 }}.
Lemma pow2_plus_1 : ∀ n,
  pow2 (n+1) = pow2 n + pow2 n.
Proof. induction n; simpl. reflexivity. omega. Qed.
Lemma pow2_le_1 : ∀ n, pow2 n ≥ 1.
Proof. induction n. simpl. constructor. simpl. omega. Qed.
Theorem dpow2_down_correct : ∀ n,
  dec_correct (dpow2_down n).
Proof.
  intro m. verify.
  -
    rewrite pow2_plus_1. rewrite ← H0. reflexivity.
  -
    rewrite ← plus_n_O.
    rewrite ← pow2_plus_1. remember (st X) as n.
    replace (pow2 (n + 1) - 1 + pow2 (n + 1))
```

```
      with (pow2 (n + 1) + pow2 (n + 1) - 1) by omega.
    rewrite ← pow2_plus_1.
    replace (n + 1 + 1) with (n + 2) by omega.
    reflexivity.
  -
    rewrite ← plus_n_O. rewrite ← pow2_plus_1.
    reflexivity.
  -
    replace (st X + 1 + 1) with (st X + 2) by omega.
    reflexivity.
Qed.
```

6.4.5 Further Exercises

Exercise: 3 stars, advanced (slow_assignment_dec) In the *slow_assignment* exercise above, we saw a roundabout way of assigning a number currently stored in X to the variable Y: start Y at 0, then decrement X until it hits 0, incrementing Y at each step. Write a formal version of this decorated program and prove it correct.

```
Example slow_assignment_dec (m : nat) : decorated
  . Admitted.

Theorem slow_assignment_dec_correct : ∀ m,
  dec_correct (slow_assignment_dec m).
Proof. Admitted.

Definition manual_grade_for_check_defn_of_slow_assignment_dec : option (nat×string) :=
None.
```

☐

Exercise: 4 stars, advanced (factorial_dec) Remember the factorial function we worked with before:

```
Fixpoint real_fact (n : nat) : nat :=
  match n with
  | O ⇒ 1
  | S n' ⇒ n × (real_fact n')
  end.
```

Following the pattern of subtract_slowly_dec, write a decorated program *factorial_dec* that implements the factorial function and prove it correct as *factorial_dec_correct*.

```
Definition manual_grade_for_factorial_dec : option (nat×string) := None.
```

☐

Exercise: 4 stars, advanced, optional (fib_eqn) The Fibonacci function is usually written like this:

Fixpoint fib n := match n with | 0 => 1 | 1 => 1 | _ => fib (pred n) + fib (pred (pred n)) end.

This doesn't pass Coq's termination checker, but here is a slightly clunkier definition that does:

```
Fixpoint fib n :=
  match n with
  | 0 ⇒ 1
  | S n' ⇒ match n' with
             | 0 ⇒ 1
             | S n'' ⇒ fib n' + fib n''
             end
  end.
```

Prove that fib satisfies the following equation:

```
Lemma fib_eqn : ∀ n,
  n > 0 →
  fib n + fib (Init.Nat.pred n) = fib (n + 1).
Proof.
  Admitted.
```

□

Exercise: 4 stars, advanced, optional (fib) The following Imp program leaves the value of fib n in the variable Y when it terminates:

X ::= 1;; Y ::= 1;; Z ::= 1;; WHILE ~(X = n + 1) DO T ::= Z;; Z ::= Z + Y;; Y ::= T;; X ::= X + 1 END

Fill in the following definition of dfib and prove that it satisfies this specification:

[219] dfib [220]

```
Definition T : string := "T".

Definition dfib (n : nat) : decorated
. Admitted.

Theorem dfib_correct : ∀ n,
  dec_correct (dfib n).
  Admitted.
```

□

Exercise: 5 stars, advanced, optional (improve_dcom) The formal decorated programs defined in this section are intended to look as similar as possible to the informal ones

[219] True

[220] Y=fibn

defined earlier in the chapter. If we drop this requirement, we can eliminate almost all annotations, just requiring final postconditions and loop invariants to be provided explicitly. Do this – i.e., define a new version of dcom with as few annotations as possible and adapt the rest of the formal development leading up to the verification_correct theorem.

Exercise: 4 stars, advanced, optional (implement_dcom) Adapt the parser for Imp presented in chapter *ImpParser* of *Logical Foundations* to parse decorated commands (either ours or, even better, the ones you defined in the previous exercise).

Chapter 7

HoareAsLogic: Hoare Logic as a Logic

The presentation of Hoare logic in chapter Hoare could be described as "model-theoretic": the proof rules for each of the constructors were presented as *theorems* about the evaluation behavior of programs, and proofs of program correctness (validity of Hoare triples) were constructed by combining these theorems directly in Coq.

Another way of presenting Hoare logic is to define a completely separate proof system – a set of axioms and inference rules that talk about commands, Hoare triples, etc. – and then say that a proof of a Hoare triple is a valid derivation in *that* logic. We can do this by giving an inductive definition of *valid derivations* in this new logic.

This chapter is optional. Before reading it, you'll want to read the *ProofObjects* chapter in *Logical Foundations* (*Software Foundations*, volume 1).

```
From PLF Require Import Imp.
From PLF Require Import Hoare.
```

7.1 Definitions

```
Inductive hoare_proof : Assertion → com → Assertion → Type :=
  | H_Skip : ∀ P,
      hoare_proof P (SKIP) P
  | H_Asgn : ∀ Q V a,
      hoare_proof (assn_sub V a Q) (V ::= a) Q
  | H_Seq : ∀ P c Q d R,
      hoare_proof P c Q → hoare_proof Q d R → hoare_proof P (c;;d) R
  | H_If : ∀ P Q b c1 c2,
    hoare_proof (fun st ⇒ P st ∧ bassn b st) c1 Q →
    hoare_proof (fun st ⇒ P st ∧ ~(bassn b st)) c2 Q →
    hoare_proof P (TEST b THEN c1 ELSE c2 FI) Q
  | H_While : ∀ P b c,
    hoare_proof (fun st ⇒ P st ∧ bassn b st) c P →
    hoare_proof P (WHILE b DO c END) (fun st ⇒ P st ∧ ¬ (bassn b st))
```

```
| H_Consequence : ∀ (P Q P' Q' : Assertion) c,
    hoare_proof P' c Q' →
    (∀ st, P st → P' st) →
    (∀ st, Q' st → Q st) →
    hoare_proof P c Q.
```

We don't need to include axioms corresponding to hoare_consequence_pre or hoare_consequence_post, because these can be proven easily from H_Consequence.

```
Lemma H_Consequence_pre : ∀ (P Q P': Assertion) c,
    hoare_proof P' c Q →
    (∀ st, P st → P' st) →
    hoare_proof P c Q.
Proof.
  intros. eapply H_Consequence.
    apply X. apply H. intros. apply H0. Qed.
Lemma H_Consequence_post : ∀ (P Q Q' : Assertion) c,
    hoare_proof P c Q' →
    (∀ st, Q' st → Q st) →
    hoare_proof P c Q.
Proof.
  intros. eapply H_Consequence.
    apply X. intros. apply H0. apply H. Qed.
```

As an example, let's construct a proof object representing a derivation for the hoare triple
[1] X::=X+1 ;; X::=X+2 [2].

We can use Coq's tactics to help us construct the proof object.

```
Example sample_proof :
  hoare_proof
    ((fun st:state ⇒ st X = 3) [X |-> X + 2] [X |-> X + 1])
    (X ::= X + 1;; X ::= X + 2)
    (fun st:state ⇒ st X = 3).
Proof.
  eapply H_Seq; apply H_Asgn.
Qed.
```

7.2 Properties

Exercise: 2 stars, standard (hoare_proof_sound) Prove that such proof objects represent true claims.

```
Theorem hoare_proof_sound : ∀ P c Q,
```

[1] (X=3) [X|->X+2] [X|->X+1]

[2] X=3

hoare_proof P c Q → {{P}} c {{Q}}.

Proof.

Admitted.

□

We can also use Coq's reasoning facilities to prove metatheorems about Hoare Logic. For example, here are the analogs of two theorems we saw in chapter Hoare – this time expressed in terms of the syntax of Hoare Logic derivations (provability) rather than directly in terms of the semantics of Hoare triples.

The first one says that, for every P and c, the assertion $\{\{P\}\}$ c $\{\{\textbf{True}\}\}$ is *provable* in Hoare Logic. Note that the proof is more complex than the semantic proof in Hoare: we actually need to perform an induction over the structure of the command c.

```
Theorem H_Post_True_deriv:
  ∀ c P, hoare_proof P c (fun _ ⇒ True).
Proof.
  intro c.
  induction c; intro P.
  -
    eapply H_Consequence.
    apply H_Skip.
    intros. apply H.
    intros. apply I.
  -
    eapply H_Consequence_pre.
    apply H_Asgn.
    intros. apply I.
  -
    eapply H_Consequence_pre.
    eapply H_Seq.
    apply (IHc1 (fun _ ⇒ True)).
    apply IHc2.
    intros. apply I.
  -
    apply H_Consequence_pre with (fun _ ⇒ True).
    apply H_If.
    apply IHc1.
    apply IHc2.
    intros. apply I.
  -
    eapply H_Consequence.
    eapply H_While.
    eapply IHc.
    intros; apply I.
```

```
    intros; apply I.
Qed.
```

Similarly, we can show that {{**False**}} c {{Q}} is provable for any c and Q.

```
Lemma False_and_P_imp: ∀ P Q,
  False ∧ P → Q.
Proof.
  intros P Q [CONTRA HP].
  destruct CONTRA.
Qed.

Tactic Notation "pre_false_helper" constr(CONSTR) :=
  eapply H_Consequence_pre;
    [eapply CONSTR | intros ? CONTRA; destruct CONTRA].

Theorem H_Pre_False_deriv:
  ∀ c Q, hoare_proof (fun _ ⇒ False) c Q.
Proof.
  intros c.
  induction c; intro Q.
  - pre_false_helper H_Skip.
  - pre_false_helper H_Asgn.
  - pre_false_helper H_Seq. apply IHc1. apply IHc2.
  -
    apply H_If; eapply H_Consequence_pre.
    apply IHc1. intro. eapply False_and_P_imp.
    apply IHc2. intro. eapply False_and_P_imp.
  -
    eapply H_Consequence_post.
    eapply H_While.
    eapply H_Consequence_pre.
      apply IHc.
      intro. eapply False_and_P_imp.
    intro. simpl. eapply False_and_P_imp.
Qed.
```

As a last step, we can show that the set of **hoare_proof** axioms is sufficient to prove any true fact about (partial) correctness. More precisely, any semantic Hoare triple that we can prove can also be proved from these axioms. Such a set of axioms is said to be *relatively complete*. Our proof is inspired by this one:

http://www.ps.uni-saarland.de/courses/sem-ws11/script/Hoare.html

To carry out the proof, we need to invent some intermediate assertions using a technical device known as *weakest preconditions*. Given a command c and a desired postcondition assertion Q, the weakest precondition wp c Q is an assertion P such that {{P}} c {{Q}} holds, and moreover, for any other assertion P', if {{P'}} c {{Q}} holds then $P' \rightarrow P$. We can more directly define this as follows:

```
Definition wp (c:com) (Q:Assertion) : Assertion :=
  fun s ⇒ ∀ s', s =[ c ]=> s' → Q s'.
```

Exercise: 1 star, standard (wp_is_precondition) Lemma wp_is_precondition: ∀ c Q,

```
  {{wp c Q}} c {{Q}}.
   Admitted.
```

☐

Exercise: 1 star, standard (wp_is_weakest) Lemma wp_is_weakest: ∀ c Q P',

```
   {{P'}} c {{Q}} → ∀ st, P' st → wp c Q st.
   Admitted.
```

The following utility lemma will also be useful.

```
Lemma bassn_eval_false : ∀ b st, ¬ bassn b st → beval st b = false.
Proof.
  intros b st H. unfold bassn in H. destruct (beval st b).
    exfalso. apply H. reflexivity.
    reflexivity.
Qed.
```

☐

Exercise: 5 stars, standard (hoare_proof_complete) Complete the proof of the theorem.

```
Theorem hoare_proof_complete: ∀ P c Q,
  {{P}} c {{Q}} → hoare_proof P c Q.
Proof.
  intros P c. generalize dependent P.
  induction c; intros P Q HT.
  -
    eapply H_Consequence.
     eapply H_Skip.
      intros. eassumption.
      intro st. apply HT. apply E_Skip.
  -
    eapply H_Consequence.
      eapply H_Asgn.
      intro st. apply HT. econstructor. reflexivity.
      intros; assumption.
  -
    apply H_Seq with (wp c2 Q).
```

```
eapply IHc1.
  intros st st' E1 H. unfold wp. intros st'' E2.
    eapply HT. econstructor; eassumption. assumption.
eapply IHc2. intros st st' E1 H. apply H; assumption.
```
Admitted.

☐

Finally, we might hope that our axiomatic Hoare logic is *decidable*; that is, that there is an (terminating) algorithm (a *decision procedure*) that can determine whether or not a given Hoare triple is valid (derivable). But such a decision procedure cannot exist!

Consider the triple {{**True**}} c {{**False**}}. This triple is valid if and only if c is non-terminating. So any algorithm that could determine validity of arbitrary triples could solve the Halting Problem.

Similarly, the triple {{**True**}} *SKIP* {{*P*}} is valid if and only if ∀ s, *P* s is valid, where *P* is an arbitrary assertion of Coq's logic. But it is known that there can be no decision procedure for this logic.

Overall, this axiomatic style of presentation gives a clearer picture of what it means to "give a proof in Hoare logic." However, it is not entirely satisfactory from the point of view of writing down such proofs in practice: it is quite verbose. The section of chapter Hoare2 on formalizing decorated programs shows how we can do even better.

Chapter 8

Smallstep: Small-step Operational Semantics

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Arith.Arith.
From Coq Require Import Arith.EqNat.
From Coq Require Import Init.Nat.
From Coq Require Import omega.Omega.
From Coq Require Import Lists.List.
Import ListNotations.
From PLF Require Import Maps.
From PLF Require Import Imp.
```

The evaluators we have seen so far (for **aexp**s, **bexp**s, commands, ...) have been formulated in a "big-step" style: they specify how a given expression can be evaluated to its final value (or a command plus a store to a final store) "all in one big step."

This style is simple and natural for many purposes – indeed, Gilles Kahn, who popularized it, called it *natural semantics*. But there are some things it does not do well. In particular, it does not give us a natural way of talking about *concurrent* programming languages, where the semantics of a program – i.e., the essence of how it behaves – is not just which input states get mapped to which output states, but also includes the intermediate states that it passes through along the way, since these states can also be observed by concurrently executing code.

Another shortcoming of the big-step style is more technical, but critical in many situations. Suppose we want to define a variant of Imp where variables could hold *either* numbers *or* lists of numbers. In the syntax of this extended language, it will be possible to write strange expressions like 2 + nil, and our semantics for arithmetic expressions will then need to say something about how such expressions behave. One possibility is to maintain the convention that every arithmetic expression evaluates to some number by choosing some way of viewing a list as a number – e.g., by specifying that a list should be interpreted as 0 when it occurs in a context expecting a number. But this is really a bit of a hack.

A much more natural approach is simply to say that the behavior of an expression like 2+nil is *undefined* – i.e., it doesn't evaluate to any result at all. And we can easily do this: we just have to formulate aeval and beval as `Inductive` propositions rather than `Fixpoint`s, so that we can make them partial functions instead of total ones.

Now, however, we encounter a serious deficiency. In this language, a command might fail to map a given starting state to any ending state for *two quite different reasons*: either because the execution gets into an infinite loop or because, at some point, the program tries to do an operation that makes no sense, such as adding a number to a list, so that none of the evaluation rules can be applied.

These two outcomes – nontermination vs. getting stuck in an erroneous configuration – should not be confused. In particular, we want to *allow* the first (permitting the possibility of infinite loops is the price we pay for the convenience of programming with general looping constructs like *while*) but *prevent* the second (which is just wrong), for example by adding some form of *typechecking* to the language. Indeed, this will be a major topic for the rest of the course. As a first step, we need a way of presenting the semantics that allows us to distinguish nontermination from erroneous "stuck states."

So, for lots of reasons, we'd often like to have a finer-grained way of defining and reasoning about program behaviors. This is the topic of the present chapter. Our goal is to replace the "big-step" `eval` relation with a "small-step" relation that specifies, for a given program, how the "atomic steps" of computation are performed.

8.1 A Toy Language

To save space in the discussion, let's go back to an incredibly simple language containing just constants and addition. (We use single letters – C and *P* (for Constant and Plus) – as constructor names, for brevity.) At the end of the chapter, we'll see how to apply the same techniques to the full Imp language.

```
Inductive tm : Type :=
  | C : nat → tm
  | P : tm → tm → tm.
```

Here is a standard evaluator for this language, written in the big-step style that we've been using up to this point.

```
Fixpoint evalF (t : tm) : nat :=
  match t with
  | C n ⇒ n
  | P a1 a2 ⇒ evalF a1 + evalF a2
  end.
```

Here is the same evaluator, written in exactly the same style, but formulated as an inductively defined relation. We use the notation t ==> n for "t evaluates to n."

(E_Const) C n ==> n

t1 ==> n1 t2 ==> n2

(E_Plus) P t1 t2 ==> n1 + n2

```
Reserved Notation " t '==>' n " (at level 50, left associativity).

Inductive eval : tm → nat → Prop :=
  | E_Const : ∀ n,
      C n ==> n
  | E_Plus : ∀ t1 t2 n1 n2,
      t1 ==> n1 →
      t2 ==> n2 →
      P t1 t2 ==> (n1 + n2)
where " t '==>' n " := (eval t n).

Module SimpleArith1.
```

Now, here is the corresponding *small-step* evaluation relation.

(ST_PlusConstConst) P (C n1) (C n2) –> C (n1 + n2)

t1 –> t1'

(ST_Plus1) P t1 t2 –> P t1' t2

t2 –> t2'

(ST_Plus2) P (C n1) t2 –> P (C n1) t2'

```
Reserved Notation " t '->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_PlusConstConst : ∀ n1 n2,
      P (C n1) (C n2) -> C (n1 + n2)
  | ST_Plus1 : ∀ t1 t1' t2,
      t1 -> t1' →
      P t1 t2 -> P t1' t2
  | ST_Plus2 : ∀ n1 t2 t2',
      t2 -> t2' →
      P (C n1) t2 -> P (C n1) t2'

  where " t '->' t' " := (step t t').
```

Things to notice:

- We are defining just a single reduction step, in which one *P* node is replaced by its value.
- Each step finds the *leftmost* *P* node that is ready to go (both of its operands are constants) and rewrites it in place. The first rule tells how to rewrite this *P* node itself; the other two rules tell how to find it.

- A term that is just a constant cannot take a step.

Let's pause and check a couple of examples of reasoning with the **step** relation...
If *t1* can take a step to *t1'*, then *P t1 t2* steps to *P t1' t2*:

```
Example test_step_1 :
      P
        (P (C 0) (C 3))
        (P (C 2) (C 4))
      ->
      P
        (C (0 + 3))
        (P (C 2) (C 4)).
Proof.
  apply ST_Plus1. apply ST_PlusConstConst. Qed.
```

Exercise: 1 star, standard (test_step_2) Right-hand sides of sums can take a step only when the left-hand side is finished: if *t2* can take a step to *t2'*, then *P* (C n) *t2* steps to *P* (C n) *t2'*:

```
Example test_step_2 :
      P
        (C 0)
        (P
          (C 2)
          (P (C 0) (C 3)))
      ->
      P
        (C 0)
        (P
          (C 2)
          (C (0 + 3))).
Proof.
   Admitted.
   □
End SimpleArith1.
```

8.2 Relations

We will be working with several different single-step relations, so it is helpful to generalize a bit and state a few definitions and theorems about relations in general. (The optional chapter *Rel.v* develops some of these ideas in a bit more detail; it may be useful if the treatment here is too dense.)

A *binary relation* on a set X is a family of propositions parameterized by two elements of X – i.e., a proposition about pairs of elements of X.

```
Definition relation (X : Type) := X → X → Prop.
```

Our main examples of such relations in this chapter will be the single-step reduction relation, –>, and its multi-step variant, –>* (defined below), but there are many other examples – e.g., the "equals," "less than," "less than or equal to," and "is the square of" relations on numbers, and the "prefix of" relation on lists and strings.

One simple property of the –> relation is that, like the big-step evaluation relation for Imp, it is *deterministic*.

Theorem: For each *t*, there is at most one *t'* such that *t* steps to *t'* (*t* –> *t'* is provable).

Proof sketch: We show that if x steps to both *y1* and *y2*, then *y1* and *y2* are equal, by induction on a derivation of **step** x *y1*. There are several cases to consider, depending on the last rule used in this derivation and the last rule in the given derivation of **step** x *y2*.

- If both are ST_PlusConstConst, the result is immediate.
- The cases when both derivations end with ST_Plus1 or ST_Plus2 follow by the induction hypothesis.
- It cannot happen that one is ST_PlusConstConst and the other is ST_Plus1 or ST_Plus2, since this would imply that x has the form P *t1 t2* where both *t1* and *t2* are constants (by ST_PlusConstConst) *and* one of *t1* or *t2* has the form P _.
- Similarly, it cannot happen that one is ST_Plus1 and the other is ST_Plus2, since this would imply that x has the form P *t1 t2* where *t1* has both the form P *t11 t12* and the form C n. □

Formally:

```
Definition deterministic {X : Type} (R : relation X) :=
  ∀ x y1 y2 : X, R x y1 → R x y2 → y1 = y2.

Module SimpleArith2.
Import SimpleArith1.

Theorem step_deterministic:
  deterministic step.
Proof.
  unfold deterministic. intros x y1 y2 Hy1 Hy2.
  generalize dependent y2.
  induction Hy1; intros y2 Hy2.
  - inversion Hy2.
    + reflexivity.
    + inversion H2.
    + inversion H2.
```

```
  - inversion Hy2.
    +
      rewrite ← H0 in Hy1. inversion Hy1.
    +
      rewrite ← (IHHy1 t1'0).
      reflexivity. assumption.
    +
      rewrite ← H in Hy1. inversion Hy1.
  - inversion Hy2.
    +
      rewrite ← H1 in Hy1. inversion Hy1.
    + inversion H2.
    +
      rewrite ← (IHHy1 t2'0).
      reflexivity. assumption.
Qed.
```

End SimpleArith2.

There is some annoying repetition in this proof. Each use of `inversion` *Hy2* results in three subcases, only one of which is relevant (the one that matches the current case in the induction on *Hy1*). The other two subcases need to be dismissed by finding the contradiction among the hypotheses and doing inversion on it.

The following custom tactic, called *solve_by_inverts*, can be helpful in such cases. It will solve the goal if it can be solved by inverting some hypothesis; otherwise, it fails.

```
Ltac solve_by_inverts n :=
  match goal with | H : ?T ⊢ _ ⇒
  match type of T with Prop ⇒
    solve [
      inversion H;
      match n with S (S (?n')) ⇒ subst; solve_by_inverts (S n') end ]
  end end.
```

The details of how this works are not important for now, but it illustrates the power of Coq's `Ltac` language for programmatically defining special-purpose tactics. It looks through the current proof state for a hypothesis *H* (the first `match`) of type `Prop` (the second `match`) such that performing inversion on *H* (followed by a recursive invocation of the same tactic, if its argument `n` is greater than one) completely solves the current goal. If no such hypothesis exists, it fails.

We will usually want to call *solve_by_inverts* with argument 1 (especially as larger arguments can lead to very slow proof checking), so we define *solve_by_invert* as a shorthand for this case.

```
Ltac solve_by_invert :=
  solve_by_inverts 1.
```

Let's see how a proof of the previous theorem can be simplified using this tactic...

```
Module SimpleArith3.
Import SimpleArith1.

Theorem step_deterministic_alt: deterministic step.
Proof.
  intros x y1 y2 Hy1 Hy2.
  generalize dependent y2.
  induction Hy1; intros y2 Hy2;
    inversion Hy2; subst; try solve_by_invert.
  - reflexivity.
  -
    apply IHHy1 in H2. rewrite H2. reflexivity.
  -
    apply IHHy1 in H2. rewrite H2. reflexivity.
Qed.

End SimpleArith3.
```

8.2.1 Values

Next, it will be useful to slightly reformulate the definition of single-step reduction by stating it in terms of "values."

It can be useful to think of the –> relation as defining an *abstract machine*:

- At any moment, the *state* of the machine is a term.
- A *step* of the machine is an atomic unit of computation – here, a single "add" operation.
- The *halting states* of the machine are ones where there is no more computation to be done.

We can then execute a term t as follows:

- Take t as the starting state of the machine.
- Repeatedly use the –> relation to find a sequence of machine states, starting with t, where each state steps to the next.
- When no more reduction is possible, "read out" the final state of the machine as the result of execution.

Intuitively, it is clear that the final states of the machine are always terms of the form C n for some n. We call such terms *values*.

```
Inductive value : tm → Prop :=
```

| v_const : ∀ n, **value** (C n).

Having introduced the idea of values, we can use it in the definition of the –> relation to write ST_Plus2 rule in a slightly more elegant way:

(ST_PlusConstConst) P (C n1) (C n2) –> C (n1 + n2)

t1 –> t1'

(ST_Plus1) P t1 t2 –> P t1' t2

value v1 t2 –> t2'

(ST_Plus2) P v1 t2 –> P v1 t2'

Again, the variable names here carry important information: by convention, *v1* ranges only over values, while *t1* and *t2* range over arbitrary terms. (Given this convention, the explicit **value** hypothesis is arguably redundant. We'll keep it for now, to maintain a close correspondence between the informal and Coq versions of the rules, but later on we'll drop it in informal rules for brevity.)

Here are the formal rules:

```
Reserved Notation " t '->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_PlusConstConst : ∀ n1 n2,
          P (C n1) (C n2)
      -> C (n1 + n2)
  | ST_Plus1 : ∀ t1 t1' t2,
        t1 -> t1' →
        P t1 t2 -> P t1' t2
  | ST_Plus2 : ∀ v1 t2 t2',
        value v1 →
        t2 -> t2' →
        P v1 t2 -> P v1 t2'

  where " t '->' t' " := (step t t').
```

Exercise: 3 stars, standard, recommended (redo_determinism) As a sanity check on this change, let's re-verify determinism. Here's an informal proof:

Proof sketch: We must show that if x steps to both *y1* and *y2*, then *y1* and *y2* are equal. Consider the final rules used in the derivations of **step** x *y1* and **step** x *y2*.

- If both are ST_PlusConstConst, the result is immediate.
- It cannot happen that one is ST_PlusConstConst and the other is ST_Plus1 or ST_Plus2, since this would imply that x has the form *P t1 t2* where both *t1* and *t2* are constants (by ST_PlusConstConst) *and* one of *t1* or *t2* has the form *P _*.

- Similarly, it cannot happen that one is ST_Plus1 and the other is ST_Plus2, since this would imply that x has the form *P t1 t2* where *t1* both has the form *P t11 t12* and is a value (hence has the form C n).

- The cases when both derivations end with ST_Plus1 or ST_Plus2 follow by the induction hypothesis. ☐

Most of this proof is the same as the one above. But to get maximum benefit from the exercise you should try to write your formal version from scratch and just use the earlier one if you get stuck.

```
Theorem step_deterministic :
  deterministic step.
Proof.
  Admitted.
  ☐
```

8.2.2 Strong Progress and Normal Forms

The definition of single-step reduction for our toy language is fairly simple, but for a larger language it would be easy to forget one of the rules and accidentally create a situation where some term cannot take a step even though it has not been completely reduced to a value. The following theorem shows that we did not, in fact, make such a mistake here.

Theorem (*Strong Progress*): If *t* is a term, then either *t* is a value or else there exists a term *t'* such that *t* –> *t'*.

Proof: By induction on *t*.

- Suppose *t* = C n. Then *t* is a value.

- Suppose *t* = *P t1 t2*, where (by the IH) *t1* either is a value or can step to some *t1'*, and where *t2* is either a value or can step to some *t2'*. We must show *P t1 t2* is either a value or steps to some *t'*.

  - If *t1* and *t2* are both values, then *t* can take a step, by ST_PlusConstConst.
  - If *t1* is a value and *t2* can take a step, then so can *t*, by ST_Plus2.
  - If *t1* can take a step, then so can *t*, by ST_Plus1. ☐

Or, formally:

```
Theorem strong_progress : ∀ t,
  value t ∨ (∃ t', t -> t').
Proof.
  induction t.
  - left. apply v_const.
  - right. destruct IHt1 as [IHt1 | [t1' Ht1]].
```

```
  + destruct IHt2 as [IHt2 | [t2' Ht2]].
    × inversion IHt1. inversion IHt2.
      ∃ (C (n + n0)).
      apply ST_PlusConstConst.
    ×
      ∃ (P t1 t2').
      apply ST_Plus2. apply IHt1. apply Ht2.
  +
    ∃ (P t1' t2).
    apply ST_Plus1. apply Ht1.
Qed.
```

This important property is called *strong progress*, because every term either is a value or can "make progress" by stepping to some other term. (The qualifier "strong" distinguishes it from a more refined version that we'll see in later chapters, called just *progress*.)

The idea of "making progress" can be extended to tell us something interesting about values: in this language, values are exactly the terms that *cannot* make progress in this sense.

To state this observation formally, let's begin by giving a name to terms that cannot make progress. We'll call them *normal forms*.

```
Definition normal_form {X : Type} (R : relation X) (t : X) : Prop :=
  ¬ ∃ t', R t t'.
```

Note that this definition specifies what it is to be a normal form for an *arbitrary* relation R over an arbitrary set X, not just for the particular single-step reduction relation over terms that we are interested in at the moment. We'll re-use the same terminology for talking about other relations later in the course.

We can use this terminology to generalize the observation we made in the strong progress theorem: in this language, normal forms and values are actually the same thing.

```
Lemma value_is_nf : ∀ v,
  value v → normal_form step v.
Proof.
  unfold normal_form. intros v H. inversion H.
  intros contra. inversion contra. inversion H1.
Qed.

Lemma nf_is_value : ∀ t,
  normal_form step t → value t.
Proof.   unfold normal_form. intros t H.
  assert (G : value t ∨ ∃ t', t -> t').
  { apply strong_progress. }
  destruct G as [G | G].
  - apply G.
  - exfalso. apply H. assumption.
```

```
Qed.

Corollary nf_same_as_value : ∀ t,
  normal_form step t ↔ value t.
Proof.
  split. apply nf_is_value. apply value_is_nf.
Qed.
```

Why is this interesting?

Because **value** is a syntactic concept – it is defined by looking at the form of a term – while normal_form is a semantic one – it is defined by looking at how the term steps.

It is not obvious that these concepts should coincide!

Indeed, we could easily have written the definitions incorrectly so that they would *not* coincide.

Exercise: 3 stars, standard, optional (value_not_same_as_normal_form1) We might, for example, wrongly define **value** so that it includes some terms that are not finished reducing.

(Even if you don't work this exercise and the following ones in Coq, make sure you can think of an example of such a term.)

```
Module Temp1.

Inductive value : tm → Prop :=
  | v_const : ∀ n, value (C n)
  | v_funny : ∀ t1 n2,
                  value (P t1 (C n2)).

Reserved Notation " t '-->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_PlusConstConst : ∀ n1 n2,
      P (C n1) (C n2) --> C (n1 + n2)
  | ST_Plus1 : ∀ t1 t1' t2,
      t1 --> t1' →
      P t1 t2 --> P t1' t2
  | ST_Plus2 : ∀ v1 t2 t2',
      value v1 →
      t2 --> t2' →
      P v1 t2 --> P v1 t2'

  where " t '-->' t' " := (step t t').

Lemma value_not_same_as_normal_form :
  ∃ v, value v ∧ ¬ normal_form step v.
Proof.
  Admitted.
```

End Temp1.

□

Exercise: 2 stars, standard, optional (value_not_same_as_normal_form2) Alternatively, we might wrongly define **step** so that it permits something designated as a value to reduce further.

```
Module Temp2.

Inductive value : tm → Prop :=
  | v_const : ∀ n, value (C n).

Reserved Notation " t '->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_Funny : ∀ n,
      C n -> P (C n) (C 0)
  | ST_PlusConstConst : ∀ n1 n2,
      P (C n1) (C n2) -> C (n1 + n2)
  | ST_Plus1 : ∀ t1 t1' t2,
      t1 -> t1' →
      P t1 t2 -> P t1' t2
  | ST_Plus2 : ∀ v1 t2 t2',
      value v1 →
      t2 -> t2' →
      P v1 t2 -> P v1 t2'

  where " t '->' t' " := (step t t').

Lemma value_not_same_as_normal_form :
  ∃ v, value v ∧ ¬ normal_form step v.
Proof.
   Admitted.

End Temp2.
```

□

Exercise: 3 stars, standard, optional (value_not_same_as_normal_form3) Finally, we might define **value** and **step** so that there is some term that is not a value but that cannot take a step in the **step** relation. Such terms are said to be *stuck*. In this case this is caused by a mistake in the semantics, but we will also see situations where, even in a correct language definition, it makes sense to allow some terms to be stuck.

```
Module Temp3.

Inductive value : tm → Prop :=
  | v_const : ∀ n, value (C n).
```

```
Reserved Notation " t '->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_PlusConstConst : ∀ n1 n2,
      P (C n1) (C n2) -> C (n1 + n2)
  | ST_Plus1 : ∀ t1 t1' t2,
      t1 -> t1' →
      P t1 t2 -> P t1' t2

  where " t '->' t' " := (step t t').
```

(Note that ST_Plus2 is missing.)

```
Lemma value_not_same_as_normal_form :
  ∃ t, ¬ value t ∧ normal_form step t.
Proof.
  Admitted.

End Temp3.
```

□

Additional Exercises

```
Module Temp4.
```

Here is another very simple language whose terms, instead of being just addition expressions and numbers, are just the booleans true and false and a conditional expression...

```
Inductive tm : Type :=
  | tru : tm
  | fls : tm
  | test : tm → tm → tm → tm.

Inductive value : tm → Prop :=
  | v_tru : value tru
  | v_fls : value fls.

Reserved Notation " t '->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_IfTrue : ∀ t1 t2,
      test tru t1 t2 -> t1
  | ST_IfFalse : ∀ t1 t2,
      test fls t1 t2 -> t2
  | ST_If : ∀ t1 t1' t2 t3,
      t1 -> t1' →
      test t1 t2 t3 -> test t1' t2 t3

  where " t '->' t' " := (step t t').
```

Exercise: 1 star, standard (smallstep_bools) Which of the following propositions are provable? (This is just a thought exercise, but for an extra challenge feel free to prove your answers in Coq.)

```
Definition bool_step_prop1 :=
  fls -> fls.

Definition bool_step_prop2 :=
     test
       tru
       (test tru tru tru)
       (test fls fls fls)
  ->
     tru.

Definition bool_step_prop3 :=
     test
       (test tru tru tru)
       (test tru tru tru)
       fls
   ->
     test
       tru
       (test tru tru tru)
       fls.

Definition manual_grade_for_smallstep_bools : option (nat×string) := None.
```

□

Exercise: 3 stars, standard, optional (progress_bool) Just as we proved a progress theorem for plus expressions, we can do so for boolean expressions, as well.

```
Theorem strong_progress : ∀ t,
  value t ∨ (∃ t', t -> t').
Proof.
   Admitted.
```

□

Exercise: 2 stars, standard, optional (step_deterministic) Theorem step_deterministic :

```
  deterministic step.
Proof.
   Admitted.
```

□

Module TEMP5.

Exercise: 2 stars, standard (smallstep_bool_shortcut) Suppose we want to add a "short circuit" to the step relation for boolean expressions, so that it can recognize when the `then` and `else` branches of a conditional are the same value (either tru or fls) and reduce the whole conditional to this value in a single step, even if the guard has not yet been reduced to a value. For example, we would like this proposition to be provable:

test (test tru tru tru) fls fls > fls.

Write an extra clause for the step relation that achieves this effect and prove bool_step_prop4.

```
Reserved Notation " t '->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_IfTrue : ∀ t1 t2,
      test tru t1 t2 -> t1
  | ST_IfFalse : ∀ t1 t2,
      test fls t1 t2 -> t2
  | ST_If : ∀ t1 t1' t2 t3,
      t1 -> t1' →
      test t1 t2 t3 -> test t1' t2 t3


  where " t '->' t' " := (step t t').

Definition bool_step_prop4 :=
         test
            (test tru tru tru)
            fls
            fls
     ->
         fls.

Example bool_step_prop4_holds :
  bool_step_prop4.
Proof.
   Admitted.
```

□

Exercise: 3 stars, standard, optional (properties_of_altered_step) It can be shown that the determinism and strong progress theorems for the step relation in the lecture notes also hold for the definition of step given above. After we add the clause *ST_ShortCircuit*...

- Is the **step** relation still deterministic? Write yes or no and briefly (1 sentence) explain your answer.

  Optional: prove your answer correct in Coq.

End TEMP5.
End TEMP4.

8.3 Multi-Step Reduction

We've been working so far with the *single-step reduction* relation –>, which formalizes the individual steps of an abstract machine for executing programs.

We can use the same machine to reduce programs to completion – to find out what final result they yield. This can be formalized as follows:

- First, we define a *multi-step reduction relation* –>*, which relates terms *t* and *t'* if *t* can reach *t'* by any number (including zero) of single reduction steps.

- Then we define a "result" of a term *t* as a normal form that *t* can reach by multi-step reduction.

Since we'll want to reuse the idea of multi-step reduction many times, let's take a little extra trouble and define it generically.

Given a relation R (which will be –> for present purposes), we define a relation **multi** R, called the *multi-step closure of R* as follows.

```
Inductive multi {X : Type} (R : relation X) : relation X :=
  | multi_refl : ∀ (x : X), multi R x x
  | multi_step : ∀ (x y z : X),
                    R x y →
                    multi R y z →
                    multi R x z.
```

(In the *Rel* chapter of *Logical Foundations* and the Coq standard library, this relation is called *clos_refl_trans_1n*. We give it a shorter name here for the sake of readability.)

The effect of this definition is that **multi** R relates two elements x and y if

- x = y, or

- R x y, or

- there is some nonempty sequence *z1*, *z2*, ..., *zn* such that

  R x z1 R z1 z2 ... R zn y.

Thus, if R describes a single-step of computation, then *z1* ... *zn* is the sequence of intermediate steps of computation between x and y.

We write –>* for the **multi step** relation on terms.

```
Notation " t '–>*' t' " := (multi step t t') (at level 40).
```

The relation **multi** R has several crucial properties.

First, it is obviously *reflexive* (that is, ∀ x, **multi** R x x). In the case of the –>* (i.e., **multi step**) relation, the intuition is that a term can execute to itself by taking zero steps of execution.

Second, it contains R – that is, single-step executions are a particular case of multi-step executions. (It is this fact that justifies the word "closure" in the term "multi-step closure of R.")

```
Theorem multi_R : ∀ (X : Type) (R : relation X) (x y : X),
    R x y → (multi R) x y.
Proof.
  intros X R x y H.
  apply multi_step with y. apply H. apply multi_refl.
Qed.
```

Third, **multi** R is *transitive*.

```
Theorem multi_trans :
  ∀ (X : Type) (R : relation X) (x y z : X),
      multi R x y →
      multi R y z →
      multi R x z.
Proof.
  intros X R x y z G H.
  induction G.
    - assumption.
    -
      apply multi_step with y. assumption.
      apply IHG. assumption.
Qed.
```

In particular, for the **multi step** relation on terms, if *t1* –>* *t2* and *t2* –>* *t3*, then *t1* –>* *t3*.

8.3.1 Examples

Here's a specific instance of the **multi step** relation:

```
Lemma test_multistep_1:
      P
        (P (C 0) (C 3))
        (P (C 2) (C 4))
   ->*
      C ((0 + 3) + (2 + 4)).
Proof.
  apply multi_step with
            (P (C (0 + 3))
```

```
                (P (C 2) (C 4))).
  { apply ST_Plus1. apply ST_PlusConstConst. }
  apply multi_step with
            (P (C (0 + 3))
                (C (2 + 4))).
  { apply ST_Plus2. apply v_const. apply ST_PlusConstConst. }
  apply multi_R.
  { apply ST_PlusConstConst. }
Qed.
```

Here's an alternate proof of the same fact that uses `eapply` to avoid explicitly constructing all the intermediate terms.

```
Lemma test_multistep_1':
      P
        (P (C 0) (C 3))
        (P (C 2) (C 4))
  ->*
      C ((0 + 3) + (2 + 4)).
Proof.
  eapply multi_step. { apply ST_Plus1. apply ST_PlusConstConst. }
  eapply multi_step. { apply ST_Plus2. apply v_const.
                       apply ST_PlusConstConst. }
  eapply multi_step. { apply ST_PlusConstConst. }
  apply multi_refl.
Qed.
```

Exercise: 1 star, standard, optional (test_multistep_2) `Lemma test_multistep_2:`

```
  C 3 ->* C 3.
Proof.
```

Admitted.

□

Exercise: 1 star, standard, optional (test_multistep_3) `Lemma test_multistep_3:`

```
      P (C 0) (C 3)
   ->*
      P (C 0) (C 3).
Proof.
```

Admitted.

□

Exercise: 2 stars, standard (test_multistep_4) `Lemma test_multistep_4:`

```
      P
        (C 0)
```

```
        (P
          (C 2)
          (P (C 0) (C 3)))
  ->*
      P
        (C 0)
        (C (2 + (0 + 3))).
Proof.
   Admitted.
   □
```

8.3.2 Normal Forms Again

If t reduces to t' in zero or more steps and t' is a normal form, we say that "t' is a normal form of t."

```
Definition step_normal_form := normal_form step.

Definition normal_form_of (t t' : tm) :=
  (t ->* t' ∧ step_normal_form t').
```

We have already seen that, for our language, single-step reduction is deterministic – i.e., a given term can take a single step in at most one way. It follows from this that, if t can reach a normal form, then this normal form is unique. In other words, we can actually pronounce normal_form t t' as "t' is *the* normal form of t."

Exercise: 3 stars, standard, optional (normal_forms_unique)

```
Theorem normal_forms_unique:
  deterministic normal_form_of.
Proof.
  unfold deterministic. unfold normal_form_of.
  intros x y1 y2 P1 P2.
  inversion P1 as [P11 P12]; clear P1.
  inversion P2 as [P21 P22]; clear P2.
  generalize dependent y2.
   Admitted.
   □
```

Indeed, something stronger is true for this language (though not for all languages): the reduction of *any* term t will eventually reach a normal form – i.e., normal_form_of is a *total* function. Formally, we say the **step** relation is *normalizing*.

```
Definition normalizing {X : Type} (R : relation X) :=
  ∀ t, ∃ t',
    (multi R) t t' ∧ normal_form R t'.
```

To prove that **step** is normalizing, we need a couple of lemmas.

First, we observe that, if t reduces to t' in many steps, then the same sequence of reduction steps within t is also possible when t appears as the left-hand child of a P node, and similarly when t appears as the right-hand child of a P node whose left-hand child is a value.

```
Lemma multistep_congr_1 : ∀ t1 t1' t2,
     t1 ->* t1' →
     P t1 t2 ->* P t1' t2.
Proof.
  intros t1 t1' t2 H. induction H.
  - apply multi_refl.
  - apply multi_step with (P y t2).
    + apply ST_Plus1. apply H.
    + apply IHmulti.
Qed.
```

Exercise: 2 stars, standard (multistep_congr_2) Lemma multistep_congr_2 : ∀ t1 t2 t2',

```
     value t1 →
     t2 ->* t2' →
     P t1 t2 ->* P t1 t2'.
Proof.
   Admitted.
```

□

With these lemmas in hand, the main proof is a straightforward induction.

Theorem: The **step** function is normalizing – i.e., for every t there exists some t' such that t steps to t' and t' is a normal form.

Proof sketch: By induction on terms. There are two cases to consider:

- t = C n for some n. Here t doesn't take a step, and we have $t' = t$. We can derive the left-hand side by reflexivity and the right-hand side by observing (a) that values are normal forms (by nf_same_as_value) and (b) that t is a value (by v_const).

- t = P $t1$ $t2$ for some $t1$ and $t2$. By the IH, $t1$ and $t2$ have normal forms $t1'$ and $t2'$. Recall that normal forms are values (by nf_same_as_value); we know that $t1'$ = C $n1$ and $t2'$ = C $n2$, for some $n1$ and $n2$. We can combine the –>* derivations for $t1$ and $t2$ using *multi_congr_1* and *multi_congr_2* to prove that P $t1$ $t2$ reduces in many steps to C $(n1 + n2)$.

  It is clear that our choice of t' = C $(n1 + n2)$ is a value, which is in turn a normal form. □

```
Theorem step_normalizing :
  normalizing step.
```

```
Proof.
  unfold normalizing.
  induction t.
  -
    ∃ (C n).
    split.
    + apply multi_refl.
    +

      rewrite nf_same_as_value. apply v_const.
  -
    destruct IHt1 as [t1' [Hsteps1 Hnormal1]].
    destruct IHt2 as [t2' [Hsteps2 Hnormal2]].
    rewrite nf_same_as_value in Hnormal1.
    rewrite nf_same_as_value in Hnormal2.
    inversion Hnormal1 as [n1 H1].
    inversion Hnormal2 as [n2 H2].
    rewrite ← H1 in Hsteps1.
    rewrite ← H2 in Hsteps2.
    ∃ (C (n1 + n2)).
    split.
    +
      apply multi_trans with (P (C n1) t2).
      × apply multistep_congr_1. apply Hsteps1.
      × apply multi_trans with
        (P (C n1) (C n2)).
        { apply multistep_congr_2. apply v_const. apply Hsteps2. }
        apply multi_R. { apply ST_PlusConstConst. }
    +
      rewrite nf_same_as_value. apply v_const.
Qed.
```

8.3.3 Equivalence of Big-Step and Small-Step

Having defined the operational semantics of our tiny programming language in two different ways (big-step and small-step), it makes sense to ask whether these definitions actually define the same thing! They do, though it takes a little work to show it. The details are left as an exercise.

Exercise: 3 stars, standard (eval__multistep) `Theorem eval__multistep : ∀ t n,`
`t ==> n → t ->* C n.`

The key ideas in the proof can be seen in the following picture:

P t1 t2 –> (by ST_Plus1) P t1' t2 –> (by ST_Plus1) P t1'' t2 –> (by ST_Plus1) ... P (C n1) t2 –> (by ST_Plus2) P (C n1) t2' –> (by ST_Plus2) P (C n1) t2'' –> (by ST_Plus2) ... P (C n1) (C n2) –> (by ST_PlusConstConst) C (n1 + n2)

That is, the multistep reduction of a term of the form P *t1* *t2* proceeds in three phases:

- First, we use ST_Plus1 some number of times to reduce *t1* to a normal form, which must (by nf_same_as_value) be a term of the form C *n1* for some *n1*.
- Next, we use ST_Plus2 some number of times to reduce *t2* to a normal form, which must again be a term of the form C *n2* for some *n2*.
- Finally, we use ST_PlusConstConst one time to reduce P (C *n1*) (C *n2*) to C (*n1* + *n2*).

To formalize this intuition, you'll need to use the congruence lemmas from above (you might want to review them now, so that you'll be able to recognize when they are useful), plus some basic properties of –>*: that it is reflexive, transitive, and includes –>.

`Proof.`
Admitted.
□

Exercise: 3 stars, advanced (eval__multistep_inf) Write a detailed informal version of the proof of eval__multistep.

`Definition` manual_grade_for_eval__multistep_inf : **option** (**nat**×**string**) := None.
□

For the other direction, we need one lemma, which establishes a relation between single-step reduction and big-step evaluation.

Exercise: 3 stars, standard (step__eval) `Lemma` step__eval : ∀ *t* *t'* *n*,
t `->` *t'* →
t' `==>` *n* →
t `==>` *n*.

```
Proof.
  intros t t' n Hs. generalize dependent n.
   Admitted.
```
□

The fact that small-step reduction implies big-step evaluation is now straightforward to prove, once it is stated correctly.

The proof proceeds by induction on the multi-step reduction sequence that is buried in the hypothesis normal_form_of *t* *t'*.

Make sure you understand the statement before you start to work on the proof.

Exercise: 3 stars, standard (multistep__eval) Theorem multistep__eval : ∀ t t',
  normal_form_of t t' → ∃ n, t' = C n ∧ t ==> n.
Proof.
  Admitted.
  □

8.3.4 Additional Exercises

Exercise: 3 stars, standard, optional (interp_tm) Remember that we also defined big-step evaluation of terms as a function evalF. Prove that it is equivalent to the existing semantics. (Hint: we just proved that eval and multistep are equivalent, so logically it doesn't matter which you choose. One will be easier than the other, though!)

```
Theorem evalF_eval : ∀ t n,
  evalF t = n ↔ t ==> n.
Proof.
  Admitted.
  □
```

Exercise: 4 stars, standard (combined_properties) We've considered arithmetic and conditional expressions separately. This exercise explores how the two interact.

```
Module Combined.

Inductive tm : Type :=
  | C : nat → tm
  | P : tm → tm → tm
  | tru : tm
  | fls : tm
  | test : tm → tm → tm → tm.

Inductive value : tm → Prop :=
  | v_const : ∀ n, value (C n)
  | v_tru : value tru
  | v_fls : value fls.

Reserved Notation " t '-->' t' " (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_PlusConstConst : ∀ n1 n2,
      P (C n1) (C n2) --> C (n1 + n2)
  | ST_Plus1 : ∀ t1 t1' t2,
      t1 --> t1' →
      P t1 t2 --> P t1' t2
  | ST_Plus2 : ∀ v1 t2 t2',
      value v1 →
      t2 --> t2' →
```

```
      P v1 t2 -> P v1 t2'
  | ST_IfTrue : ∀ t1 t2,
      test tru t1 t2 -> t1
  | ST_IfFalse : ∀ t1 t2,
      test fls t1 t2 -> t2
  | ST_If : ∀ t1 t1' t2 t3,
      t1 -> t1' →
      test t1 t2 t3 -> test t1' t2 t3

  where " t '->' t' " := (step t t').
```

Earlier, we separately proved for both plus- and if-expressions...

- that the step relation was deterministic, and
- a strong progress lemma, stating that every term is either a value or can take a step.

Formally prove or disprove these two properties for the combined language. (That is, state a theorem saying that the property holds or does not hold, and prove your theorem.)

```
End Combined.
Definition manual_grade_for_combined_properties : option (nat×string) := None.
```

□

8.4 Small-Step Imp

Now for a more serious example: a small-step version of the Imp operational semantics.

The small-step reduction relations for arithmetic and boolean expressions are straightforward extensions of the tiny language we've been working up to now. To make them easier to read, we introduce the symbolic notations ->a and ->b for the arithmetic and boolean step relations.

```
Inductive aval : aexp → Prop :=
  | av_num : ∀ n, aval (ANum n).
```

We are not actually going to bother to define boolean values, since they aren't needed in the definition of ->b below (why?), though they might be if our language were a bit larger (why?).

```
Reserved Notation " t '/' st '->a' t' "
                  (at level 40, st at level 39).

Inductive astep : state → aexp → aexp → Prop :=
  | AS_Id : ∀ st i,
      AId i / st ->a ANum (st i)
  | AS_Plus1 : ∀ st a1 a1' a2,
```

```
      a1 / st ->a a1' →
      (APlus a1 a2) / st ->a (APlus a1' a2)
  | AS_Plus2 : ∀ st v1 a2 a2',
      aval v1 →
      a2 / st ->a a2' →
      (APlus v1 a2) / st ->a (APlus v1 a2')
  | AS_Plus : ∀ st n1 n2,
      APlus (ANum n1) (ANum n2) / st ->a ANum (n1 + n2)
  | AS_Minus1 : ∀ st a1 a1' a2,
      a1 / st ->a a1' →
      (AMinus a1 a2) / st ->a (AMinus a1' a2)
  | AS_Minus2 : ∀ st v1 a2 a2',
      aval v1 →
      a2 / st ->a a2' →
      (AMinus v1 a2) / st ->a (AMinus v1 a2')
  | AS_Minus : ∀ st n1 n2,
      (AMinus (ANum n1) (ANum n2)) / st ->a (ANum (minus n1 n2))
  | AS_Mult1 : ∀ st a1 a1' a2,
      a1 / st ->a a1' →
      (AMult a1 a2) / st ->a (AMult a1' a2)
  | AS_Mult2 : ∀ st v1 a2 a2',
      aval v1 →
      a2 / st ->a a2' →
      (AMult v1 a2) / st ->a (AMult v1 a2')
  | AS_Mult : ∀ st n1 n2,
      (AMult (ANum n1) (ANum n2)) / st ->a (ANum (mult n1 n2))

    where " t '/' st '->a' t' " := (astep st t t').

Reserved Notation " t '/' st '->b' t' "
                  (at level 40, st at level 39).

Inductive bstep : state → bexp → bexp → Prop :=
| BS_Eq1 : ∀ st a1 a1' a2,
    a1 / st ->a a1' →
    (BEq a1 a2) / st ->b (BEq a1' a2)
| BS_Eq2 : ∀ st v1 a2 a2',
    aval v1 →
    a2 / st ->a a2' →
    (BEq v1 a2) / st ->b (BEq v1 a2')
| BS_Eq : ∀ st n1 n2,
    (BEq (ANum n1) (ANum n2)) / st ->b
    (if (n1 =? n2) then BTrue else BFalse)
| BS_LtEq1 : ∀ st a1 a1' a2,
```

```
    a1 / st ->a a1' →
    (BLe a1 a2) / st ->b (BLe a1' a2)
| BS_LtEq2 : ∀ st v1 a2 a2',
    aval v1 →
    a2 / st ->a a2' →
    (BLe v1 a2) / st ->b (BLe v1 a2')
| BS_LtEq : ∀ st n1 n2,
    (BLe (ANum n1) (ANum n2)) / st ->b
             (if (n1 <=? n2) then BTrue else BFalse)
| BS_NotStep : ∀ st b1 b1',
    b1 / st ->b b1' →
    (BNot b1) / st ->b (BNot b1')
| BS_NotTrue : ∀ st,
    (BNot BTrue) / st ->b BFalse
| BS_NotFalse : ∀ st,
    (BNot BFalse) / st ->b BTrue
| BS_AndTrueStep : ∀ st b2 b2',
    b2 / st ->b b2' →
    (BAnd BTrue b2) / st ->b (BAnd BTrue b2')
| BS_AndStep : ∀ st b1 b1' b2,
    b1 / st ->b b1' →
    (BAnd b1 b2) / st ->b (BAnd b1' b2)
| BS_AndTrueTrue : ∀ st,
    (BAnd BTrue BTrue) / st ->b BTrue
| BS_AndTrueFalse : ∀ st,
    (BAnd BTrue BFalse) / st ->b BFalse
| BS_AndFalse : ∀ st b2,
    (BAnd BFalse b2) / st ->b BFalse

where " t '/' st '–>b' t' " := (bstep st t t').
```

The semantics of commands is the interesting part. We need two small tricks to make it work:

- We use *SKIP* as a “command value” – i.e., a command that has reached a normal form.
    - An assignment command reduces to *SKIP* (and an updated state).
    - The sequencing command waits until its left-hand subcommand has reduced to *SKIP*, then throws it away so that reduction can continue with the right-hand subcommand.
- We reduce a *WHILE* command by transforming it into a conditional followed by the same *WHILE*.

(There are other ways of achieving the effect of the latter trick, but they all share the feature that the original *WHILE* command needs to be saved somewhere while a single copy of the loop body is being reduced.)

```
Reserved Notation " t '/' st '->' t' '/' st' "
                  (at level 40, st at level 39, t' at level 39).

Open Scope imp_scope.
Inductive cstep : (com × state) → (com × state) → Prop :=
  | CS_AssStep : ∀ st i a a',
      a / st ->a a' →
      (i ::= a) / st -> (i ::= a') / st
  | CS_Ass : ∀ st i n,
      (i ::= (ANum n)) / st -> SKIP / (i !-> n ; st)
  | CS_SeqStep : ∀ st c1 c1' st' c2,
      c1 / st -> c1' / st' →
      (c1 ;; c2) / st -> (c1' ;; c2) / st'
  | CS_SeqFinish : ∀ st c2,
      (SKIP ;; c2) / st -> c2 / st
  | CS_IfStep : ∀ st b b' c1 c2,
      b / st ->b b' →
      TEST b THEN c1 ELSE c2 FI / st
      ->
      (TEST b' THEN c1 ELSE c2 FI) / st
  | CS_IfTrue : ∀ st c1 c2,
      TEST BTrue THEN c1 ELSE c2 FI / st -> c1 / st
  | CS_IfFalse : ∀ st c1 c2,
      TEST BFalse THEN c1 ELSE c2 FI / st -> c2 / st
  | CS_While : ∀ st b c1,
      WHILE b DO c1 END / st
      ->
      (TEST b THEN c1 ;; WHILE b DO c1 END ELSE SKIP FI) / st

  where " t '/' st '->' t' '/' st' " := (cstep (t,st) (t',st')).

Close Scope imp_scope.
```

8.5 Concurrent Imp

Finally, to show the power of this definitional style, let's enrich Imp with a new form of command that runs two subcommands in parallel and terminates when both have terminated. To reflect the unpredictability of scheduling, the actions of the subcommands may be interleaved in any order, but they share the same memory and can communicate by reading and writing the same variables.

```
Module CImp.

Inductive com : Type :=
  | CSkip : com
  | CAss : string → aexp → com
  | CSeq : com → com → com
  | CIf : bexp → com → com → com
  | CWhile : bexp → com → com
  | CPar : com → com → com.

Notation "'SKIP'" :=
  CSkip.
Notation "x '::=' a" :=
  (CAss x a) (at level 60).
Notation "c1 ;; c2" :=
  (CSeq c1 c2) (at level 80, right associativity).
Notation "'WHILE' b 'DO' c 'END'" :=
  (CWhile b c) (at level 80, right associativity).
Notation "'TEST' b 'THEN' c1 'ELSE' c2 'FI'" :=
  (CIf b c1 c2) (at level 80, right associativity).
Notation "'PAR' c1 'WITH' c2 'END'" :=
  (CPar c1 c2) (at level 80, right associativity).

Inductive cstep : (com × state) → (com × state) → Prop :=

  | CS_AssStep : ∀ st i a a',
      a / st ->a a' →
      (i ::= a) / st -> (i ::= a') / st
  | CS_Ass : ∀ st i n,
      (i ::= (ANum n)) / st -> SKIP / (i !-> n ; st)
  | CS_SeqStep : ∀ st c1 c1' st' c2,
      c1 / st -> c1' / st' →
      (c1 ;; c2) / st -> (c1' ;; c2) / st'
  | CS_SeqFinish : ∀ st c2,
      (SKIP ;; c2) / st -> c2 / st
  | CS_IfStep : ∀ st b b' c1 c2,
      b /st ->b b' →
          (TEST b THEN c1 ELSE c2 FI) / st
      -> (TEST b' THEN c1 ELSE c2 FI) / st
  | CS_IfTrue : ∀ st c1 c2,
      (TEST BTrue THEN c1 ELSE c2 FI) / st -> c1 / st
  | CS_IfFalse : ∀ st c1 c2,
      (TEST BFalse THEN c1 ELSE c2 FI) / st -> c2 / st
  | CS_While : ∀ st b c1,
          (WHILE b DO c1 END) / st
```

```
      -> (TEST b THEN (c1;; (WHILE b DO c1 END)) ELSE SKIP FI) / st

  | CS_Par1 : ∀ st c1 c1' c2 st',
      c1 / st -> c1' / st' →
      (PAR c1 WITH c2 END) / st -> (PAR c1' WITH c2 END) / st'
  | CS_Par2 : ∀ st c1 c2 c2' st',
      c2 / st -> c2' / st' →
      (PAR c1 WITH c2 END) / st -> (PAR c1 WITH c2' END) / st'
  | CS_ParDone : ∀ st,
      (PAR SKIP WITH SKIP END) / st -> SKIP / st
  where " t '/' st '–>' t' '/' st' " := (cstep (t,st) (t',st')).
Definition cmultistep := multi cstep.

Notation " t '/' st '–>*' t' '/' st' " :=
   (multi cstep (t,st) (t',st'))
   (at level 40, st at level 39, t' at level 39).
```

Among the (many) interesting properties of this language is the fact that the following program can terminate with the variable X set to any value.

```
Definition par_loop : com :=
  PAR
    Y ::= 1
  WITH
    WHILE Y = 0 DO
      X ::= X + 1
    END
  END.
```

In particular, it can terminate with X set to 0:

```
Example par_loop_example_0:
  ∃ st',
       par_loop / empty_st ->* SKIP / st'
    ∧ st' X = 0.
Proof.
  eapply ex_intro. split.
  unfold par_loop.
  eapply multi_step. apply CS_Par1.
    apply CS_Ass.
  eapply multi_step. apply CS_Par2. apply CS_While.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq1. apply AS_Id.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq. simpl.
  eapply multi_step. apply CS_Par2. apply CS_IfFalse.
```

```
  eapply multi_step. apply CS_ParDone.
  eapply multi_refl.
  reflexivity. Qed.
```

It can also terminate with X set to 2:

```
Example par_loop_example_2:
  ∃ st',
       par_loop / empty_st ->* SKIP / st'
    ∧ st' X = 2.
Proof.
  eapply ex_intro. split.
  eapply multi_step. apply CS_Par2. apply CS_While.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq1. apply AS_Id.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq. simpl.
  eapply multi_step. apply CS_Par2. apply CS_IfTrue.
  eapply multi_step. apply CS_Par2. apply CS_SeqStep.
    apply CS_AssStep. apply AS_Plus1. apply AS_Id.
  eapply multi_step. apply CS_Par2. apply CS_SeqStep.
    apply CS_AssStep. apply AS_Plus.
  eapply multi_step. apply CS_Par2. apply CS_SeqStep.
    apply CS_Ass.
  eapply multi_step. apply CS_Par2. apply CS_SeqFinish.

  eapply multi_step. apply CS_Par2. apply CS_While.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq1. apply AS_Id.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq. simpl.
  eapply multi_step. apply CS_Par2. apply CS_IfTrue.
  eapply multi_step. apply CS_Par2. apply CS_SeqStep.
    apply CS_AssStep. apply AS_Plus1. apply AS_Id.
  eapply multi_step. apply CS_Par2. apply CS_SeqStep.
    apply CS_AssStep. apply AS_Plus.
  eapply multi_step. apply CS_Par2. apply CS_SeqStep.
    apply CS_Ass.

  eapply multi_step. apply CS_Par1. apply CS_Ass.
  eapply multi_step. apply CS_Par2. apply CS_SeqFinish.
  eapply multi_step. apply CS_Par2. apply CS_While.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq1. apply AS_Id.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
```

```
    apply BS_Eq. simpl.
  eapply multi_step. apply CS_Par2. apply CS_IfFalse.
  eapply multi_step. apply CS_ParDone.
  eapply multi_refl.
  reflexivity. Qed.
```

More generally...

Exercise: 3 stars, standard, optional (par_body_n__Sn)

```
Lemma par_body_n__Sn : ∀ n st,
  st X = n ∧ st Y = 0 →
  par_loop / st ->* par_loop / (X !-> S n ; st).
Proof.
  Admitted.
```

□

Exercise: 3 stars, standard, optional (par_body_n)

```
Lemma par_body_n : ∀ n st,
  st X = 0 ∧ st Y = 0 →
  ∃ st',
    par_loop / st ->* par_loop / st' ∧ st' X = n ∧ st' Y = 0.
Proof.
  Admitted.
```

□

... the above loop can exit with X having any value whatsoever.

```
Theorem par_loop_any_X:
  ∀ n, ∃ st',
    par_loop / empty_st ->* SKIP / st'
    ∧ st' X = n.
Proof.
  intros n.
  destruct (par_body_n n empty_st).
    split; unfold t_update; reflexivity.
  rename x into st.
  inversion H as [H' [HX HY]]; clear H.
  ∃ (Y !-> 1 ; st). split.
  eapply multi_trans with (par_loop,st). apply H'.
  eapply multi_step. apply CS_Par1. apply CS_Ass.
  eapply multi_step. apply CS_Par2. apply CS_While.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq1. apply AS_Id. rewrite t_update_eq.
  eapply multi_step. apply CS_Par2. apply CS_IfStep.
    apply BS_Eq. simpl.
```

```
  eapply multi_step. apply CS_Par2. apply CS_IfFalse.
  eapply multi_step. apply CS_ParDone.
  apply multi_refl.
  rewrite t_update_neq. assumption. intro X; inversion X.
Qed.
End CImp.
```

8.6 A Small-Step Stack Machine

Our last example is a small-step semantics for the stack machine example from the Imp chapter of *Logical Foundations*.

```
Definition stack := list nat.
Definition prog := list sinstr.
Inductive stack_step : state → prog × stack → prog × stack → Prop :=
  | SS_Push : ∀ st stk n p',
    stack_step st (SPush n :: p', stk) (p', n :: stk)
  | SS_Load : ∀ st stk i p',
    stack_step st (SLoad i :: p', stk) (p', st i :: stk)
  | SS_Plus : ∀ st stk n m p',
    stack_step st (SPlus :: p', n::m::stk) (p', (m+n)::stk)
  | SS_Minus : ∀ st stk n m p',
    stack_step st (SMinus :: p', n::m::stk) (p', (m-n)::stk)
  | SS_Mult : ∀ st stk n m p',
    stack_step st (SMult :: p', n::m::stk) (p', (m×n)::stk).
Theorem stack_step_deterministic : ∀ st,
  deterministic (stack_step st).
Proof.
  unfold deterministic. intros st x y1 y2 H1 H2.
  induction H1; inversion H2; reflexivity.
Qed.
Definition stack_multistep st := multi (stack_step st).
```

Exercise: 3 stars, advanced (compiler_is_correct) Remember the definition of *compile* for **aexp** given in the Imp chapter of *Logical Foundations*. We want now to prove s_compile correct with respect to the stack machine.

State what it means for the compiler to be correct according to the stack machine small step semantics and then prove it.

```
Definition compiler_is_correct_statement : Prop
  . Admitted.
Theorem compiler_is_correct : compiler_is_correct_statement.
```

```
Proof.
  Admitted.
  □
```

8.7 Aside: A *normalize* Tactic

When experimenting with definitions of programming languages in Coq, we often want to see what a particular concrete term steps to – i.e., we want to find proofs for goals of the form *t* –>* *t'*, where *t* is a completely concrete term and *t'* is unknown. These proofs are quite tedious to do by hand. Consider, for example, reducing an arithmetic expression using the small-step relation **astep**.

```
Example step_example1 :
  (P (C 3) (P (C 3) (C 4)))
  ->* (C 10).
Proof.
  apply multi_step with (P (C 3) (C 7)).
    apply ST_Plus2.
      apply v_const.
      apply ST_PlusConstConst.
  apply multi_step with (C 10).
    apply ST_PlusConstConst.
  apply multi_refl.
Qed.
```

The proof repeatedly applies multi_step until the term reaches a normal form. Fortunately The sub-proofs for the intermediate steps are simple enough that `auto`, with appropriate hints, can solve them.

```
Hint Constructors step value.
Example step_example1' :
  (P (C 3) (P (C 3) (C 4)))
  ->* (C 10).
Proof.
  eapply multi_step. auto. simpl.
  eapply multi_step. auto. simpl.
  apply multi_refl.
Qed.
```

The following custom `Tactic Notation` definition captures this pattern. In addition, before each step, we print out the current goal, so that we can follow how the term is being reduced.

```
Tactic Notation "print_goal" :=
  match goal with ⊢ ?x ⇒ idtac x end.
```

```
Tactic Notation "normalize" :=
  repeat (print_goal; eapply multi_step ;
            [ (eauto 10; fail) | (instantiate; simpl)]);
  apply multi_refl.
```

```
Example step_example1'' :
  (P (C 3) (P (C 3) (C 4)))
  ->* (C 10).
Proof.
  normalize.
Qed.
```

The *normalize* tactic also provides a simple way to calculate the normal form of a term, by starting with a goal with an existentially bound variable.

```
Example step_example1''' : ∃ e',
  (P (C 3) (P (C 3) (C 4)))
  ->* e'.
Proof.
  eapply ex_intro. normalize.
Qed.
```

Exercise: 1 star, standard (normalize_ex)

```
Theorem normalize_ex : ∃ e',
  (P (C 3) (P (C 2) (C 1)))
  ->* e' ∧ value e'.
Proof.
  Admitted.
```

□

Exercise: 1 star, standard, optional (normalize_ex') For comparison, prove it using `apply` instead of `eapply`.

```
Theorem normalize_ex' : ∃ e',
  (P (C 3) (P (C 2) (C 1)))
  ->* e' ∧ value e'.
Proof.
  Admitted.
```

□

Chapter 9

Types: Type Systems

Our next major topic is *type systems* – static program analyses that classify expressions according to the "shapes" of their results. We'll begin with a typed version of the simplest imaginable language, to introduce the basic ideas of types and typing rules and the fundamental theorems about type systems: *type preservation* and *progress*. In chapter Stlc we'll move on to the *simply typed lambda-calculus*, which lives at the core of every modern functional programming language (including Coq!).

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Arith.Arith.
From PLF Require Import Maps.
From PLF Require Import Imp.
From PLF Require Import Smallstep.
```

```
Hint Constructors multi.
```

9.1 Typed Arithmetic Expressions

To motivate the discussion of type systems, let's begin as usual with a tiny toy language. We want it to have the potential for programs to go wrong because of runtime type errors, so we need something a tiny bit more complex than the language of constants and addition that we used in chapter Smallstep: a single kind of data (e.g., numbers) is too simple, but just two kinds (numbers and booleans) gives us enough material to tell an interesting story.

The language definition is completely routine.

9.1.1 Syntax

Here is the syntax, informally:

t ::= tru | fls | test t then t else t | zro | scc t | prd t | iszro t

And here it is formally:

```
Inductive tm : Type :=
```

```
  | tru : tm
  | fls : tm
  | test : tm → tm → tm → tm
  | zro : tm
  | scc : tm → tm
  | prd : tm → tm
  | iszro : tm → tm.
```

Values are tru, fls, and numeric values...

```
Inductive bvalue : tm → Prop :=
  | bv_tru : bvalue tru
  | bv_fls : bvalue fls.

Inductive nvalue : tm → Prop :=
  | nv_zro : nvalue zro
  | nv_scc : ∀ t, nvalue t → nvalue (scc t).

Definition value (t : tm) := bvalue t ∨ nvalue t.

Hint Constructors bvalue nvalue.
Hint Unfold value.
Hint Unfold update.
```

9.1.2 Operational Semantics

Here is the single-step relation, first informally...

(ST_TestTru) test tru then t1 else t2 –> t1

(ST_TestFls) test fls then t1 else t2 –> t2
t1 –> t1'

(ST_Test) test t1 then t2 else t3 –> test t1' then t2 else t3
t1 –> t1'

(ST_Scc) scc t1 –> scc t1'

(ST_PrdZro) prd zro –> zro
numeric value v1

(ST_PrdScc) prd (scc v1) –> v1
t1 –> t1'

(ST_Prd) prd t1 –> prd t1'

(ST_IszroZro) iszro zro –> tru
numeric value v1

(ST_IszroScc) iszro (scc v1) –> fls
t1 –> t1'

(ST_Iszro) iszro t1 –> iszro t1'
... and then formally:

```
Reserved Notation "t1 '->' t2" (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_TestTru : ∀ t1 t2,
      (test tru t1 t2) -> t1
  | ST_TestFls : ∀ t1 t2,
      (test fls t1 t2) -> t2
  | ST_Test : ∀ t1 t1' t2 t3,
      t1 -> t1' →
      (test t1 t2 t3) -> (test t1' t2 t3)
  | ST_Scc : ∀ t1 t1',
      t1 -> t1' →
      (scc t1) -> (scc t1')
  | ST_PrdZro :
      (prd zro) -> zro
  | ST_PrdScc : ∀ t1,
      nvalue t1 →
      (prd (scc t1)) -> t1
  | ST_Prd : ∀ t1 t1',
      t1 -> t1' →
      (prd t1) -> (prd t1')
  | ST_IszroZro :
      (iszro zro) -> tru
  | ST_IszroScc : ∀ t1,
       nvalue t1 →
      (iszro (scc t1)) -> fls
  | ST_Iszro : ∀ t1 t1',
      t1 -> t1' →
      (iszro t1) -> (iszro t1')

where "t1 '->' t2" := (step t1 t2).

Hint Constructors step.
```

Notice that the **step** relation doesn't care about whether the expression being stepped makes global sense – it just checks that the operation in the *next* reduction step is being applied to the right kinds of operands. For example, the term scc tru cannot take a step, but the almost as obviously nonsensical term

scc (test tru then tru else tru)

can take a step (once, before becoming stuck).

9.1.3 Normal Forms and Values

The first interesting thing to notice about this **step** relation is that the strong progress theorem from the Smallstep chapter fails here. That is, there are terms that are normal forms (they can't take a step) but not values (because we have not included them in our definition of possible "results of reduction"). Such terms are *stuck*.

```
Notation step_normal_form := (normal_form step).

Definition stuck (t : tm) : Prop :=
  step_normal_form t ∧ ¬ value t.

Hint Unfold stuck.
```

Exercise: 2 stars, standard (some_term_is_stuck)

```
Example some_term_is_stuck :
  ∃ t, stuck t.
Proof.
  Admitted.
```

□

However, although values and normal forms are *not* the same in this language, the set of values is a subset of the set of normal forms. This is important because it shows we did not accidentally define things so that some value could still take a step.

Exercise: 3 stars, standard (value_is_nf)

```
Lemma value_is_nf : ∀ t,
  value t → step_normal_form t.
Proof.
  Admitted.
```

(Hint: You will reach a point in this proof where you need to use an induction to reason about a term that is known to be a numeric value. This induction can be performed either over the term itself or over the evidence that it is a numeric value. The proof goes through in either case, but you will find that one way is quite a bit shorter than the other. For the sake of the exercise, try to complete the proof both ways.)

□

Exercise: 3 stars, standard, optional (step_deterministic) Use value_is_nf to show that the **step** relation is also deterministic.

```
Theorem step_deterministic:
  deterministic step.
Proof with eauto.
```

Admitted.
□

9.1.4 Typing

The next critical observation is that, although this language has stuck terms, they are always nonsensical, mixing booleans and numbers in a way that we don't even *want* to have a meaning. We can easily exclude such ill-typed terms by defining a *typing relation* that relates terms to the types (either numeric or boolean) of their final results.

```
Inductive ty : Type :=
  | Bool : ty
  | Nat : ty.
```

In informal notation, the typing relation is often written ⊢ *t* \in T and pronounced "*t* has type T." The ⊢ symbol is called a "turnstile." Below, we're going to see richer typing relations where one or more additional "context" arguments are written to the left of the turnstile. For the moment, the context is always empty.

(T_Tru) |- tru \in Bool

(T_Fls) |- fls \in Bool
|- t1 \in Bool |- t2 \in T |- t3 \in T

(T_Test) |- test t1 then t2 else t3 \in T

(T_Zro) |- zro \in Nat
|- t1 \in Nat

(T_Scc) |- scc t1 \in Nat
|- t1 \in Nat

(T_Prd) |- prd t1 \in Nat
|- t1 \in Nat

(T_IsZro) |- iszro t1 \in Bool

```
Reserved Notation "'|-' t '\in' T" (at level 40).

Inductive has_type : tm → ty → Prop :=
  | T_Tru :
       ⊢ tru \in Bool
  | T_Fls :
       ⊢ fls \in Bool
  | T_Test : ∀ t1 t2 t3 T,
```

```
        ⊢ t1 \in Bool →
        ⊢ t2 \in T →
        ⊢ t3 \in T →
        ⊢ test t1 t2 t3 \in T
  | T_Zro :
        ⊢ zro \in Nat
  | T_Scc : ∀ t1,
        ⊢ t1 \in Nat →
        ⊢ scc t1 \in Nat
  | T_Prd : ∀ t1,
        ⊢ t1 \in Nat →
        ⊢ prd t1 \in Nat
  | T_Iszro : ∀ t1,
        ⊢ t1 \in Nat →
        ⊢ iszro t1 \in Bool

where "'|-' t '\in' T" := (has_type t T).

Hint Constructors has_type.

Example has_type_1 :
  ⊢ test fls zro (scc zro) \in Nat.
Proof.
  apply T_Test.
    - apply T_Fls.
    - apply T_Zro.
    - apply T_Scc.
      + apply T_Zro.
Qed.
```

(Since we've included all the constructors of the typing relation in the hint database, the `auto` tactic can actually find this proof automatically.)

It's important to realize that the typing relation is a *conservative* (or *static*) approximation: it does not consider what happens when the term is reduced – in particular, it does not calculate the type of its normal form.

```
Example has_type_not :
  ¬ ( ⊢ test fls zro tru \in Bool ).
Proof.
  intros Contra. solve_by_inverts 2. Qed.
```

Exercise: 1 star, standard, optional (scc_hastype_nat__hastype_nat)

```
Example scc_hastype_nat__hastype_nat : ∀ t,
  ⊢ scc t \in Nat →
  ⊢ t \in Nat.
```

```
Proof.
  Admitted.
```

Canonical forms

The following two lemmas capture the fundamental property that the definitions of boolean and numeric values agree with the typing relation.

```
Lemma bool_canonical : ∀ t,
  ⊢ t \in Bool → value t → bvalue t.
Proof.
  intros t HT [Hb | Hn].
  - assumption.
  - induction Hn; inversion HT; auto.
Qed.

Lemma nat_canonical : ∀ t,
  ⊢ t \in Nat → value t → nvalue t.
Proof.
  intros t HT [Hb | Hn].
  - inversion Hb; subst; inversion HT.
  - assumption.
Qed.
```

9.1.5 Progress

The typing relation enjoys two critical properties. The first is that well-typed normal forms are not stuck – or conversely, if a term is well typed, then either it is a value or it can take at least one step. We call this *progress*.

Exercise: 3 stars, standard (finish_progress) `Theorem progress : ∀ t T,`

```
  ⊢ t \in T →
  value t ∨ ∃ t', t -> t'.
```

Complete the formal proof of the `progress` property. (Make sure you understand the parts we've given of the informal proof in the following exercise before starting – this will save you a lot of time.) `Proof with auto.`

```
  intros t T HT.
  induction HT...
  -
    right. inversion IHHT1; clear IHHT1.
    +
    apply (bool_canonical t1 HT1) in H.
    inversion H; subst; clear H.
```

∃ *t2*...
∃ *t3*...
+
inversion *H* as [*t1' H1*].
∃ (test *t1' t2 t3*)...
Admitted.
□

Exercise: 3 stars, advanced (finish_progress_informal) Complete the corresponding informal proof:

Theorem: If ⊢ *t* \in T, then either *t* is a value or else *t* –> *t'* for some *t'*.

Proof: By induction on a derivation of ⊢ *t* \in T.

- If the last rule in the derivation is T_Test, then *t* = test *t1* then *t2* else *t3*, with ⊢ *t1* \in Bool, ⊢ *t2* \in T and ⊢ *t3* \in T. By the IH, either *t1* is a value or else *t1* can step to some *t1'*.
  - If *t1* is a value, then by the canonical forms lemmas and the fact that ⊢ *t1* \in Bool we have that *t1* is a **bvalue** – i.e., it is either tru or fls. If *t1* = tru, then *t* steps to *t2* by ST_TestTru, while if *t1* = fls, then *t* steps to *t3* by ST_TestFls. Either way, *t* can step, which is what we wanted to show.
  - If *t1* itself can take a step, then, by ST_Test, so can *t*.
-

Definition manual_grade_for_finish_progress_informal : **option** (**nat**×**string**) := None.
□

This theorem is more interesting than the strong progress theorem that we saw in the Smallstep chapter, where *all* normal forms were values. Here a term can be stuck, but only if it is ill typed.

9.1.6 Type Preservation

The second critical property of typing is that, when a well-typed term takes a step, the result is also a well-typed term.

Exercise: 2 stars, standard (finish_preservation) Theorem preservation : ∀ *t t' T*,
⊢ *t* \in *T* →
t -> *t'* →
⊢ *t'* \in *T*.

Complete the formal proof of the preservation property. (Again, make sure you understand the informal proof fragment in the following exercise first.)

```
Proof with auto.
  intros t t' T HT HE.
  generalize dependent t'.
  induction HT;

         intros t' HE;

         try solve_by_invert.
    - inversion HE; subst; clear HE.
      + assumption.
      + assumption.
      + apply T_Test; try assumption.
        apply IHHT1; assumption.
   Admitted.
```
☐

Exercise: 3 stars, advanced (finish_preservation_informal) Complete the following informal proof:

Theorem: If ⊢ *t* \in T and *t* –> *t'*, then ⊢ *t'* \in T.

Proof: By induction on a derivation of ⊢ *t* \in T.

- If the last rule in the derivation is T_Test, then *t* = test *t1* then *t2* else *t3*, with ⊢ *t1* \in Bool, ⊢ *t2* \in T and ⊢ *t3* \in T.

  Inspecting the rules for the small-step reduction relation and remembering that *t* has the form test ..., we see that the only ones that could have been used to prove *t* –> *t'* are ST_TestTru, ST_TestFls, or ST_Test.

  - If the last rule was ST_TestTru, then *t'* = *t2*. But we know that ⊢ *t2* \in T, so we are done.
  - If the last rule was ST_TestFls, then *t'* = *t3*. But we know that ⊢ *t3* \in T, so we are done.
  - If the last rule was ST_Test, then *t'* = test *t1'* then *t2* else *t3*, where *t1* –> *t1'*. We know ⊢ *t1* \in Bool so, by the IH, ⊢ *t1'* \in Bool. The T_Test rule then gives us ⊢ test *t1'* then *t2* else *t3* \in T, as required.

-

```
Definition manual_grade_for_finish_preservation_informal : option (nat×string) := None.
```
☐

Exercise: 3 stars, standard (preservation_alternate_proof) Now prove the same property again by induction on the *evaluation* derivation instead of on the typing derivation. Begin by carefully reading and thinking about the first few lines of the above proofs to make sure you understand what each one is doing. The set-up for this proof is similar, but not exactly the same.

```
Theorem preservation' : ∀ t t' T,
  ⊢ t \in T →
  t -> t' →
  ⊢ t' \in T.
Proof with eauto.
   Admitted.
```

□

The preservation theorem is often called *subject reduction*, because it tells us what happens when the "subject" of the typing relation is reduced. This terminology comes from thinking of typing statements as sentences, where the term is the subject and the type is the predicate.

9.1.7 Type Soundness

Putting progress and preservation together, we see that a well-typed term can never reach a stuck state.

```
Definition multistep := (multi step).
Notation "t1 '–>*' t2" := (multistep t1 t2) (at level 40).

Corollary soundness : ∀ t t' T,
  ⊢ t \in T →
  t ->* t' →
  ~(stuck t').
Proof.
  intros t t' T HT P. induction P; intros [R S].
  destruct (progress x T HT); auto.
  apply IHP. apply (preservation x y T HT H).
  unfold stuck. split; auto. Qed.
```

9.1.8 Additional Exercises

Exercise: 2 stars, standard, recommended (subject_expansion) Having seen the subject reduction property, one might wonder whether the opposity property – subject *expansion* – also holds. That is, is it always the case that, if t –> t' and ⊢ t' \in T, then ⊢ t \in T? If so, prove it. If not, give a counter-example. (You do not need to prove your counter-example in Coq, but feel free to do so.)

```
Definition manual_grade_for_subject_expansion : option (nat×string) := None.
```

□

Exercise: 2 stars, standard (variation1) Suppose that we add this new rule to the typing relation:

| T_SccBool : forall t, |- t \in Bool -> |- scc t \in Bool

Which of the following properties remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

`Definition` manual_grade_for_variation1 : **option** (**nat**×**string**) := None.

□

Exercise: 2 stars, standard (variation2) Suppose, instead, that we add this new rule to the **step** relation:

| ST_Funny1 : forall t2 t3, (test tru t2 t3) –> t3

Which of the above properties become false in the presence of this rule? For each one that does, give a counter-example. `Definition` manual_grade_for_variation2 : **option** (**nat**×**string**) := None.

□

Exercise: 2 stars, standard, optional (variation3) Suppose instead that we add this rule:

| ST_Funny2 : forall t1 t2 t2' t3, t2 –> t2' -> (test t1 t2 t3) –> (test t1 t2' t3)

Which of the above properties become false in the presence of this rule? For each one that does, give a counter-example.

□

Exercise: 2 stars, standard, optional (variation4) Suppose instead that we add this rule:

| ST_Funny3 : (prd fls) –> (prd (prd fls))

Which of the above properties become false in the presence of this rule? For each one that does, give a counter-example.

□

Exercise: 2 stars, standard, optional (variation5) Suppose instead that we add this rule:

| T_Funny4 : |- zro \in Bool

Which of the above properties become false in the presence of this rule? For each one that does, give a counter-example.

□

Exercise: 2 stars, standard, optional (variation6) Suppose instead that we add this rule:

| T_Funny5 : |- prd zro \in Bool

Which of the above properties become false in the presence of this rule? For each one that does, give a counter-example.

□

Exercise: 3 stars, standard, optional (more_variations) Make up some exercises of your own along the same lines as the ones above. Try to find ways of selectively breaking properties – i.e., ways of changing the definitions that break just one of the properties and leave the others alone.

Exercise: 1 star, standard (remove_prdzro) The reduction rule ST_PrdZro is a bit counter-intuitive: we might feel that it makes more sense for the predecessor of zro to be undefined, rather than being defined to be zro. Can we achieve this simply by removing the rule from the definition of **step**? Would doing so create any problems elsewhere?

```
Definition manual_grade_for_remove_predzro : option (nat×string) := None.
```

□

Exercise: 4 stars, advanced (prog_pres_bigstep) Suppose our evaluation relation is defined in the big-step style. State appropriate analogs of the progress and preservation properties. (You do not need to prove them.)

Can you see any limitations of either of your properties? Do they allow for nonterminating commands? Why might we prefer the small-step semantics for stating preservation and progress?

```
Definition manual_grade_for_prog_pres_bigstep : option (nat×string) := None.
```

□

Chapter 10

Stlc: The Simply Typed Lambda-Calculus

The simply typed lambda-calculus (STLC) is a tiny core calculus embodying the key concept of *functional abstraction*, which shows up in pretty much every real-world programming language in some form (functions, procedures, methods, etc.).

We will follow exactly the same pattern as in the previous chapter when formalizing this calculus (syntax, small-step semantics, typing rules) and its main properties (progress and preservation). The new technical challenges arise from the mechanisms of *variable binding* and *substitution*. It will take some work to deal with these.

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Strings.String.
From PLF Require Import Maps.
From PLF Require Import Smallstep.
```

10.1 Overview

The STLC is built on some collection of *base types*: booleans, numbers, strings, etc. The exact choice of base types doesn't matter much – the construction of the language and its theoretical properties work out the same no matter what we choose – so for the sake of brevity let's take just Bool for the moment. At the end of the chapter we'll see how to add more base types, and in later chapters we'll enrich the pure STLC with other useful constructs like pairs, records, subtyping, and mutable state.

Starting from boolean constants and conditionals, we add three things:

- variables
- function abstractions
- application

This gives us the following collection of abstract syntax constructors (written out first in informal BNF notation – we'll formalize it below).

t ::= x variable | \x:T1.t2 abstraction | t1 t2 application | tru constant true | fls constant false | test t1 then t2 else t3 conditional

The \ symbol in a function abstraction \x:*T.t* is generally written as a Greek letter "lambda" (hence the name of the calculus). The variable x is called the *parameter* to the function; the term *t* is its *body*. The annotation :*T1* specifies the type of arguments that the function can be applied to.

Some examples:

- \x:Bool. x

  The identity function for booleans.

- (\x:Bool. x) tru

  The identity function for booleans, applied to the boolean tru.

- \x:Bool. test x then fls else tru

  The boolean "not" function.

- \x:Bool. tru

  The constant function that takes every (boolean) argument to tru.

- \x:Bool. \y:Bool. x

  A two-argument function that takes two booleans and returns the first one. (As in Coq, a two-argument function is really a one-argument function whose body is also a one-argument function.)

- (\x:Bool. \y:Bool. x) fls tru

  A two-argument function that takes two booleans and returns the first one, applied to the booleans fls and tru.

  As in Coq, application associates to the left – i.e., this expression is parsed as ((\x:Bool. \y:Bool. x) fls) tru.

- \f:Bool→Bool. f (f tru)

  A higher-order function that takes a *function* f (from booleans to booleans) as an argument, applies f to tru, and applies f again to the result.

- (\f:Bool→Bool. f (f tru)) (\x:Bool. fls)

  The same higher-order function, applied to the constantly fls function.

As the last several examples show, the STLC is a language of *higher-order* functions: we can write down functions that take other functions as arguments and/or return other functions as results.

The STLC doesn't provide any primitive syntax for defining *named* functions – all functions are "anonymous." We'll see in chapter MoreStlc that it is easy to add named functions to what we've got – indeed, the fundamental naming and binding mechanisms are exactly the same.

The *types* of the STLC include Bool, which classifies the boolean constants tru and fls as well as more complex computations that yield booleans, plus *arrow types* that classify functions.

T ::= Bool | T -> T

For example:

- \x:Bool. fls has type Bool→Bool
- \x:Bool. x has type Bool→Bool
- (\x:Bool. x) tru has type Bool
- \x:Bool. \y:Bool. x has type Bool→Bool→Bool (i.e., Bool → (Bool→Bool))
- (\x:Bool. \y:Bool. x) fls has type Bool→Bool
- (\x:Bool. \y:Bool. x) fls tru has type Bool

10.2 Syntax

We next formalize the syntax of the STLC.

```
Module STLC.
```

10.2.1 Types

```
Inductive ty : Type :=
  | Bool : ty
  | Arrow : ty → ty → ty.
```

10.2.2 Terms

```
Inductive tm : Type :=
  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm
```

```
| tru : tm
| fls : tm
| test : tm → tm → tm → tm.
```

Note that an abstraction \x:$T.t$ (formally, abs x T t) is always annotated with the type T of its parameter, in contrast to Coq (and other functional languages like ML, Haskell, etc.), which use type inference to fill in missing annotations. We're not considering type inference here.

Some examples...

```
Open Scope string_scope.

Definition x := "x".
Definition y := "y".
Definition z := "z".

Hint Unfold x.
Hint Unfold y.
Hint Unfold z.
```

idB = \x:Bool. x

```
Notation idB :=
  (abs x Bool (var x)).
```

idBB = \x:Bool→Bool. x

```
Notation idBB :=
  (abs x (Arrow Bool Bool) (var x)).
```

idBBBB = \x:(Bool→Bool) → (Bool→Bool). x

```
Notation idBBBB :=
  (abs x (Arrow (Arrow Bool Bool)
                      (Arrow Bool Bool))
    (var x)).
```

k = \x:Bool. \y:Bool. x

```
Notation k := (abs x Bool (abs y Bool (var x))).
```

notB = \x:Bool. test x then fls else tru

```
Notation notB := (abs x Bool (test (var x) fls tru)).
```

(We write these as `Notation`s rather than `Definition`s to make things easier for `auto`.)

10.3 Operational Semantics

To define the small-step semantics of STLC terms, we begin, as always, by defining the set of values. Next, we define the critical notions of *free variables* and *substitution*, which are used in the reduction rule for application expressions. And finally we give the small-step relation itself.

10.3.1 Values

To define the values of the STLC, we have a few cases to consider.

First, for the boolean part of the language, the situation is clear: tru and fls are the only values. A test expression is never a value.

Second, an application is not a value: it represents a function being invoked on some argument, which clearly still has work left to do.

Third, for abstractions, we have a choice:

- We can say that \x:T. t is a value only when t is a value – i.e., only if the function's body has been reduced (as much as it can be without knowing what argument it is going to be applied to).
- Or we can say that \x:T. t is always a value, no matter whether t is one or not – in other words, we can say that reduction stops at abstractions.

Our usual way of evaluating expressions in Coq makes the first choice – for example,

Compute (fun x:bool => 3 + 4)

yields:

fun x:bool => 7

Most real-world functional programming languages make the second choice – reduction of a function's body only begins when the function is actually applied to an argument. We also make the second choice here.

```
Inductive value : tm → Prop :=
  | v_abs : ∀ x T t,
      value (abs x T t)
  | v_tru :
      value tru
  | v_fls :
      value fls.
Hint Constructors value.
```

Finally, we must consider what constitutes a *complete* program.

Intuitively, a "complete program" must not refer to any undefined variables. We'll see shortly how to define the *free* variables in a STLC term. A complete program is *closed* – that is, it contains no free variables.

(Conversely, a term with free variables is often called an *open term*.)

Having made the choice not to reduce under abstractions, we don't need to worry about whether variables are values, since we'll always be reducing programs "from the outside in," and that means the **step** relation will always be working with closed terms.

10.3.2 Substitution

Now we come to the heart of the STLC: the operation of substituting one term for a variable in another term. This operation is used below to define the operational semantics of function application, where we will need to substitute the argument term for the function parameter in the function's body. For example, we reduce

(\x:Bool. test x then tru else x) fls

to

test fls then tru else fls

by substituting fls for the parameter x in the body of the function.

In general, we need to be able to substitute some given term s for occurrences of some variable x in another term *t*. In informal discussions, this is usually written [x:=s]*t* and pronounced "substitute s for x in *t*."

Here are some examples:

x:=tru (test x then x else fls) yields test tru then tru else fls

x:=tru x yields tru

x:=tru (test x then x else y) yields test tru then tru else y

x:=tru y yields y

x:=tru fls yields fls (vacuous substitution)

x:=tru (\y:Bool. test y then x else fls) yields \y:Bool. test y then tru else fls

x:=tru (\y:Bool. x) yields \y:Bool. tru

x:=tru (\y:Bool. y) yields \y:Bool. y

x:=tru (\x:Bool. x) yields \x:Bool. x

The last example is very important: substituting x with tru in \x:Bool. x does *not* yield \x:Bool. tru! The reason for this is that the x in the body of \x:Bool. x is *bound* by the abstraction: it is a new, local name that just happens to be spelled the same as some global name x.

Here is the definition, informally...

x:=sx = s x:=sy = y if x <> y x:=s(\x:T11. t12) = \x:T11. t12 x:=s(\y:T11. t12) = \y:T11. x:=st12 if x <> y x:=s(t1 t2) = (x:=st1) (x:=st2) x:=stru = tru x:=sfls = fls x:=s(test t1 then t2 else t3) = test x:=st1 then x:=st2 else x:=st3

... and formally:

```
Reserved Notation "'[' x ':=' s ']' t" (at level 20).

Fixpoint subst (x : string) (s : tm) (t : tm) : tm :=
  match t with
  | var x' ⇒
      if eqb_string x x' then s else t
  | abs x' T t1 ⇒
      abs x' T (if eqb_string x x' then t1 else ([x:=s] t1))
```

```
  | app t1 t2 ⇒
      app ([x:=s] t1) ([x:=s] t2)
  | tru ⇒
      tru
  | fls ⇒
      fls
  | test t1 t2 t3 ⇒
      test ([x:=s] t1) ([x:=s] t2) ([x:=s] t3)
  end

where "'[' x ':=' s ']' t" := (subst x s t).
```

Technical note: Substitution becomes trickier to define if we consider the case where s, the term being substituted for a variable in some other term, may itself contain free variables. Since we are only interested here in defining the **step** relation on *closed* terms (i.e., terms like \x:Bool. x that include binders for all of the variables they mention), we can sidestep this extra complexity, but it must be dealt with when formalizing richer languages.

For example, using the definition of substitution above to substitute the *open* term s = \x:Bool. r, where r is a *free* reference to some global resource, for the variable z in the term t = \r:Bool. z, where r is a bound variable, we would get \r:Bool. \x:Bool. r, where the free reference to r in s has been "captured" by the binder at the beginning of t.

Why would this be bad? Because it violates the principle that the names of bound variables do not matter. For example, if we rename the bound variable in t, e.g., let t' = \\$w$:Bool. z, then [x:=s]$t'$ is \\$w$:Bool. \x:Bool. r, which does not behave the same as [x:=s]t = \r:Bool. \x:Bool. r. That is, renaming a bound variable changes how t behaves under substitution.

See, for example, *Aydemir* 2008 (in Bib.v) for further discussion of this issue.

Exercise: 3 stars, standard (substi_correct) The definition that we gave above uses Coq's `Fixpoint` facility to define substitution as a *function*. Suppose, instead, we wanted to define substitution as an inductive *relation* **substi**. We've begun the definition by providing the `Inductive` header and one of the constructors; your job is to fill in the rest of the constructors and prove that the relation you've defined coincides with the function given above.

```
Inductive substi (s : tm) (x : string) : tm → tm → Prop :=
  | s_var1 :
      substi s x (var x) s

.

Hint Constructors substi.

Theorem substi_correct : ∀ s x t t',
  [x:=s] t = t' ↔ substi s x t t'.
```

```
Proof.
  Admitted.
  □
```

10.3.3 Reduction

The small-step reduction relation for STLC now follows the same pattern as the ones we have seen before. Intuitively, to reduce a function application, we first reduce its left-hand side (the function) until it becomes an abstraction; then we reduce its right-hand side (the argument) until it is also a value; and finally we substitute the argument for the bound variable in the body of the abstraction. This last rule, written informally as

(\x:T.t12) v2 –> x:=*v2*t12

is traditionally called *beta-reduction*.

value v2

(ST_AppAbs) (\x:T.t12) v2 –> x:=*v2*t12

t1 –> t1'

(ST_App1) t1 t2 –> t1' t2

value v1 t2 –> t2'

(ST_App2) v1 t2 –> v1 t2'

... plus the usual rules for conditionals:

(ST_TestTru) (test tru then t1 else t2) –> t1

(ST_TestFls) (test fls then t1 else t2) –> t2

t1 –> t1'

(ST_Test) (test t1 then t2 else t3) –> (test t1' then t2 else t3)

Formally:

```
Reserved Notation "t1 '–>' t2" (at level 40).

Inductive step : tm → tm → Prop :=
  | ST_AppAbs : ∀ x T t12 v2,
         value v2 →
         (app (abs x T t12) v2) -> [x:=v2]t12
  | ST_App1 : ∀ t1 t1' t2,
         t1 -> t1' →
         app t1 t2 -> app t1' t2
  | ST_App2 : ∀ v1 t2 t2',
         value v1 →
         t2 -> t2' →
```

```
        app v1 t2 -> app v1 t2'
  | ST_TestTru : ∀ t1 t2,
      (test tru t1 t2) -> t1
  | ST_TestFls : ∀ t1 t2,
      (test fls t1 t2) -> t2
  | ST_Test : ∀ t1 t1' t2 t3,
      t1 -> t1' →
      (test t1 t2 t3) -> (test t1' t2 t3)

where "t1 '->' t2" := (step t1 t2).

Hint Constructors step.

Notation multistep := (multi step).
Notation "t1 '->*' t2" := (multistep t1 t2) (at level 40).
```

10.3.4 Examples

Example:
(\x:Bool->Bool. x) (\x:Bool. x) –>* \x:Bool. x
i.e.,
idBB idB –>* idB

```
Lemma step_example1 :
  (app idBB idB) ->* idB.
Proof.
  eapply multi_step.
    apply ST_AppAbs.
    apply v_abs.
  simpl.
  apply multi_refl. Qed.
```

Example:
(\x:Bool->Bool. x) ((\x:Bool->Bool. x) (\x:Bool. x)) >* \x:Bool. x
i.e.,
(idBB (idBB idB)) –>* idB.

```
Lemma step_example2 :
  (app idBB (app idBB idB)) ->* idB.
Proof.
  eapply multi_step.
    apply ST_App2. auto.
    apply ST_AppAbs. auto.
  eapply multi_step.
    apply ST_AppAbs. simpl. auto.
  simpl. apply multi_refl. Qed.
```

Example:
(\x:Bool->Bool. x) (\x:Bool. test x then fls else tru) tru >* fls
i.e.,
(idBB notB) tru –>* fls.

```
Lemma step_example3 :
  app (app idBB notB) tru ->* fls.
Proof.
  eapply multi_step.
    apply ST_App1. apply ST_AppAbs. auto. simpl.
  eapply multi_step.
    apply ST_AppAbs. auto. simpl.
  eapply multi_step.
    apply ST_TestTru. apply multi_refl. Qed.
```

Example:
(\x:Bool -> Bool. x) ((\x:Bool. test x then fls else tru) tru) >* fls
i.e.,
idBB (notB tru) –>* fls.
(Note that this term doesn't actually typecheck; even so, we can ask how it reduces.)

```
Lemma step_example4 :
  app idBB (app notB tru) ->* fls.
Proof.
  eapply multi_step.
    apply ST_App2. auto.
    apply ST_AppAbs. auto. simpl.
  eapply multi_step.
    apply ST_App2. auto.
    apply ST_TestTru.
  eapply multi_step.
    apply ST_AppAbs. auto. simpl.
  apply multi_refl. Qed.
```

We can use the *normalize* tactic defined in the Smallstep chapter to simplify these proofs.

```
Lemma step_example1' :
  app idBB idB ->* idB.
Proof. normalize. Qed.

Lemma step_example2' :
  app idBB (app idBB idB) ->* idB.
Proof. normalize. Qed.

Lemma step_example3' :
  app (app idBB notB) tru ->* fls.
Proof. normalize. Qed.
```

```
Lemma step_example4' :
  app idBB (app notB tru) ->* fls.
Proof. normalize. Qed.
```

Exercise: 2 stars, standard (step_example5) Try to do this one both with and without *normalize*.

```
Lemma step_example5 :
       app (app idBBBB idBB) idB
  ->* idB.
Proof.
   Admitted.
```

```
Lemma step_example5_with_normalize :
       app (app idBBBB idBB) idB
  ->* idB.
Proof.
   Admitted.
```

□

10.4 Typing

Next we consider the typing relation of the STLC.

10.4.1 Contexts

Question: What is the type of the term "x y"?

Answer: It depends on the types of x and y!

I.e., in order to assign a type to a term, we need to know what assumptions we should make about the types of its free variables.

This leads us to a three-place *typing judgment*, informally written *Gamma* ⊢ *t* \in T, where *Gamma* is a "typing context" – a mapping from variables to their types.

Following the usual notation for partial maps, we write (X |-> *T11*, *Gamma*) for "update the partial function *Gamma* so that it maps x to T."

```
Definition context := partial_map ty.
```

10.4.2 Typing Relation

Gamma x = T

(T_Var) Gamma |- x \in T

(x |-> T11 ; Gamma) |- t12 \in T12

(T_Abs) Gamma |- \x:T11.t12 \in T11->T12
Gamma |- t1 \in T11->T12 Gamma |- t2 \in T11

(T_App) Gamma |- t1 t2 \in T12

(T_Tru) Gamma |- tru \in Bool

(T_Fls) Gamma |- fls \in Bool
Gamma |- t1 \in Bool Gamma |- t2 \in T Gamma |- t3 \in T

(T_Test) Gamma |- test t1 then t2 else t3 \in T

We can read the three-place relation *Gamma* ⊢ *t* \in T as: "under the assumptions in Gamma, the term *t* has the type T."

```
Reserved Notation "Gamma '|-' t '\in' T" (at level 40).

Inductive has_type : context → tm → ty → Prop :=
  | T_Var : ∀ Gamma x T,
      Gamma x = Some T →
      Gamma ⊢ var x \in T
  | T_Abs : ∀ Gamma x T11 T12 t12,
      (x |-> T11 ; Gamma) ⊢ t12 \in T12 →
      Gamma ⊢ abs x T11 t12 \in Arrow T11 T12
  | T_App : ∀ T11 T12 Gamma t1 t2,
      Gamma ⊢ t1 \in Arrow T11 T12 →
      Gamma ⊢ t2 \in T11 →
      Gamma ⊢ app t1 t2 \in T12
  | T_Tru : ∀ Gamma,
       Gamma ⊢ tru \in Bool
  | T_Fls : ∀ Gamma,
       Gamma ⊢ fls \in Bool
  | T_Test : ∀ t1 t2 t3 T Gamma,
       Gamma ⊢ t1 \in Bool →
       Gamma ⊢ t2 \in T →
       Gamma ⊢ t3 \in T →
       Gamma ⊢ test t1 t2 t3 \in T

where "Gamma '|-' t '\in' T" := (has_type Gamma t T).

Hint Constructors has_type.
```

10.4.3 Examples

```
Example typing_example_1 :
```

```
  empty ⊢ abs x Bool (var x) \in Arrow Bool Bool.
Proof.
  apply T_Abs. apply T_Var. reflexivity. Qed.
```

Note that, since we added the **has_type** constructors to the hints database, auto can actually solve this one immediately.

```
Example typing_example_1' :
  empty ⊢ abs x Bool (var x) \in Arrow Bool Bool.
Proof. auto. Qed.
```

More examples:

empty |- \x:A. \y:A->A. y (y x) \in A -> (A->A) -> A.

```
Example typing_example_2 :
  empty ⊢
    (abs x Bool
       (abs y (Arrow Bool Bool)
          (app (var y) (app (var y) (var x))))) \in
    (Arrow Bool (Arrow (Arrow Bool Bool) Bool)).
Proof with auto using update_eq.
  apply T_Abs.
  apply T_Abs.
  eapply T_App. apply T_Var...
  eapply T_App. apply T_Var...
  apply T_Var...
Qed.
```

Exercise: 2 stars, standard, optional (typing_example_2_full) Prove the same result without using auto, eauto, or eapply (or ...).

```
Example typing_example_2_full :
  empty ⊢
    (abs x Bool
       (abs y (Arrow Bool Bool)
          (app (var y) (app (var y) (var x))))) \in
    (Arrow Bool (Arrow (Arrow Bool Bool) Bool)).
Proof.
```

Admitted.

□

Exercise: 2 stars, standard (typing_example_3) Formally prove the following typing derivation holds:

empty |- \x:Bool->B. \y:Bool->Bool. \z:Bool. y (x z) \in T.

```
Example typing_example_3 :
```

```
  ∃ T,
     empty ⊢
       (abs x (Arrow Bool Bool)
          (abs y (Arrow Bool Bool)
             (abs z Bool
                (app (var y) (app (var x) (var z)))))) \in
       T.
Proof with auto.
```

Admitted.

☐

We can also show that some terms are *not* typable. For example, let's check that there is no typing derivation assigning a type to the term \x:Bool. \y:Bool, x y – i.e.,

~ exists T, empty |- \x:Bool. \y:Bool, x y \in T.

```
Example typing_nonexample_1 :
  ¬ ∃ T,
       empty ⊢
         (abs x Bool
             (abs y Bool
                (app (var x) (var y)))) \in
         T.
Proof.
  intros Hc. inversion Hc.
  inversion H. subst. clear H.
  inversion H5. subst. clear H5.
  inversion H4. subst. clear H4.
  inversion H2. subst. clear H2.
  inversion H5. subst. clear H5.
  inversion H1. Qed.
```

Exercise: 3 stars, standard, optional (typing_nonexample_3) Another nonexample:

~ (exists S T, empty |- \x:S. x x \in T).

```
Example typing_nonexample_3 :
  ¬ (∃ S T,
         empty ⊢
           (abs x S
              (app (var x) (var x))) \in
           T).
Proof.
```

Admitted.

☐

End STLC.

Chapter 11

StlcProp: Properties of STLC

```
Set Warnings "-notation-overridden,-parsing".
From PLF Require Import Maps.
From PLF Require Import Types.
From PLF Require Import Stlc.
From PLF Require Import Smallstep.
Module STLCProp.
Import STLC.
```

In this chapter, we develop the fundamental theory of the Simply Typed Lambda Calculus – in particular, the type safety theorem.

11.1 Canonical Forms

As we saw for the simple calculus in the Types chapter, the first step in establishing basic properties of reduction and types is to identify the possible *canonical forms* (i.e., well-typed closed values) belonging to each type. For Bool, these are the boolean values tru and fls; for arrow types, they are lambda-abstractions.

```
Lemma canonical_forms_bool : ∀ t,
  empty ⊢ t \in Bool →
  value t →
  (t = tru) ∨ (t = fls).
Proof.
  intros t HT HVal.
  inversion HVal; intros; subst; try inversion HT; auto.
Qed.
```

```
Lemma canonical_forms_fun : ∀ t T1 T2,
  empty ⊢ t \in (Arrow T1 T2) →
  value t →
  ∃ x u, t = abs x T1 u.
```

```
Proof.
  intros t T1 T2 HT HVal.
  inversion HVal; intros; subst; try inversion HT; subst; auto.
  ∃ x0, t0. auto.
Qed.
```

11.2 Progress

The *progress* theorem tells us that closed, well-typed terms are not stuck: either a well-typed term is a value, or it can take a reduction step. The proof is a relatively straightforward extension of the progress proof we saw in the Types chapter. We give the proof in English first, then the formal version.

```
Theorem progress : ∀ t T,
  empty ⊢ t \in T →
  value t ∨ ∃ t', t -> t'.
```

Proof: By induction on the derivation of ⊢ t \in T.

- The last rule of the derivation cannot be T_Var, since a variable is never well typed in an empty context.
- The T_Tru, T_Fls, and T_Abs cases are trivial, since in each of these cases we can see by inspecting the rule that t is a value.
- If the last rule of the derivation is T_App, then t has the form t1 t2 for some t1 and t2, where ⊢ t1 \in T2 → T and ⊢ t2 \in T2 for some type T2. The induction hypothesis for the first subderivation says that either t1 is a value or else it can take a reduction step.
    - If t1 is a value, then consider t2, which by the induction hypothesis for the second subderivation must also either be a value or take a step.
        - Suppose t2 is a value. Since t1 is a value with an arrow type, it must be a lambda abstraction; hence t1 t2 can take a step by ST_AppAbs.
        - Otherwise, t2 can take a step, and hence so can t1 t2 by ST_App2.
    - If t1 can take a step, then so can t1 t2 by ST_App1.
- If the last rule of the derivation is T_Test, then t = test t1 then t2 else t3, where t1 has type Bool. The first IH says that t1 either is a value or takes a step.
    - If t1 is a value, then since it has type Bool it must be either tru or fls. If it is tru, then t steps to t2; otherwise it steps to t3.
    - Otherwise, t1 takes a step, and therefore so does t (by ST_Test).

```
Proof with eauto.
  intros t T Ht.
  remember (@empty ty) as Gamma.
  induction Ht; subst Gamma...
  -

    inversion H.

  -

    right. destruct IHHt1...
    +
      destruct IHHt2...
      ×
        assert (∃ x0 t0, t1 = abs x0 T11 t0).
        eapply canonical_forms_fun; eauto.
        destruct H1 as [x0 [t0 Heq]]. subst.
        ∃ ([x0:=t2]t0)...

      ×
        inversion H0 as [t2' Hstp]. ∃ (app t1 t2')...

    +
      inversion H as [t1' Hstp]. ∃ (app t1' t2)...

  -
    right. destruct IHHt1...

    +
      destruct (canonical_forms_bool t1); subst; eauto.

    +
      inversion H as [t1' Hstp]. ∃ (test t1' t2 t3)...
Qed.
```

Exercise: 3 stars, advanced (progress_from_term_ind) Show that progress can also be proved by induction on terms instead of induction on typing derivations.

```
Theorem progress' : ∀ t T,
     empty ⊢ t \in T →
     value t ∨ ∃ t', t -> t'.
Proof.
  intros t.
  induction t; intros T Ht; auto.
   Admitted.
   □
```

11.3 Preservation

The other half of the type soundness property is the preservation of types during reduction. For this part, we'll need to develop some technical machinery for reasoning about variables and substitution. Working from top to bottom (from the high-level property we are actually interested in to the lowest-level technical lemmas that are needed by various cases of the more interesting proofs), the story goes like this:

- The *preservation theorem* is proved by induction on a typing derivation, pretty much as we did in the `Types` chapter. The one case that is significantly different is the one for the ST_AppAbs rule, whose definition uses the substitution operation. To see that this step preserves typing, we need to know that the substitution itself does. So we prove a...

- *substitution lemma*, stating that substituting a (closed) term s for a variable x in a term *t* preserves the type of *t*. The proof goes by induction on the form of *t* and requires looking at all the different cases in the definition of substitition. This time, the tricky cases are the ones for variables and for function abstractions. In both, we discover that we need to take a term s that has been shown to be well-typed in some context *Gamma* and consider the same term s in a slightly different context *Gamma'*. For this we prove a...

- *context invariance* lemma, showing that typing is preserved under "inessential changes" to the context *Gamma* – in particular, changes that do not affect any of the free variables of the term. And finally, for this, we need a careful definition of...

- the *free variables* in a term – i.e., variables that are used in the term in positions that are *not* in the scope of an enclosing function abstraction binding a variable of the same name.

To make Coq happy, of course, we need to formalize the story in the opposite order...

11.3.1 Free Occurrences

A variable x *appears free in* a term *t* if *t* contains some occurrence of x that is not under an abstraction labeled x. For example:

- y appears free, but x does not, in \x:T→*U*. x y

- both x and y appear free in (\x:T→*U*. x y) x

- no variables appear free in \x:T→*U*. \y:T. x y

Formally:

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
      appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
      appears_free_in x t1 →
      appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
      appears_free_in x t2 →
      appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
      y ≠ x →
      appears_free_in x t12 →
      appears_free_in x (abs y T11 t12)
  | afi_test1 : ∀ x t1 t2 t3,
      appears_free_in x t1 →
      appears_free_in x (test t1 t2 t3)
  | afi_test2 : ∀ x t1 t2 t3,
      appears_free_in x t2 →
      appears_free_in x (test t1 t2 t3)
  | afi_test3 : ∀ x t1 t2 t3,
      appears_free_in x t3 →
      appears_free_in x (test t1 t2 t3).

Hint Constructors appears_free_in.
```

The *free variables* of a term are just the variables that appear free in it. A term with no free variables is said to be *closed*.

```
Definition closed (t:tm) :=
  ∀ x, ¬ appears_free_in x t.
```

An *open* term is one that may contain free variables. (I.e., every term is an open term; the closed terms are a subset of the open ones. "Open" precisely means "possibly containing free variables.")

Exercise: 1 star, standard (afi) In the space below, write out the rules of the **appears_free_in** relation in informal inference-rule notation. (Use whatever notational conventions you like – the point of the exercise is just for you to think a bit about the meaning of each rule.) Although this is a rather low-level, technical definition, understanding it is crucial to understanding substitution and its properties, which are really the crux of the lambda-calculus.

```
Definition manual_grade_for_afi : option (nat×string) := None.
```

□

11.3.2 Substitution

To prove that substitution preserves typing, we first need a technical lemma connecting free variables and typing contexts: If a variable x appears free in a term *t*, and if we know *t* is well typed in context *Gamma*, then it must be the case that *Gamma* assigns a type to x.

```
Lemma free_in_context : ∀ x t T Gamma,
  appears_free_in x t →
  Gamma ⊢ t \in T →
  ∃ T', Gamma x = Some T'.
```

Proof: We show, by induction on the proof that x appears free in *t*, that, for all contexts *Gamma*, if *t* is well typed under *Gamma*, then *Gamma* assigns some type to x.

- If the last rule used is afi_var, then $t = \mathsf{x}$, and from the assumption that *t* is well typed under *Gamma* we have immediately that *Gamma* assigns a type to x.

- If the last rule used is afi_app1, then $t = t1\ t2$ and x appears free in *t1*. Since *t* is well typed under *Gamma*, we can see from the typing rules that *t1* must also be, and the IH then tells us that *Gamma* assigns x a type.

- Almost all the other cases are similar: x appears free in a subterm of *t*, and since *t* is well typed under *Gamma*, we know the subterm of *t* in which x appears is well typed under *Gamma* as well, and the IH gives us exactly the conclusion we want.

- The only remaining case is afi_abs. In this case $t = \backslash \mathsf{y}{:}T11.t12$ and x appears free in *t12*, and we also know that x is different from y. The difference from the previous cases is that, whereas *t* is well typed under *Gamma*, its body *t12* is well typed under (y|->*T11*; *Gamma*, so the IH allows us to conclude that x is assigned some type by the extended context (y|->*T11*; *Gamma*. To conclude that *Gamma* assigns a type to x, we appeal to lemma update_neq, noting that x and y are different variables.

```
Proof.
  intros x t T Gamma H H0. generalize dependent Gamma.
  generalize dependent T.
  induction H;
         intros; try solve [inversion H0; eauto].
  -
    inversion H1; subst.
    apply IHappears_free_in in H7.
    rewrite update_neq in H7; assumption.
Qed.
```

From the free_in_context lemma, it immediately follows that any term *t* that is well typed in the empty context is closed (it has no free variables).

Exercise: 2 stars, standard, optional (typable_empty__closed) Corollary typable_empty__closed : ∀ t T,

  empty ⊢ t \in T →
  closed t.

Proof.
  Admitted.
  □

Sometimes, when we have a proof of some typing relation *Gamma* ⊢ t \in T, we will need to replace *Gamma* by a different context *Gamma'*. When is it safe to do this? Intuitively, it must at least be the case that *Gamma'* assigns the same types as *Gamma* to all the variables that appear free in t. In fact, this is the only condition that is needed.

Lemma context_invariance : ∀ *Gamma* *Gamma'* t T,
  Gamma ⊢ t \in T →
  (∀ x, **appears_free_in** x t → *Gamma* x = *Gamma'* x) →
  Gamma' ⊢ t \in T.

Proof: By induction on the derivation of *Gamma* ⊢ t \in T.

- If the last rule in the derivation was T_Var, then t = x and *Gamma* x = T. By assumption, *Gamma'* x = T as well, and hence *Gamma'* ⊢ t \in T by T_Var.

- If the last rule was T_Abs, then t = \y:*T11*. *t12*, with T = *T11* → *T12* and y|->*T11*; *Gamma* ⊢ *t12* \in *T12*. The induction hypothesis is that, for any context *Gamma''*, if y|->*T11*; *Gamma* and *Gamma''* assign the same types to all the free variables in *t12*, then *t12* has type *T12* under *Gamma''*. Let *Gamma'* be a context which agrees with *Gamma* on the free variables in t; we must show *Gamma'* ⊢ \y:*T11*. *t12* \in *T11* → *T12*.

  By T_Abs, it suffices to show that y|->*T11*; *Gamma'* ⊢ *t12* \in *T12*. By the IH (setting *Gamma''* = y|->*T11*;*Gamma'*), it suffices to show that y|->*T11*;*Gamma* and y|->*T11*;*Gamma'* agree on all the variables that appear free in *t12*.

  Any variable occurring free in *t12* must be either y or some other variable. y|->*T11*; *Gamma* and y|->*T11*; *Gamma'* clearly agree on y. Otherwise, note that any variable other than y that occurs free in *t12* also occurs free in t = \y:*T11*. *t12*, and by assumption *Gamma* and *Gamma'* agree on all such variables; hence so do y|->*T11*; *Gamma* and y|->*T11*; *Gamma'*.

- If the last rule was T_App, then t = *t1* *t2*, with *Gamma* ⊢ *t1* \in *T2* → T and *Gamma* ⊢ *t2* \in *T2*. One induction hypothesis states that for all contexts *Gamma'*, if *Gamma'* agrees with *Gamma* on the free variables in *t1*, then *t1* has type *T2* → T under *Gamma'*; there is a similar IH for *t2*. We must show that *t1* *t2* also has type T under *Gamma'*, given the assumption that *Gamma'* agrees with *Gamma* on all the free variables in *t1* *t2*. By T_App, it suffices to show that *t1* and *t2* each have the same type under *Gamma'* as under *Gamma*. But all free variables in *t1* are also free in *t1* *t2*, and similarly for *t2*; hence the desired result follows from the induction hypotheses.

```
Proof with eauto.
  intros.
  generalize dependent Gamma'.
  induction H; intros; auto.
  -
    apply T_Var. rewrite <- H0...
  -
    apply T_Abs.
    apply IHhas_type. intros x1 Hafi.
    unfold update. unfold t_update. destruct (eqb_string x0 x1) eqn: Hx0x1...
    rewrite eqb_string_false_iff in Hx0x1. auto.
  -
    apply T_App with T11...
Qed.
```

Now we come to the conceptual heart of the proof that reduction preserves types – namely, the observation that *substitution* preserves types.

Formally, the so-called *substitution lemma* says this: Suppose we have a term t with a free variable x, and suppose we've assigned a type T to t under the assumption that x has some type U. Also, suppose that we have some other term v and that we've shown that v has type U. Then, since v satisfies the assumption we made about x when typing t, we can substitute v for each of the occurrences of x in t and obtain a new term that still has type T.

Lemma: If x|->U; $Gamma$ ⊢ t \in T and ⊢ v \in U, then $Gamma$ ⊢ [x:=v]t \in T.

```
Lemma substitution_preserves_typing : ∀ Gamma x U t v T,
  (x |-> U ; Gamma) ⊢ t \in T →
  empty ⊢ v \in U →
  Gamma ⊢ [x:=v]t \in T.
```

One technical subtlety in the statement of the lemma is that we assume v has type U in the *empty* context – in other words, we assume v is closed. This assumption considerably simplifies the T_Abs case of the proof (compared to assuming $Gamma$ ⊢ v \in U, which would be the other reasonable assumption at this point) because the context invariance lemma then tells us that v has type U in any context at all – we don't have to worry about free variables in v clashing with the variable being introduced into the context by T_Abs.

The substitution lemma can be viewed as a kind of "commutation property." Intuitively, it says that substitution and typing can be done in either order: we can either assign types to the terms t and v separately (under suitable contexts) and then combine them using substitution, or we can substitute first and then assign a type to [x:=v] t – the result is the same either way.

Proof: We show, by induction on t, that for all T and $Gamma$, if x|->U; $Gamma$ ⊢ t \in T and ⊢ v \in U, then $Gamma$ ⊢ [x:=v]t \in T.

- If t is a variable there are two cases to consider, depending on whether t is x or some other variable.

- If $t = \mathsf{x}$, then from the fact that x|->U; *Gamma* ⊢ x \in T we conclude that U = T. We must show that [x:=v]x = v has type T under *Gamma*, given the assumption that v has type U = T under the empty context. This follows from context invariance: if a closed term has type T in the empty context, it has that type in any context.
- If t is some variable y that is not equal to x, then we need only note that y has the same type under x|->U; *Gamma* as under *Gamma*.

- If t is an abstraction \y:*T11*. *t12*, then the IH tells us, for all *Gamma'* and *T'*, that if x|->U; *Gamma'* ⊢ *t12* \in *T'* and ⊢ v \in U, then *Gamma'* ⊢ [x:=v]*t12* \in *T'*.

  The substitution in the conclusion behaves differently depending on whether x and y are the same variable.

  First, suppose x = y. Then, by the definition of substitution, [x:=v]t = t, so we just need to show *Gamma* ⊢ t \in T. But we know x|->U; *Gamma* ⊢ t \in T, and, since y does not appear free in \y:*T11*. *t12*, the context invariance lemma yields *Gamma* ⊢ t \in T.

  Second, suppose x ≠ y. We know x|->U; y|->*T11*; *Gamma* ⊢ *t12* \in *T12* by inversion of the typing relation, from which y|->*T11*; x|->U; *Gamma* ⊢ *t12* \in *T12* follows by the context invariance lemma, so the IH applies, giving us y|->*T11*; *Gamma* ⊢ [x:=v]*t12* \in *T12*. By T_Abs, *Gamma* ⊢ \y:*T11*. [x:=v]*t12* \in *T11*→*T12*, and by the definition of substitution (noting that x ≠ y), *Gamma* ⊢ \y:*T11*. [x:=v]*t12* \in *T11*→*T12* as required.

- If t is an application *t1 t2*, the result follows straightforwardly from the definition of substitution and the induction hypotheses.

- The remaining cases are similar to the application case.

Technical note: This proof is a rare case where an induction on terms, rather than typing derivations, yields a simpler argument. The reason for this is that the assumption x|->U; *Gamma* ⊢ t \in T is not completely generic, in the sense that one of the "slots" in the typing relation – namely the context – is not just a variable, and this means that Coq's native induction tactic does not give us the induction hypothesis that we want. It is possible to work around this, but the needed generalization is a little tricky. The term t, on the other hand, is completely generic.

```
Proof with eauto.
  intros Gamma x U t v T Ht Ht'.
  generalize dependent Gamma. generalize dependent T.
  induction t; intros T Gamma H;
```

```
    inversion H; subst; simpl...
  -
    rename s into y. destruct (eqb_stringP x y) as [Hxy|Hxy].
    +
      subst.
      rewrite update_eq in H2.
      inversion H2; subst.
      eapply context_invariance. eassumption.
      apply typable_empty__closed in Ht'. unfold closed in Ht'.
      intros. apply (Ht' x0) in H0. inversion H0.
    +
      apply T_Var. rewrite update_neq in H2...
  -
    rename s into y. rename t into T. apply T_Abs.
    destruct (eqb_stringP x y) as [Hxy | Hxy].
    +
      subst. rewrite update_shadow in H5. apply H5.
    +
      apply IHt. eapply context_invariance...
      intros z Hafi. unfold update, t_update.
      destruct (eqb_stringP y z) as [Hyz | Hyz]; subst; trivial.
      rewrite ← eqb_string_false_iff in Hxy.
      rewrite Hxy...
Qed.
```

11.3.3 Main Theorem

We now have the tools we need to prove preservation: if a closed term *t* has type T and takes a step to *t'*, then *t'* is also a closed term with type T. In other words, the small-step reduction relation preserves types.

```
Theorem preservation : ∀ t t' T,
  empty ⊢ t \in T →
  t -> t' →
  empty ⊢ t' \in T.
```

Proof: By induction on the derivation of ⊢ *t* \in T.

- We can immediately rule out T_Var, T_Abs, T_Tru, and T_Fls as final rules in the derivation, since in each of these cases *t* cannot take a step.

- If the last rule in the derivation is T_App, then *t* = *t1 t2*, and there are subderivations showing that ⊢ *t1* \in *T11*→T and ⊢ *t2* \in *T11* plus two induction hypotheses: (1) *t1* –> *t1'* implies ⊢ *t1'* \in *T11*→T and (2) *t2* –> *t2'* implies ⊢ *t2'* \in *T11*. There are now three subcases to consider, one for each rule that could be used to show that *t1 t2* takes a step to *t'*.

  - If *t1 t2* takes a step by ST_App1, with *t1* stepping to *t1'*, then, by the first IH, *t1'* has the same type as *t1* (⊢ *t1* \in *T11*→T), and hence by T_App *t1' t2* has type T.
  - The ST_App2 case is similar, using the second IH.
  - If *t1 t2* takes a step by ST_AppAbs, then *t1* = \x:*T11*.*t12* and *t1 t2* steps to [x:=*t2*]*t12*; the desired result now follows from the substitution lemma.

- If the last rule in the derivation is T_Test, then *t* = test *t1* then *t2* else *t3*, with ⊢ *t1* \in Bool, ⊢ *t2* \in T, and ⊢ *t3* \in T, and with three induction hypotheses: (1) *t1* –> *t1'* implies ⊢ *t1'* \in Bool, (2) *t2* –> *t2'* implies ⊢ *t2'* \in T, and (3) *t3* –> *t3'* implies ⊢ *t3'* \in T.

  There are again three subcases to consider, depending on how *t* steps.

  - If *t* steps to *t2* or *t3* by ST_TestTru or ST_TestFalse, the result is immediate, since *t2* and *t3* have the same type as *t*.
  - Otherwise, *t* steps by ST_Test, and the desired conclusion follows directly from the first induction hypothesis.

```
Proof with eauto.
  remember (@empty ty) as Gamma.
  intros t t' T HT. generalize dependent t'.
  induction HT;
       intros t' HE; subst Gamma; subst;
       try solve [inversion HE; subst; auto].
  -
    inversion HE; subst...
    +
      apply substitution_preserves_typing with T11...
      inversion HT1...
Qed.
```

Exercise: 2 stars, standard, recommended (subject_expansion_stlc) An exercise in the Types chapter asked about the *subject expansion* property for the simple language of arithmetic and boolean expressions. This property did not hold for that language, and it also fails for STLC. That is, it is not always the case that, if *t* –> *t'* and **has_type** *t'* T, then empty ⊢ *t* \in T. Show this by giving a counter-example that does *not involve conditionals*.

You can state your counterexample informally in words, with a brief explanation.

```
Definition manual_grade_for_subject_expansion_stlc : option (nat×string) := None.
```

□

11.4 Type Soundness

Exercise: 2 stars, standard, optional (type_soundness) Put progress and preservation together and show that a well-typed term can *never* reach a stuck state.

```
Definition stuck (t:tm) : Prop :=
  (normal_form step) t ∧ ¬ value t.

Corollary soundness : ∀ t t' T,
  empty ⊢ t \in T →
  t ->* t' →
  ~(stuck t').
Proof.
  intros t t' T Hhas_type Hmulti. unfold stuck.
  intros [Hnf Hnot_val]. unfold normal_form in Hnf.
  induction Hmulti.
   Admitted.
```

□

11.5 Uniqueness of Types

Exercise: 3 stars, standard (unique_types) Another nice property of the STLC is that types are unique: a given term (in a given context) has at most one type.

```
Theorem unique_types : ∀ Gamma e T T',
  Gamma ⊢ e \in T →
  Gamma ⊢ e \in T' →
  T = T'.
Proof.
   Admitted.
```

□

11.6 Additional Exercises

Exercise: 1 star, standard (progress_preservation_statement) Without peeking at their statements above, write down the progress and preservation theorems for the simply typed lambda-calculus (as Coq theorems). You can write *Admitted* for the proofs.

```
Definition manual_grade_for_progress_preservation_statement : option (nat×string) := None.
```

□

Exercise: 2 stars, standard (stlc_variation1) Suppose we add a new term *zap* with the following reduction rule

(ST_Zap) t –> zap

and the following typing rule:

(T_Zap) Gamma |- zap \in T

Which of the following properties of the STLC remain true in the presence of these rules? For each property, write either "remains true" or "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

Definition manual_grade_for_stlc_variation1 : **option** (**nat**×**string**) := None.

□

Exercise: 2 stars, standard (stlc_variation2) Suppose instead that we add a new term foo with the following reduction rules:

(ST_Foo1) (\x:A. x) –> foo

(ST_Foo2) foo –> tru

Which of the following properties of the STLC remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

Definition manual_grade_for_stlc_variation2 : **option** (**nat**×**string**) := None.

□

Exercise: 2 stars, standard (stlc_variation3) Suppose instead that we remove the rule ST_App1 from the **step** relation. Which of the following properties of the STLC remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

Definition manual_grade_for_stlc_variation3 : **option** (**nat**×**string**) := None.

□

Exercise: 2 stars, standard, optional (stlc_variation4) Suppose instead that we add the following new rule to the reduction relation:

(ST_FunnyTestTru) (test tru then t1 else t2) –> tru

Which of the following properties of the STLC remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

☐

Exercise: 2 stars, standard, optional (stlc_variation5) Suppose instead that we add the following new rule to the typing relation:

Gamma |- t1 \in Bool->Bool->Bool Gamma |- t2 \in Bool

(T_FunnyApp) Gamma |- t1 t2 \in Bool

Which of the following properties of the STLC remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

☐

Exercise: 2 stars, standard, optional (stlc_variation6) Suppose instead that we add the following new rule to the typing relation:

Gamma |- t1 \in Bool Gamma |- t2 \in Bool

(T_FunnyApp') Gamma |- t1 t2 \in Bool

Which of the following properties of the STLC remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

☐

Exercise: 2 stars, standard, optional (stlc_variation7) Suppose we add the following new rule to the typing relation of the STLC:

(T_FunnyAbs) |- \x:Bool.t \in Bool

Which of the following properties of the STLC remain true in the presence of this rule? For each one, write either "remains true" or else "becomes false." If a property becomes false, give a counterexample.

- Determinism of **step**
- Progress
- Preservation

□

End STLCProp.

11.6.1 Exercise: STLC with Arithmetic

To see how the STLC might function as the core of a real programming language, let's extend it with a concrete base type of numbers and some constants and primitive operators.

```
Module STLCArith.
Import STLC.
```

To types, we add a base type of natural numbers (and remove booleans, for brevity).

```
Inductive ty : Type :=
  | Arrow : ty → ty → ty
  | Nat : ty.
```

To terms, we add natural number constants, along with successor, predecessor, multiplication, and zero-testing.

```
Inductive tm : Type :=
  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm
  | const : nat → tm
  | scc : tm → tm
  | prd : tm → tm
  | mlt : tm → tm → tm
  | test0 : tm → tm → tm → tm.
```

Exercise: 5 stars, standard (stlc_arith) Finish formalizing the definition and properties of the STLC extended with arithmetic. This is a longer exercise. Specifically:

1. Copy the core definitions for STLC that we went through, as well as the key lemmas and theorems, and paste them into the file at this point. Do not copy examples, exercises, etc. (In particular, make sure you don't copy any of the □ comments at the end of exercises, to avoid confusing the autograder.)

You should copy over five definitions:

- Fixpoint susbt
- Inductive value
- Inductive step
- Inductive has_type
- Inductive appears_free_in

And five theorems, with their proofs:

- Lemma context_invariance
- Lemma free_in_context
- Lemma substitution_preserves_typing
- Theorem preservation
- Theorem progress

It will be helpful to also copy over "Reserved Notation", "Notation", and "Hint Constructors" for these things.

2. Edit and extend the five definitions (subst, value, step, has_type, and appears_free_in) so they are appropriate for the new STLC extended with arithmetic.

3. Extend the proofs of all the five properties of the original STLC to deal with the new syntactic forms. Make sure Coq accepts the whole file.

```
Definition manual_grade_for_stlc_arith : option (nat×string) := None.
```

□

```
End STLCArith.
```

Chapter 12

MoreStlc: More on the Simply Typed Lambda-Calculus

```
Set Warnings "-notation-overridden,-parsing".
From PLF Require Import Maps.
From PLF Require Import Types.
From PLF Require Import Smallstep.
From PLF Require Import Stlc.
From Coq Require Import Strings.String.
```

12.1 Simple Extensions to STLC

The simply typed lambda-calculus has enough structure to make its theoretical properties interesting, but it is not much of a programming language.

In this chapter, we begin to close the gap with real-world languages by introducing a number of familiar features that have straightforward treatments at the level of typing.

12.1.1 Numbers

As we saw in exercise *stlc_arith* at the end of the StlcProp chapter, adding types, constants, and primitive operations for natural numbers is easy – basically just a matter of combining the Types and Stlc chapters. Adding more realistic numeric types like machine integers and floats is also straightforward, though of course the specifications of the numeric primitives become more fiddly.

12.1.2 Let Bindings

When writing a complex expression, it is useful to be able to give names to some of its subexpressions to avoid repetition and increase readability. Most languages provide one or more ways of doing this. In OCaml (and Coq), for example, we can write `let x=t1 in t2` to mean "reduce the expression *t1* to a value and bind the name `x` to this value while reducing *t2*."

Our `let`-binder follows OCaml in choosing a standard *call-by-value* evaluation order, where the `let`-bound term must be fully reduced before reduction of the `let`-body can begin. The typing rule *T_Let* tells us that the type of a `let` can be calculated by calculating the type of the `let`-bound term, extending the context with a binding with this type, and in this enriched context calculating the type of the body (which is then the type of the whole `let` expression).

At this point in the book, it's probably easier simply to look at the rules defining this new feature than to wade through a lot of English text conveying the same information. Here they are:

Syntax:

t ::= Terms | ... (other terms same as before) | let x=t in t let-binding

Reduction:

t1 –> t1'

(ST_Let1) let x=t1 in t2 –> let x=t1' in t2

(ST_LetValue) let x=v1 in t2 –> `x`:=*v1*t2

Typing:

Gamma |- t1 \in T1 x|->T1; Gamma |- t2 \in T2

(T_Let) Gamma |- let x=t1 in t2 \in T2

12.1.3 Pairs

Our functional programming examples in Coq have made frequent use of *pairs* of values. The type of such a pair is called a *product type*.

The formalization of pairs is almost too simple to be worth discussing. However, let's look briefly at the various parts of the definition to emphasize the common pattern.

In Coq, the primitive way of extracting the components of a pair is *pattern matching*. An alternative is to take `fst` and `snd` – the first- and second-projection operators – as primitives. Just for fun, let's do our pairs this way. For example, here's how we'd write a function that takes a pair of numbers and returns the pair of their sum and difference:

\x : Nat*Nat. let sum = x.fst + x.snd in let diff = x.fst - x.snd in (sum,diff)

Adding pairs to the simply typed lambda-calculus, then, involves adding two new forms of term – pairing, written (*t1*,*t2*), and projection, written *t.fst* for the first projection from *t* and *t.snd* for the second projection – plus one new type constructor, $T1 \times T2$, called the *product* of *T1* and *T2*.

Syntax:

t ::= Terms | ... | (t,t) pair | t.fst first projection | t.snd second projection

v ::= Values | ... | (v,v) pair value

T ::= Types | ... | T * T product type

For reduction, we need several new rules specifying how pairs and projection behave.

t1 –> t1'

(ST_Pair1) (t1,t2) –> (t1',t2)

t2 –> t2'

(ST_Pair2) (v1,t2) –> (v1,t2')

t1 –> t1'

(ST_Fst1) t1.fst –> t1'.fst

(ST_FstPair) (v1,v2).fst –> v1

t1 –> t1'

(ST_Snd1) t1.snd –> t1'.snd

(ST_SndPair) (v1,v2).snd –> v2

Rules ST_FstPair and ST_SndPair say that, when a fully reduced pair meets a first or second projection, the result is the appropriate component. The congruence rules *ST_Fst1* and *ST_Snd1* allow reduction to proceed under projections, when the term being projected from has not yet been fully reduced. ST_Pair1 and ST_Pair2 reduce the parts of pairs: first the left part, and then – when a value appears on the left – the right part. The ordering arising from the use of the metavariables v and t in these rules enforces a left-to-right evaluation strategy for pairs. (Note the implicit convention that metavariables like v and *v1* can only denote values.) We've also added a clause to the definition of values, above, specifying that (*v1*,*v2*) is a value. The fact that the components of a pair value must themselves be values ensures that a pair passed as an argument to a function will be fully reduced before the function body starts executing.

The typing rules for pairs and projections are straightforward.

Gamma |- t1 \in T1 Gamma |- t2 \in T2

(T_Pair) Gamma |- (t1,t2) \in T1*T2

Gamma |- t \in T1*T2

(T_Fst) Gamma |- t.fst \in T1

Gamma |- t \in T1*T2

(T_Snd) Gamma |- t.snd \in T2

T_Pair says that (*t1*,*t2*) has type $T1 \times T2$ if *t1* has type *T1* and *t2* has type *T2*. Conversely, T_Fst and T_Snd tell us that, if t has a product type $T1 \times T2$ (i.e., if it will reduce to a pair), then the types of the projections from this pair are *T1* and *T2*.

12.1.4 Unit

Another handy base type, found especially in languages in the ML family, is the singleton type Unit.

It has a single element – the term constant unit (with a small u) – and a typing rule making unit an element of Unit. We also add unit to the set of possible values – indeed, unit is the *only* possible result of reducing an expression of type Unit.

Syntax:

t ::= Terms | ... (other terms same as before) | unit unit

v ::= Values | ... | unit unit value

T ::= Types | ... | Unit unit type

Typing:

(T_Unit) Gamma |- unit \in Unit

It may seem a little strange to bother defining a type that has just one element – after all, wouldn't every computation living in such a type be trivial?

This is a fair question, and indeed in the STLC the Unit type is not especially critical (though we'll see two uses for it below). Where Unit really comes in handy is in richer languages with *side effects* – e.g., assignment statements that mutate variables or pointers, exceptions and other sorts of nonlocal control structures, etc. In such languages, it is convenient to have a type for the (trivial) result of an expression that is evaluated only for its effect.

12.1.5 Sums

Many programs need to deal with values that can take two distinct forms. For example, we might identify students in a university database using *either* their name *or* their id number. A search function might return *either* a matching value *or* an error code.

These are specific examples of a binary *sum type* (sometimes called a *disjoint union*), which describes a set of values drawn from one of two given types, e.g.:

Nat + Bool

We create elements of these types by *tagging* elements of the component types. For example, if n is a Nat then *inl* n is an element of Nat+Bool; similarly, if b is a Bool then *inr* b is a Nat+Bool. The names of the tags *inl* and *inr* arise from thinking of them as functions

inl \in Nat -> Nat + Bool inr \in Bool -> Nat + Bool

that "inject" elements of Nat or Bool into the left and right components of the sum type Nat+Bool. (But note that we don't actually treat them as functions in the way we formalize them: *inl* and *inr* are keywords, and *inl t* and *inr t* are primitive syntactic forms, not function applications.)

In general, the elements of a type *T1 + T2* consist of the elements of *T1* tagged with the token *inl*, plus the elements of *T2* tagged with *inr*.

As we've seen in Coq programming, one important use of sums is signaling errors:

div \in Nat -> Nat -> (Nat + Unit) div = \x:Nat. \y:Nat. test iszero y then inr unit else inl ...

The type Nat + Unit above is in fact isomorphic to **option nat** in Coq – i.e., it's easy to write functions that translate back and forth.

To *use* elements of sum types, we introduce a `case` construct (a very simplified form of Coq's `match`) to destruct them. For example, the following procedure converts a Nat+Bool into a Nat:

getNat \in Nat+Bool -> Nat getNat = \x:Nat+Bool. case x of inl n => n | inr b => test b then 1 else 0

More formally...

Syntax:

t ::= Terms | ... (other terms same as before) | inl T t tagging (left) | inr T t tagging (right) | case t of case inl x => t | inr x => t

v ::= Values | ... | inl T v tagged value (left) | inr T v tagged value (right)

T ::= Types | ... | T + T sum type

Reduction:

t1 –> t1'

(ST_Inl) inl T2 t1 –> inl T2 t1'

t2 –> t2'

(ST_Inr) inr T1 t2 –> inr T1 t2'

t0 –> t0'

(ST_Case) case t0 of inl x1 => t1 | inr x2 => t2 –> case t0' of inl x1 => t1 | inr x2 => t2

(ST_CaseInl) case (inl T2 v1) of inl x1 => t1 | inr x2 => t2 > *x1:=v1*t1

(ST_CaseInr) case (inr T1 v2) of inl x1 => t1 | inr x2 => t2 > *x2:=v1*t2

Typing:

Gamma |- t1 \in T1

(T_Inl) Gamma |- inl T2 t1 \in T1 + T2

Gamma |- t2 \in T2

(T_Inr) Gamma |- inr T1 t2 \in T1 + T2

Gamma |- t \in T1+T2 x1|->T1; Gamma |- t1 \in T x2|->T2; Gamma |- t2 \in T

(T_Case) Gamma |- case t of inl x1 => t1 | inr x2 => t2 \in T

We use the type annotation in *inl* and *inr* to make the typing relation simpler, similarly to what we did for functions.

Without this extra information, the typing rule T_Inl, for example, would have to say that, once we have shown that *t1* is an element of type *T1*, we can derive that *inl t1* is an element of *T1* + *T2* for *any* type *T2*. For example, we could derive both *inl* 5 : Nat + Nat and *inl* 5 : Nat + Bool (and infinitely many other types). This peculiarity (technically, a failure of uniqueness of types) would mean that we cannot build a typechecking algorithm simply by "reading the rules from bottom to top" as we could for all the other features seen so far.

There are various ways to deal with this difficulty. One simple one – which we've adopted here – forces the programmer to explicitly annotate the "other side" of a sum type when performing an injection. This is a bit heavy for programmers (so real languages adopt other solutions), but it is easy to understand and formalize.

12.1.6 Lists

The typing features we have seen can be classified into *base types* like Bool, and *type constructors* like → and × that build new types from old ones. Another useful type constructor is List. For every type T, the type List T describes finite-length lists whose elements are drawn from T.

In principle, we could encode lists using pairs, sums and *recursive* types. But giving semantics to recursive types is non-trivial. Instead, we'll just discuss the special case of lists directly.

Below we give the syntax, semantics, and typing rules for lists. Except for the fact that explicit type annotations are mandatory on nil and cannot appear on cons, these lists are essentially identical to those we built in Coq. We use *lcase* to destruct lists, to avoid dealing with questions like "what is the *head* of the empty list?"

For example, here is a function that calculates the sum of the first two elements of a list of numbers:

\x:List Nat. lcase x of nil => 0 | a::x' => lcase x' of nil => a | b::x" => a+b

Syntax:

t ::= Terms | ... | nil T | cons t t | lcase t of nil => t | x::x => t

v ::= Values | ... | nil T nil value | cons v v cons value

T ::= Types | ... | List T list of Ts

Reduction:

t1 –> t1'

(ST_Cons1) cons t1 t2 –> cons t1' t2

t2 –> t2'

(ST_Cons2) cons v1 t2 –> cons v1 t2'

t1 –> t1'

(ST_Lcase1) (lcase t1 of nil => t2 | xh::xt => t3) –> (lcase t1' of nil => t2 | xh::xt => t3)

(ST_LcaseNil) (lcase nil T of nil => t2 | xh::xt => t3) > t2

(ST_LcaseCons) (lcase (cons vh vt) of nil => t2 | xh::xt => t3) > xh:=vh,xt:=vtt3

Typing:

(T_Nil) Gamma |- nil T \in List T

Gamma |- t1 \in T Gamma |- t2 \in List T

(T_Cons) Gamma |- cons t1 t2 \in List T

Gamma |- t1 \in List T1 Gamma |- t2 \in T (h|->T1; t|->List T1; Gamma) |- t3 \in T

(T_Lcase) Gamma |- (lcase t1 of nil => t2 | h::t => t3) \in T

12.1.7 General Recursion

Another facility found in most programming languages (including Coq) is the ability to define recursive functions. For example, we would like to be able to define the factorial function like this:

fact = \x:Nat. test x=0 then 1 else x * (fact (pred x)))

Note that the right-hand side of this binder mentions the variable being bound – something that is not allowed by our formalization of `let` above.

Directly formalizing this "recursive definition" mechanism is possible, but it requires some extra effort: in particular, we'd have to pass around an "environment" of recursive function definitions in the definition of the **step** relation.

Here is another way of presenting recursive functions that is a bit more verbose but equally powerful and much more straightforward to formalize: instead of writing recursive definitions, we will define a *fixed-point operator* called `fix` that performs the "unfolding" of the recursive definition in the right-hand side as needed, during reduction.

For example, instead of

fact = \x:Nat. test x=0 then 1 else x * (fact (pred x)))

we will write:

fact = fix (\f:Nat->Nat. \x:Nat. test x=0 then 1 else x * (f (pred x)))

We can derive the latter from the former as follows:

- In the right-hand side of the definition of fact, replace recursive references to fact by a fresh variable f.
- Add an abstraction binding f at the front, with an appropriate type annotation. (Since we are using f in place of fact, which had type Nat→Nat, we should require f to have the same type.) The new abstraction has type (Nat→Nat) → (Nat→Nat).
- Apply `fix` to this abstraction. This application has type Nat→Nat.

- Use all of this as the right-hand side of an ordinary `let`-binding for fact.

The intuition is that the higher-order function f passed to `fix` is a *generator* for the fact function: if f is applied to a function that "approximates" the desired behavior of fact up to some number n (that is, a function that returns correct results on inputs less than or equal to n but we don't care what it does on inputs greater than n), then f returns a slightly better approximation to fact – a function that returns correct results for inputs up to n+1. Applying `fix` to this generator returns its *fixed point*, which is a function that gives the desired behavior for all inputs n.

(The term "fixed point" is used here in exactly the same sense as in ordinary mathematics, where a fixed point of a function f is an input x such that f(x) = x. Here, a fixed point of a function F of type (Nat→Nat)->(Nat→Nat) is a function f of type Nat→Nat such that F f behaves the same as f.)

Syntax:

t ::= Terms | ... | fix t fixed-point operator

Reduction:

t1 –> t1'

(ST_Fix1) fix t1 –> fix t1'

(ST_FixAbs) fix (\xf:T1.t2) –> *xf*:=`fix` (*xf*:*T1.t2*) t2

Typing:

Gamma |- t1 \in T1->T1

(T_Fix) Gamma |- fix t1 \in T1

Let's see how *ST_FixAbs* works by reducing fact 3 = `fix` F 3, where

F = (\f. \x. test x=0 then 1 else x * (f (pred x)))

(type annotations are omitted for brevity).

fix F 3

–> *ST_FixAbs* + ST_App1

(\x. test x=0 then 1 else x * (fix F (pred x))) 3

–> ST_AppAbs

test 3=0 then 1 else 3 * (fix F (pred 3))

–> *ST_Test0_Nonzero*

3 * (fix F (pred 3))

–> *ST_FixAbs* + ST_Mult2

3 * ((\x. test x=0 then 1 else x * (fix F (pred x))) (pred 3))

–> ST_PredNat + ST_Mult2 + ST_App2

3 * ((\x. test x=0 then 1 else x * (fix F (pred x))) 2)

–> ST_AppAbs + ST_Mult2

3 * (test 2=0 then 1 else 2 * (fix F (pred 2)))

–> *ST_Test0_Nonzero* + ST_Mult2

3 * (2 * (fix F (pred 2)))

–> *ST_FixAbs* + 2 x ST_Mult2
3 * (2 * ((\x. test x=0 then 1 else x * (fix F (pred x))) (pred 2)))
–> ST_PredNat + 2 x ST_Mult2 + ST_App2
3 * (2 * ((\x. test x=0 then 1 else x * (fix F (pred x))) 1))
–> ST_AppAbs + 2 x ST_Mult2
3 * (2 * (test 1=0 then 1 else 1 * (fix F (pred 1))))
–> *ST_Test0_Nonzero* + 2 x ST_Mult2
3 * (2 * (1 * (fix F (pred 1))))
–> *ST_FixAbs* + 3 x ST_Mult2
3 * (2 * (1 * ((\x. test x=0 then 1 else x * (fix F (pred x))) (pred 1))))
–> ST_PredNat + 3 x ST_Mult2 + ST_App2
3 * (2 * (1 * ((\x. test x=0 then 1 else x * (fix F (pred x))) 0)))
–> ST_AppAbs + 3 x ST_Mult2
3 * (2 * (1 * (test 0=0 then 1 else 0 * (fix F (pred 0)))))
–> ST_Test0Zero + 3 x ST_Mult2
3 * (2 * (1 * 1))
–> ST_MultNats + 2 x ST_Mult2
3 * (2 * 1)
–> ST_MultNats + ST_Mult2
3 * 2
–> ST_MultNats
6

One important point to note is that, unlike `Fixpoint` definitions in Coq, there is nothing to prevent functions defined using `fix` from diverging.

Exercise: 1 star, standard, optional (halve_fix) Translate this informal recursive definition into one using `fix`:

halve = \x:Nat. test x=0 then 0 else test (pred x)=0 then 0 else 1 + (halve (pred (pred x)))

□

Exercise: 1 star, standard, optional (fact_steps) Write down the sequence of steps that the term fact 1 goes through to reduce to a normal form (assuming the usual reduction rules for arithmetic operations).

□

The ability to form the fixed point of a function of type T→T for any T has some surprising consequences. In particular, it implies that *every* type is inhabited by some term. To see this, observe that, for every type T, we can define the term

fix (\x:T.x)

By *T_Fix* and T_Abs, this term has type T. By *ST_FixAbs* it reduces to itself, over and over again. Thus it is a *diverging element* of T.

More usefully, here's an example using `fix` to define a two-argument recursive function:

equal = fix (\eq:Nat->Nat->Bool. \m:Nat. \n:Nat. test m=0 then iszero n else test n=0 then fls else eq (pred m) (pred n))

And finally, here is an example where `fix` is used to define a *pair* of recursive functions (illustrating the fact that the type *T1* in the rule *T_Fix* need not be a function type):

evenodd = fix (\eo: (Nat->Bool * Nat->Bool). let e = \n:Nat. test n=0 then tru else eo.snd (pred n) in let o = \n:Nat. test n=0 then fls else eo.fst (pred n) in (e,o))

even = evenodd.fst odd = evenodd.snd

12.1.8 Records

As a final example of a basic extension of the STLC, let's look briefly at how to define *records* and their types. Intuitively, records can be obtained from pairs by two straightforward generalizations: they are n-ary (rather than just binary) and their fields are accessed by *label* (rather than position).

Syntax:

t ::= Terms | ... | {i1=t1, ..., in=tn} record | t.i projection

v ::= Values | ... | {i1=v1, ..., in=vn} record value

T ::= Types | ... | {i1:T1, ..., in:Tn} record type

The generalization from products should be pretty obvious. But it's worth noticing the ways in which what we've actually written is even *more* informal than the informal syntax we've used in previous sections and chapters: we've used "..." in several places to mean "any number of these," and we've omitted explicit mention of the usual side condition that the labels of a record should not contain any repetitions.

Reduction:

ti –> ti'

(ST_Rcd) {i1=v1, ..., im=vm, in=ti , ...} > {i1=v1, ..., im=vm, in=ti', ...}

t1 –> t1'

(ST_Proj1) t1.i –> t1'.i

(ST_ProjRcd) {..., i=vi, ...}.i –> vi

Again, these rules are a bit informal. For example, the first rule is intended to be read "if *ti* is the leftmost field that is not a value and if *ti* steps to *ti'*, then the whole record steps..." In the last rule, the intention is that there should be only one field called `i`, and that all the other fields must contain values.

The typing rules are also simple:

Gamma |- t1 \in T1 ... Gamma |- tn \in Tn

(T_Rcd) Gamma |- {i1=t1, ..., in=tn} \in {i1:T1, ..., in:Tn}

Gamma |- t \in {..., i:Ti, ...}

(T_Proj) Gamma |- t.i \in Ti

There are several ways to approach formalizing the above definitions.

- We can directly formalize the syntactic forms and inference rules, staying as close as possible to the form we've given them above. This is conceptually straightforward, and it's probably what we'd want to do if we were building a real compiler (in particular, it will allow us to print error messages in the form that programmers will find easy to understand). But the formal versions of the rules will not be very pretty or easy to work with, because all the ...s above will have to be replaced with explicit quantifications or comprehensions. For this reason, records are not included in the extended exercise at the end of this chapter. (It is still useful to discuss them informally here because they will help motivate the addition of subtyping to the type system when we get to the Sub chapter.)

- Alternatively, we could look for a smoother way of presenting records – for example, a binary presentation with one constructor for the empty record and another constructor for adding a single field to an existing record, instead of a single monolithic constructor that builds a whole record at once. This is the right way to go if we are primarily interested in studying the metatheory of the calculi with records, since it leads to clean and elegant definitions and proofs. Chapter Records shows how this can be done.

- Finally, if we like, we can avoid formalizing records altogether, by stipulating that record notations are just informal shorthands for more complex expressions involving pairs and product types. We sketch this approach in the next section.

Encoding Records (Optional)

Let's see how records can be encoded using just pairs and unit. (This clever encoding, as well as the observation that it also extends to systems with subtyping, is due to Luca Cardelli.)

First, observe that we can encode arbitrary-size *tuples* using nested pairs and the unit value. To avoid overloading the pair notation (*t1*,*t2*), we'll use curly braces without labels to write down tuples, so {} is the empty tuple, {5} is a singleton tuple, {5,6} is a 2-tuple (morally the same as a pair), {5,6,7} is a triple, etc.

{} —-> unit {t1, t2, ..., tn} —-> (t1, trest) where {t2, ..., tn} —-> trest

Similarly, we can encode tuple types using nested product types:

{} —-> Unit {T1, T2, ..., Tn} —-> T1 * TRest where {T2, ..., Tn} —-> TRest

The operation of projecting a field from a tuple can be encoded using a sequence of second projections followed by a first projection:

t.0 —-> t.fst t.(n+1) —-> (t.snd).n

Next, suppose that there is some total ordering on record labels, so that we can associate each label with a unique natural number. This number is called the *position* of the label. For example, we might assign positions like this:

LABEL POSITION a 0 b 1 c 2 ... ... bar 1395 ... ... foo 4460 ... ...

We use these positions to encode record values as tuples (i.e., as nested pairs) by sorting the fields according to their positions. For example:

{a=5,b=6} —-> {5,6} {a=5,c=7} —-> {5,unit,7} {c=7,a=5} —-> {5,unit,7} {c=5,b=3} —-> {unit,3,5} {f=8,c=5,a=7} —-> {7,unit,5,unit,unit,8} {f=8,c=5} —-> {unit,unit,5,unit,unit,8}

Note that each field appears in the position associated with its label, that the size of the tuple is determined by the label with the highest position, and that we fill in unused positions with unit.

We do exactly the same thing with record types:

{a:Nat,b:Nat} —-> {Nat,Nat} {c:Nat,a:Nat} —-> {Nat,Unit,Nat} {f:Nat,c:Nat} —-> {Unit,Unit,Nat,Unit,Unit,Nat}

Finally, record projection is encoded as a tuple projection from the appropriate position:

t.l —-> t.(position of l)

It is not hard to check that all the typing rules for the original "direct" presentation of records are validated by this encoding. (The reduction rules are "almost validated" – not quite, because the encoding reorders fields.)

Of course, this encoding will not be very efficient if we happen to use a record with label foo! But things are not actually as bad as they might seem: for example, if we assume that our compiler can see the whole program at the same time, we can *choose* the numbering of labels so that we assign small positions to the most frequently used labels. Indeed, there are industrial compilers that essentially do this!

Variants (Optional)

Just as products can be generalized to records, sums can be generalized to n-ary labeled types called *variants*. Instead of *T1+T2*, we can write something like <l1:*T1*,l2:*T2*,...*ln*:*Tn*> where l1,l2,... are field labels which are used both to build instances and as case arm labels.

These n-ary variants give us almost enough mechanism to build arbitrary inductive data types like lists and trees from scratch – the only thing missing is a way to allow *recursion* in type definitions. We won't cover this here, but detailed treatments can be found in many textbooks – e.g., Types and Programming Languages *Pierce* 2002 (in Bib.v).

12.2 Exercise: Formalizing the Extensions

`Module` STLCExtended.

Exercise: 3 stars, standard (STLCE_definitions) In this series of exercises, you will formalize some of the extensions described in this chapter. We've provided the necessary additions to the syntax of terms and types, and we've included a few examples that you can test your definitions with to make sure they are working as expected. You'll fill in the rest of the definitions and extend all the proofs accordingly.

To get you started, we've provided implementations for:

- numbers
- sums
- lists
- unit

You need to complete the implementations for:

- pairs
- let (which involves binding)
- `fix`

A good strategy is to work on the extensions one at a time, in two passes, rather than trying to work through the file from start to finish in a single pass. For each definition or proof, begin by reading carefully through the parts that are provided for you, referring to the text in the Stlc chapter for high-level intuitions and the embedded comments for detailed mechanics.

Syntax

```
Inductive ty : Type :=
  | Arrow : ty → ty → ty
  | Nat : ty
  | Sum : ty → ty → ty
  | List : ty → ty
  | Unit : ty
  | Prod : ty → ty → ty.
Inductive tm : Type :=

  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm

  | const : nat → tm
  | scc : tm → tm
  | prd : tm → tm
  | mlt : tm → tm → tm
  | test0 : tm → tm → tm → tm

  | tinl : ty → tm → tm
```

```
| tinr : ty → tm → tm
| tcase : tm → string → tm → string → tm → tm

| tnil : ty → tm
| tcons : tm → tm → tm
| tlcase : tm → tm → string → string → tm → tm

| unit : tm

| pair : tm → tm → tm
| fst : tm → tm
| snd : tm → tm

| tlet : string → tm → tm → tm

| tfix : tm → tm.
```

Note that, for brevity, we've omitted booleans and instead provided a single test0 form combining a zero test and a conditional. That is, instead of writing

test x = 0 then ... else ...

we'll write this:

test0 x then ... else ...

Substitution

```
Fixpoint subst (x : string) (s : tm) (t : tm) : tm :=
  match t with

  | var y ⇒
      if eqb_string x y then s else t
  | abs y T t1 ⇒
      abs y T (if eqb_string x y then t1 else (subst x s t1))
  | app t1 t2 ⇒
      app (subst x s t1) (subst x s t2)

  | const n ⇒
      const n
```

```
  | scc t1 ⇒
      scc (subst x s t1)
  | prd t1 ⇒
      prd (subst x s t1)
  | mlt t1 t2 ⇒
      mlt (subst x s t1) (subst x s t2)
  | test0 t1 t2 t3 ⇒
      test0 (subst x s t1) (subst x s t2) (subst x s t3)

  | tinl T t1 ⇒
      tinl T (subst x s t1)
  | tinr T t1 ⇒
      tinr T (subst x s t1)
  | tcase t0 y1 t1 y2 t2 ⇒
      tcase (subst x s t0)
          y1 (if eqb_string x y1 then t1 else (subst x s t1))
          y2 (if eqb_string x y2 then t2 else (subst x s t2))

  | tnil T ⇒
      tnil T
  | tcons t1 t2 ⇒
      tcons (subst x s t1) (subst x s t2)
  | tlcase t1 t2 y1 y2 t3 ⇒
      tlcase (subst x s t1) (subst x s t2) y1 y2
        (if eqb_string x y1 then
           t3
         else if eqb_string x y2 then t3
              else (subst x s t3))

  | unit ⇒ unit

  | _ ⇒ t
  end.
Notation "'[' x ':=' s ']' t" := (subst x s t) (at level 20).
```

Reduction

Next we define the values of our language.

```
Inductive value : tm → Prop :=

  | v_abs : ∀ x T11 t12,
      value (abs x T11 t12)

  | v_nat : ∀ n1,
      value (const n1)

  | v_inl : ∀ v T,
      value v →
      value (tinl T v)
  | v_inr : ∀ v T,
      value v →
      value (tinr T v)

  | v_lnil : ∀ T, value (tnil T)
  | v_lcons : ∀ v1 vl,
      value v1 →
      value vl →
      value (tcons v1 vl)

  | v_unit : value unit

  | v_pair : ∀ v1 v2,
      value v1 →
      value v2 →
      value (pair v1 v2).
Hint Constructors value.
Reserved Notation "t1 '–>' t2" (at level 40).
Inductive step : tm → tm → Prop :=

  | ST_AppAbs : ∀ x T11 t12 v2,
         value v2 →
         (app (abs x T11 t12) v2) -> [x:=v2]t12
  | ST_App1 : ∀ t1 t1' t2,
         t1 -> t1' →
         (app t1 t2) -> (app t1' t2)
  | ST_App2 : ∀ v1 t2 t2',
         value v1 →
```

```
          t2 -> t2' →
          (app v1 t2) -> (app v1 t2')

  | ST_Succ1 : ∀ t1 t1',
        t1 -> t1' →
        (scc t1) -> (scc t1')
  | ST_SuccNat : ∀ n1,
        (scc (const n1)) -> (const (S n1))
  | ST_Pred : ∀ t1 t1',
        t1 -> t1' →
        (prd t1) -> (prd t1')
  | ST_PredNat : ∀ n1,
        (prd (const n1)) -> (const (pred n1))
  | ST_Mult1 : ∀ t1 t1' t2,
        t1 -> t1' →
        (mlt t1 t2) -> (mlt t1' t2)
  | ST_Mult2 : ∀ v1 t2 t2',
        value v1 →
        t2 -> t2' →
        (mlt v1 t2) -> (mlt v1 t2')
  | ST_Mulconsts : ∀ n1 n2,
        (mlt (const n1) (const n2)) -> (const (mult n1 n2))
  | ST_Test01 : ∀ t1 t1' t2 t3,
        t1 -> t1' →
        (test0 t1 t2 t3) -> (test0 t1' t2 t3)
  | ST_Test0Zero : ∀ t2 t3,
        (test0 (const 0) t2 t3) -> t2
  | ST_Test0Nonzero : ∀ n t2 t3,
        (test0 (const (S n)) t2 t3) -> t3

  | ST_Inl : ∀ t1 t1' T,
         t1 -> t1' →
         (tinl T t1) -> (tinl T t1')
  | ST_Inr : ∀ t1 t1' T,
         t1 -> t1' →
         (tinr T t1) -> (tinr T t1')
  | ST_Case : ∀ t0 t0' x1 t1 x2 t2,
         t0 -> t0' →
         (tcase t0 x1 t1 x2 t2) -> (tcase t0' x1 t1 x2 t2)
  | ST_CaseInl : ∀ v0 x1 t1 x2 t2 T,
         value v0 →
         (tcase (tinl T v0) x1 t1 x2 t2) -> [x1:=v0]t1
```

```
  | ST_CaseInr : ∀ v0 x1 t1 x2 t2 T,
        value v0 →
        (tcase (tinr T v0) x1 t1 x2 t2) -> [x2:=v0]t2

  | ST_Cons1 : ∀ t1 t1' t2,
       t1 -> t1' →
       (tcons t1 t2) -> (tcons t1' t2)
  | ST_Cons2 : ∀ v1 t2 t2',
       value v1 →
       t2 -> t2' →
       (tcons v1 t2) -> (tcons v1 t2')
  | ST_Lcase1 : ∀ t1 t1' t2 x1 x2 t3,
       t1 -> t1' →
       (tlcase t1 t2 x1 x2 t3) -> (tlcase t1' t2 x1 x2 t3)
  | ST_LcaseNil : ∀ T t2 x1 x2 t3,
       (tlcase (tnil T) t2 x1 x2 t3) -> t2
  | ST_LcaseCons : ∀ v1 vl t2 x1 x2 t3,
       value v1 →
       value vl →
       (tlcase (tcons v1 vl) t2 x1 x2 t3)
         -> (subst x2 vl (subst x1 v1 t3))
```

```
where "t1 '->' t2" := (step t1 t2).
Notation multistep := (multi step).
Notation "t1 '->*' t2" := (multistep t1 t2) (at level 40).
Hint Constructors step.
```

Typing

```
Definition context := partial_map ty.
```

Next we define the typing rules. These are nearly direct transcriptions of the inference rules shown above.

```
Reserved Notation "Gamma '|-' t '\in' T" (at level 40).
```

```
Inductive has_type : context → tm → ty → Prop :=

  | T_Var : ∀ Gamma x T,
      Gamma x = Some T →
      Gamma ⊢ (var x) \in T
  | T_Abs : ∀ Gamma x T11 T12 t12,
      (update Gamma x T11) ⊢ t12 \in T12 →
      Gamma ⊢ (abs x T11 t12) \in (Arrow T11 T12)
  | T_App : ∀ T1 T2 Gamma t1 t2,
      Gamma ⊢ t1 \in (Arrow T1 T2) →
      Gamma ⊢ t2 \in T1 →
      Gamma ⊢ (app t1 t2) \in T2

  | T_Nat : ∀ Gamma n1,
      Gamma ⊢ (const n1) \in Nat
  | T_Succ : ∀ Gamma t1,
      Gamma ⊢ t1 \in Nat →
      Gamma ⊢ (scc t1) \in Nat
  | T_Pred : ∀ Gamma t1,
      Gamma ⊢ t1 \in Nat →
      Gamma ⊢ (prd t1) \in Nat
  | T_Mult : ∀ Gamma t1 t2,
      Gamma ⊢ t1 \in Nat →
      Gamma ⊢ t2 \in Nat →
      Gamma ⊢ (mlt t1 t2) \in Nat
  | T_Test0 : ∀ Gamma t1 t2 t3 T1,
      Gamma ⊢ t1 \in Nat →
      Gamma ⊢ t2 \in T1 →
      Gamma ⊢ t3 \in T1 →
      Gamma ⊢ (test0 t1 t2 t3) \in T1

  | T_Inl : ∀ Gamma t1 T1 T2,
      Gamma ⊢ t1 \in T1 →
      Gamma ⊢ (tinl T2 t1) \in (Sum T1 T2)
  | T_Inr : ∀ Gamma t2 T1 T2,
      Gamma ⊢ t2 \in T2 →
      Gamma ⊢ (tinr T1 t2) \in (Sum T1 T2)
  | T_Case : ∀ Gamma t0 x1 T1 t1 x2 T2 t2 T,
      Gamma ⊢ t0 \in (Sum T1 T2) →
      (update Gamma x1 T1) ⊢ t1 \in T →
      (update Gamma x2 T2) ⊢ t2 \in T →
      Gamma ⊢ (tcase t0 x1 t1 x2 t2) \in T
```

```
| T_Nil : ∀ Gamma T,
    Gamma ⊢ (tnil T) \in (List T)
| T_Cons : ∀ Gamma t1 t2 T1,
    Gamma ⊢ t1 \in T1 →
    Gamma ⊢ t2 \in (List T1) →
    Gamma ⊢ (tcons t1 t2) \in (List T1)
| T_Lcase : ∀ Gamma t1 T1 t2 x1 x2 t3 T2,
    Gamma ⊢ t1 \in (List T1) →
    Gamma ⊢ t2 \in T2 →
    (update (update Gamma x2 (List T1)) x1 T1) ⊢ t3 \in T2 →
    Gamma ⊢ (tlcase t1 t2 x1 x2 t3) \in T2

| T_Unit : ∀ Gamma,
    Gamma ⊢ unit \in Unit
```

```
where "Gamma '|-' t '\in' T" := (has_type Gamma t T).
Hint Constructors has_type.
Definition manual_grade_for_extensions_definition : option (nat×string) := None.
```
□

12.2.1 Examples

Exercise: 3 stars, standard (STLCE_examples) This section presents formalized versions of the examples from above (plus several more).

For each example, uncomment proofs and replace *Admitted* by Qed once you've implemented enough of the definitions for the tests to pass.

The examples at the beginning focus on specific features; you can use these to make sure your definition of a given feature is reasonable before moving on to extending the proofs later in the file with the cases relating to this feature. The later examples require all the features together, so you'll need to come back to these when you've got all the definitions filled in.

```
Module Examples.
```

Preliminaries

First, let's define a few variable names:

```
Open Scope string_scope.
Notation x := "x".
Notation y := "y".
Notation a := "a".
Notation f := "f".
Notation g := "g".
Notation l := "l".
Notation k := "k".
Notation i1 := "i1".
Notation i2 := "i2".
Notation processSum := "processSum".
Notation n := "n".
Notation eq := "eq".
Notation m := "m".
Notation evenodd := "evenodd".
Notation even := "even".
Notation odd := "odd".
Notation eo := "eo".
```

Next, a bit of Coq hackery to automate searching for typing derivations. You don't need to understand this bit in detail – just have a look over it so that you'll know what to look for if you ever find yourself needing to make custom extensions to `auto`.

The following `Hint` declarations say that, whenever `auto` arrives at a goal of the form (*Gamma* ⊢ (app *e1* *e1*) \in T), it should consider `eapply T_App`, leaving an existential variable for the middle type T1, and similar for *lcase*. That variable will then be filled in during the search for type derivations for *e1* and *e2*. We also include a hint to "try harder" when solving equality goals; this is useful to automate uses of T_Var (which includes an equality as a precondition).

```
Hint Extern 2 (has_type _ (app _ _) _) ⇒
  eapply T_App; auto.
Hint Extern 2 (has_type _ (tlcase _ _ _ _ _) _) ⇒
  eapply T_Lcase; auto.
Hint Extern 2 (_ = _) ⇒ compute; reflexivity.
```

Numbers

```
Module NumTest.
```

```
Definition test :=
  test0
    (prd
```

```
        (scc
          (prd
            (mlt
              (const 2)
              (const 0)))))
      (const 5)
      (const 6).

Example typechecks :
  empty ⊢ test \in Nat.
Proof.
  unfold test.
  auto 10.
   Admitted.

Example numtest_reduces :
  test ->* const 5.
Proof.
   Admitted.

End NUMTEST.
```

Products

```
Module PRODTEST.

Definition test :=
  snd
    (fst
      (pair
        (pair
          (const 5)
          (const 6))
        (const 7))).

Example typechecks :
  empty ⊢ test \in Nat.
Proof. unfold test. eauto 15. Admitted.

Example reduces :
  test ->* const 6.
Proof.
   Admitted.

End PRODTEST.
```

let

```
Module LetTest.

Definition test :=
  tlet
    x
    (prd (const 6))
    (scc (var x)).

Example typechecks :
  empty ⊢ test \in Nat.
Proof. unfold test. eauto 15. Admitted.

Example reduces :
  test ->* const 6.
Proof.
   Admitted.

End LetTest.
```

Sums

```
Module SumTest1.

Definition test :=
  tcase (tinl Nat (const 5))
    x (var x)
    y (var y).

Example typechecks :
  empty ⊢ test \in Nat.
Proof. unfold test. eauto 15. Admitted.

Example reduces :
  test ->* (const 5).
Proof.
   Admitted.

End SumTest1.

Module SumTest2.

Definition test :=
  tlet
    processSum
    (abs x (Sum Nat Nat)
      (tcase (var x)
         n (var n)
```

```
          n (test0 (var n) (const 1) (const 0))))
    (pair
      (app (var processSum) (tinl Nat (const 5)))
      (app (var processSum) (tinr Nat (const 5)))).
```

```
Example typechecks :
  empty ⊢ test \in (Prod Nat Nat).
Proof. unfold test. eauto 15. Admitted.
```

```
Example reduces :
  test ->* (pair (const 5) (const 0)).
Proof.
   Admitted.
```

```
End SumTest2.
```

Lists

```
Module ListTest.
```

```
Definition test :=
  tlet l
    (tcons (const 5) (tcons (const 6) (tnil Nat)))
    (tlcase (var l)
       (const 0)
       x y (mlt (var x) (var x))).
```

```
Example typechecks :
  empty ⊢ test \in Nat.
Proof. unfold test. eauto 20. Admitted.
```

```
Example reduces :
  test ->* (const 25).
Proof.
   Admitted.
```

```
End ListTest.
```

fix

```
Module FixTest1.
```

```
Definition fact :=
  tfix
    (abs f (Arrow Nat Nat)
      (abs a Nat
        (test0
           (var a)
```

```
            (const 1)
            (mlt
              (var a)
              (app (var f) (prd (var a))))))).
```

(Warning: you may be able to typecheck fact but still have some rules wrong!)

```
Example typechecks :
  empty ⊢ fact \in (Arrow Nat Nat).
Proof. unfold fact. auto 10. Admitted.

Example reduces :
  (app fact (const 4)) ->* (const 24).
Proof.
  Admitted.

End FixTest1.

Module FixTest2.

Definition map :=
  abs g (Arrow Nat Nat)
    (tfix
      (abs f (Arrow (List Nat) (List Nat))
        (abs l (List Nat)
          (tlcase (var l)
            (tnil Nat)
            a l (tcons (app (var g) (var a))
                          (app (var f) (var l))))))).

Example typechecks :
  empty ⊢ map \in
    (Arrow (Arrow Nat Nat)
      (Arrow (List Nat)
        (List Nat))).
Proof. unfold map. auto 10. Admitted.

Example reduces :
  app (app map (abs a Nat (scc (var a))))
         (tcons (const 1) (tcons (const 2) (tnil Nat)))
  ->* (tcons (const 2) (tcons (const 3) (tnil Nat))).
Proof.
  Admitted.

End FixTest2.

Module FixTest3.

Definition equal :=
  tfix
```

```
      (abs eq (Arrow Nat (Arrow Nat Nat))
        (abs m Nat
          (abs n Nat
            (test0 (var m)
              (test0 (var n) (const 1) (const 0))
              (test0 (var n)
                (const 0)
                (app (app (var eq)
                              (prd (var m)))
                        (prd (var n)))))))).
Example typechecks :
  empty ⊢ equal \in (Arrow Nat (Arrow Nat Nat)).
Proof. unfold equal. auto 10. Admitted.

Example reduces :
  (app (app equal (const 4)) (const 4)) ->* (const 1).
Proof.
  Admitted.

Example reduces2 :
  (app (app equal (const 4)) (const 5)) ->* (const 0).
Proof.
  Admitted.

End FixTest3.

Module FixTest4.

Definition eotest :=
  tlet evenodd
    (tfix
      (abs eo (Prod (Arrow Nat Nat) (Arrow Nat Nat))
        (pair
          (abs n Nat
            (test0 (var n)
              (const 1)
              (app (snd (var eo)) (prd (var n)))))
          (abs n Nat
            (test0 (var n)
              (const 0)
              (app (fst (var eo)) (prd (var n))))))))
  (tlet even (fst (var evenodd))
  (tlet odd (snd (var evenodd))
  (pair
    (app (var even) (const 3))
    (app (var even) (const 4))))).
```

```
Example typechecks :
  empty ⊢ eotest \in (Prod Nat Nat).
Proof. unfold eotest. eauto 30. Admitted.

Example reduces :
  eotest ->* (pair (const 0) (const 1)).
Proof.
    Admitted.

End FixTest4.

End Examples.
```
☐

12.2.2 Properties of Typing

The proofs of progress and preservation for this enriched system are essentially the same (though of course longer) as for the pure STLC.

Progress

Exercise: 3 stars, standard (STLCE_progress) Complete the proof of `progress`.

Theorem: Suppose empty |- t \in T. Then either 1. t is a value, or 2. t –> t' for some t'.

Proof: By induction on the given typing derivation.

```
Theorem progress : ∀ t T,
     empty ⊢ t \in T →
     value t ∨ ∃ t', t -> t'.
Proof with eauto.
  intros t T Ht.
  remember empty as Gamma.
  generalize dependent HeqGamma.
  induction Ht; intros HeqGamma; subst.
  -

    inversion H.
  -

    left...
  -

    right.
    destruct IHHt1; subst...
    +
      destruct IHHt2; subst...
      ×
```

```
      inversion H; subst; try solve_by_invert.
      ∃ (subst x t2 t12)...
    ×
      inversion H0 as [t2' Hstp]. ∃ (app t1 t2')...
  +
    inversion H as [t1' Hstp]. ∃ (app t1' t2)...
-
  left...
-
  right.
  destruct IHHt...
  +
    inversion H; subst; try solve_by_invert.
    ∃ (const (S n1))...
  +
    inversion H as [t1' Hstp].
    ∃ (scc t1')...
-
  right.
  destruct IHHt...
  +
    inversion H; subst; try solve_by_invert.
    ∃ (const (pred n1))...
  +
    inversion H as [t1' Hstp].
    ∃ (prd t1')...
-
  right.
  destruct IHHt1...
  +
    destruct IHHt2...
    ×
      inversion H; subst; try solve_by_invert.
      inversion H0; subst; try solve_by_invert.
      ∃ (const (mult n1 n0))...
    ×
      inversion H0 as [t2' Hstp].
      ∃ (mlt t1 t2')...
  +
```

```
    inversion H as [t1' Hstp].
    ∃ (mlt t1' t2)...
-
  right.
  destruct IHHt1...
  +
    inversion H; subst; try solve_by_invert.
    destruct n1 as [|n1'].
    ×
      ∃ t2...
    ×
      ∃ t3...
  +
    inversion H as [t1' H0].
    ∃ (test0 t1' t2 t3)...
-
  destruct IHHt...
  +
    right. inversion H as [t1' Hstp]...
-
  destruct IHHt...
  +
    right. inversion H as [t1' Hstp]...
-
  right.
  destruct IHHt1...
  +
    inversion H; subst; try solve_by_invert.
    ×
      ∃ ([x1:=v]t1)...
    ×
      ∃ ([x2:=v]t2)...
  +
    inversion H as [t0' Hstp].
    ∃ (tcase t0' x1 t1 x2 t2)...
-
  left...
-
  destruct IHHt1...
  +
    destruct IHHt2...
    ×
```

```
      right. inversion H0 as [t2' Hstp].
      ∃ (tcons t1 t2')...
    +
     right. inversion H as [t1' Hstp].
     ∃ (tcons t1' t2)...
  -
    right.
    destruct IHHt1...
    +
      inversion H; subst; try solve_by_invert.
      ×
        ∃ t2...
      ×
        ∃ ([x2:=vl]([x1:=v1]t3))...
    +
      inversion H as [t1' Hstp].
      ∃ (tlcase t1' t2 x1 x2 t3)...
  -
    left...
  Admitted.
Definition manual_grade_for_progress : option (nat×string) := None.
```
☐

Context Invariance

Exercise: 3 stars, standard (STLCE_context_invariance) Complete the definition of **appears_free_in**, and the proofs of context_invariance and free_in_context.

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
      appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
      appears_free_in x t1 → appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
      appears_free_in x t2 → appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
        y ≠ x →
        appears_free_in x t12 →
        appears_free_in x (abs y T11 t12)

  | afi_succ : ∀ x t,
     appears_free_in x t →
     appears_free_in x (scc t)
```

```
| afi_pred : ∀ x t,
    appears_free_in x t →
    appears_free_in x (prd t)
| afi_mult1 : ∀ x t1 t2,
    appears_free_in x t1 →
    appears_free_in x (mlt t1 t2)
| afi_mult2 : ∀ x t1 t2,
    appears_free_in x t2 →
    appears_free_in x (mlt t1 t2)
| afi_test01 : ∀ x t1 t2 t3,
    appears_free_in x t1 →
    appears_free_in x (test0 t1 t2 t3)
| afi_test02 : ∀ x t1 t2 t3,
    appears_free_in x t2 →
    appears_free_in x (test0 t1 t2 t3)
| afi_test03 : ∀ x t1 t2 t3,
    appears_free_in x t3 →
    appears_free_in x (test0 t1 t2 t3)

| afi_inl : ∀ x t T,
     appears_free_in x t →
     appears_free_in x (tinl T t)
| afi_inr : ∀ x t T,
     appears_free_in x t →
     appears_free_in x (tinr T t)
| afi_case0 : ∀ x t0 x1 t1 x2 t2,
     appears_free_in x t0 →
     appears_free_in x (tcase t0 x1 t1 x2 t2)
| afi_case1 : ∀ x t0 x1 t1 x2 t2,
     x1 ≠ x →
     appears_free_in x t1 →
     appears_free_in x (tcase t0 x1 t1 x2 t2)
| afi_case2 : ∀ x t0 x1 t1 x2 t2,
     x2 ≠ x →
     appears_free_in x t2 →
     appears_free_in x (tcase t0 x1 t1 x2 t2)

| afi_cons1 : ∀ x t1 t2,
    appears_free_in x t1 →
    appears_free_in x (tcons t1 t2)
| afi_cons2 : ∀ x t1 t2,
    appears_free_in x t2 →
```

```
    appears_free_in x (tcons t1 t2)
  | afi_lcase1 : ∀ x t1 t2 y1 y2 t3,
    appears_free_in x t1 →
    appears_free_in x (tlcase t1 t2 y1 y2 t3)
  | afi_lcase2 : ∀ x t1 t2 y1 y2 t3,
    appears_free_in x t2 →
    appears_free_in x (tlcase t1 t2 y1 y2 t3)
  | afi_lcase3 : ∀ x t1 t2 y1 y2 t3,
    y1 ≠ x →
    y2 ≠ x →
    appears_free_in x t3 →
    appears_free_in x (tlcase t1 t2 y1 y2 t3)
```

.

```
Hint Constructors appears_free_in.

Lemma context_invariance : ∀ Gamma Gamma' t S,
     Gamma ⊢ t \in S →
     (∀ x, appears_free_in x t → Gamma x = Gamma' x) →
     Gamma' ⊢ t \in S.
Proof with eauto 30.
  intros. generalize dependent Gamma'.
  induction H;
    intros Gamma' Heqv...
  -
    apply T_Var... rewrite ← Heqv...
  -
    apply T_Abs... apply IHhas_type. intros y Hafi.
    unfold update, t_update.
    destruct (eqb_stringP x y)...
  -
    eapply T_Case...
    + apply IHhas_type2. intros y Hafi.
      unfold update, t_update.
      destruct (eqb_stringP x1 y)...
```

```
    + apply IHhas_type3. intros y Hafi.
      unfold update, t_update.
      destruct (eqb_stringP x2 y)...
  -
    eapply T_Lcase... apply IHhas_type3. intros y Hafi.
    unfold update, t_update.
    destruct (eqb_stringP x1 y)...
    destruct (eqb_stringP x2 y)...

  Admitted.
```

```
Lemma free_in_context : ∀ x t T Gamma,
   appears_free_in x t →
   Gamma ⊢ t \in T →
   ∃ T', Gamma x = Some T'.
Proof with eauto.
  intros x t T Gamma Hafi Htyp.
  induction Htyp; inversion Hafi; subst...
  -
    destruct IHHtyp as [T' Hctx]... ∃ T'.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx...
  -
    destruct IHHtyp2 as [T' Hctx]... ∃ T'.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx...
  -
    destruct IHHtyp3 as [T' Hctx]... ∃ T'.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx...
  -
    clear Htyp1 IHHtyp1 Htyp2 IHHtyp2.
    destruct IHHtyp3 as [T' Hctx]... ∃ T'.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx...
    rewrite false_eqb_string in Hctx...

  Admitted.
```

```
Definition manual_grade_for_context_invariance : option (nat×string) := None.
```

□

Substitution

Exercise: 2 stars, standard (STLCE_subst_preserves_typing) Complete the proof of substitution_preserves_typing.

```
Lemma substitution_preserves_typing : ∀ Gamma x U v t S,
     (update Gamma x U) ⊢ t \in S →
     empty ⊢ v \in U →
     Gamma ⊢ ([x:=v]t) \in S.
Proof with eauto.
  intros Gamma x U v t S Htypt Htypv.
  generalize dependent Gamma. generalize dependent S.
  induction t;
    intros S Gamma Htypt; simpl; inversion Htypt; subst...
  -
    simpl. rename s into y.
    unfold update, t_update in H1.
    destruct (eqb_stringP x y).
    +

      subst.
      inversion H1; subst. clear H1.
      eapply context_invariance...
      intros x Hcontra.
      destruct (free_in_context _ _ S empty Hcontra)
        as [T' HT']...
      inversion HT'.
    +

      apply T_Var...
  -
    rename s into y. rename t into T11.
    apply T_Abs...
    destruct (eqb_stringP x y) as [Hxy|Hxy].
    +

      eapply context_invariance...
      subst.
      intros x Hafi. unfold update, t_update.
      destruct (eqb_string y x)...
    +

      apply IHt. eapply context_invariance...
      intros z Hafi. unfold update, t_update.
```

```
      destruct (eqb_stringP y z) as [Hyz|Hyz]...
      subst.
      rewrite false_eqb_string...
-
    rename s into x1. rename s0 into x2.
    eapply T_Case...
    +
      destruct (eqb_stringP x x1) as [Hxx1|Hxx1].
      ×
        eapply context_invariance...
        subst.
        intros z Hafi. unfold update, t_update.
        destruct (eqb_string x1 z)...
      ×
        apply IHt2. eapply context_invariance...
        intros z Hafi. unfold update, t_update.
        destruct (eqb_stringP x1 z) as [Hx1z|Hx1z]...
        subst. rewrite false_eqb_string...
    +
      destruct (eqb_stringP x x2) as [Hxx2|Hxx2].
      ×
        eapply context_invariance...
        subst.
        intros z Hafi. unfold update, t_update.
        destruct (eqb_string x2 z)...
      ×
        apply IHt3. eapply context_invariance...
        intros z Hafi. unfold update, t_update.
        destruct (eqb_stringP x2 z)...
        subst. rewrite false_eqb_string...
-
    rename s into y1. rename s0 into y2.
    eapply T_Lcase...
    destruct (eqb_stringP x y1).
    +
      simpl.
      eapply context_invariance...
      subst.
      intros z Hafi. unfold update, t_update.
      destruct (eqb_stringP y1 z)...
    +
      destruct (eqb_stringP x y2).
```

```
  ×
    eapply context_invariance...
    subst.
    intros z Hafi. unfold update, t_update.
    destruct (eqb_stringP y2 z)...
  ×
    apply IHt3. eapply context_invariance...
    intros z Hafi. unfold update, t_update.
    destruct (eqb_stringP y1 z)...
    subst. rewrite false_eqb_string...
    destruct (eqb_stringP y2 z)...
    subst. rewrite false_eqb_string...
Admitted.
Definition manual_grade_for_substitution_preserves_typing : option (nat×string) := None.
```
□

Preservation

Exercise: 3 stars, standard (STLCE_preservation) Complete the proof of preservation.

```
Theorem preservation : ∀ t t' T,
     empty ⊢ t \in T →
     t -> t' →
     empty ⊢ t' \in T.
Proof with eauto.
  intros t t' T HT.
  remember empty as Gamma. generalize dependent HeqGamma.
  generalize dependent t'.
  induction HT;
    intros t' HeqGamma HE; subst; inversion HE; subst...
  -

    inversion HE; subst...
    +

      apply substitution_preserves_typing with T1...
      inversion HT1...
  -
    inversion HT1; subst.
    eapply substitution_preserves_typing...
  -
    inversion HT1; subst.
```

```
    eapply substitution_preserves_typing...
  -
    +
      inversion HT1; subst.
      apply substitution_preserves_typing with (List T1)...
      apply substitution_preserves_typing with T1...
```

Admitted.

```
Definition manual_grade_for_preservation : option (nat×string) := None.
```

□

```
End STLCExtended.
```

Chapter 13

Sub: Subtyping

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Strings.String.
From PLF Require Import Maps.
From PLF Require Import Types.
From PLF Require Import Smallstep.
```

13.1 Concepts

We now turn to the study of *subtyping*, a key feature needed to support the object-oriented programming style.

13.1.1 A Motivating Example

Suppose we are writing a program involving two record types defined as follows:

Person = {name:String, age:Nat} Student = {name:String, age:Nat, gpa:Nat}

In the simply typed lamdba-calculus with records, the term

(\r:Person. (r.age)+1) {name="Pat",age=21,gpa=1}

is not typable, since it applies a function that wants a two-field record to an argument that actually provides three fields, while the T_App rule demands that the domain type of the function being applied must match the type of the argument precisely.

But this is silly: we're passing the function a *better* argument than it needs! The only thing the body of the function can possibly do with its record argument r is project the field *age* from it: nothing else is allowed by the type, and the presence or absence of an extra *gpa* field makes no difference at all. So, intuitively, it seems that this function should be applicable to any record value that has at least an *age* field.

More generally, a record with more fields is "at least as good in any context" as one with just a subset of these fields, in the sense that any value belonging to the longer record type can be used *safely* in any context expecting the shorter record type. If the context expects something with the shorter type but we actually give it something with the longer type, nothing bad will happen (formally, the program will not get stuck).

The principle at work here is called *subtyping*. We say that "S is a subtype of T", written S <: T, if a value of type S can safely be used in any context where a value of type T is expected. The idea of subtyping applies not only to records, but to all of the type constructors in the language – functions, pairs, etc.

Safe substitution principle:

- S is a subtype of T, written S <: T, if a value of type S can safely be used in any context where a value of type T is expected.

13.1.2 Subtyping and Object-Oriented Languages

Subtyping plays a fundamental role in many programming languages – in particular, it is closely related to the notion of *subclassing* in object-oriented languages.

An *object* in Java, C#, etc. can be thought of as a record, some of whose fields are functions ("methods") and some of whose fields are data values ("fields" or "instance variables"). Invoking a method m of an object *o* on some arguments *a1..an* roughly consists of projecting out the m field of *o* and applying it to *a1..an*.

The type of an object is called a *class* – or, in some languages, an *interface*. It describes which methods and which data fields the object offers. Classes and interfaces are related by the *subclass* and *subinterface* relations. An object belonging to a subclass (or subinterface) is required to provide all the methods and fields of one belonging to a superclass (or superinterface), plus possibly some more.

The fact that an object from a subclass can be used in place of one from a superclass provides a degree of flexibility that is extremely handy for organizing complex libraries. For example, a GUI toolkit like Java's Swing framework might define an abstract interface *Component* that collects together the common fields and methods of all objects having a graphical representation that can be displayed on the screen and interact with the user, such as the buttons, checkboxes, and scrollbars of a typical GUI. A method that relies only on this common interface can now be applied to any of these objects.

Of course, real object-oriented languages include many other features besides these. For example, fields can be updated. Fields and methods can be declared *private*. Classes can give *initializers* that are used when constructing objects. Code in subclasses can cooperate with code in superclasses via *inheritance*. Classes can have static methods and fields. Etc., etc.

To keep things simple here, we won't deal with any of these issues – in fact, we won't even talk any more about objects or classes. (There is a lot of discussion in *Pierce* 2002 (in Bib.v), if you are interested.) Instead, we'll study the core concepts behind the subclass / subinterface relation in the simplified setting of the STLC.

13.1.3 The Subsumption Rule

Our goal for this chapter is to add subtyping to the simply typed lambda-calculus (with some of the basic extensions from MoreStlc). This involves two steps:

- Defining a binary *subtype relation* between types.
- Enriching the typing relation to take subtyping into account.

The second step is actually very simple. We add just a single rule to the typing relation: the so-called *rule of subsumption*:

Gamma |- t \in S S <: T

(T_Sub) Gamma |- t \in T

This rule says, intuitively, that it is OK to "forget" some of what we know about a term.

For example, we may know that *t* is a record with two fields (e.g., S = {x:A→A, y:B→B}), but choose to forget about one of the fields (T = {y:B→B}) so that we can pass *t* to a function that requires just a single-field record.

13.1.4 The Subtype Relation

The first step – the definition of the relation S <: T – is where all the action is. Let's look at each of the clauses of its definition.

Structural Rules

To start off, we impose two "structural rules" that are independent of any particular type constructor: a rule of *transitivity*, which says intuitively that, if S is better (richer, safer) than *U* and *U* is better than T, then S is better than T...

S <: U U <: T

(S_Trans) S <: T

... and a rule of *reflexivity*, since certainly any type T is as good as itself:

(S_Refl) T <: T

Products

Now we consider the individual type constructors, one by one, beginning with product types. We consider one pair to be a subtype of another if each of its components is.

S1 <: T1 S2 <: T2

(S_Prod) S1 * S2 <: T1 * T2

Arrows

The subtyping rule for arrows is a little less intuitive. Suppose we have functions f and g with these types:

f : C -> Student g : (C->Person) -> D

That is, f is a function that yields a record of type Student, and g is a (higher-order) function that expects its argument to be a function yielding a record of type Person. Also suppose that Student is a subtype of Person. Then the application g f is safe even though their types do not match up precisely, because the only thing g can do with f is to apply it to some argument (of type C); the result will actually be a Student, while g will be expecting a Person, but this is safe because the only thing g can then do is to project out the two fields that it knows about (*name* and *age*), and these will certainly be among the fields that are present.

This example suggests that the subtyping rule for arrow types should say that two arrow types are in the subtype relation if their results are:

S2 <: T2

(S_Arrow_Co) S1 -> S2 <: S1 -> T2

We can generalize this to allow the arguments of the two arrow types to be in the subtype relation as well:

T1 <: S1 S2 <: T2

(S_Arrow) S1 -> S2 <: T1 -> T2

But notice that the argument types are subtypes "the other way round": in order to conclude that $S1 \rightarrow S2$ to be a subtype of $T1 \rightarrow T2$, it must be the case that $T1$ is a subtype of $S1$. The arrow constructor is said to be *contravariant* in its first argument and *covariant* in its second.

Here is an example that illustrates this:

f : Person -> C g : (Student -> C) -> D

The application g f is safe, because the only thing the body of g can do with f is to apply it to some argument of type Student. Since f requires records having (at least) the fields of a Person, this will always work. So Person $\rightarrow$ C is a subtype of Student $\rightarrow$ C since Student is a subtype of Person.

The intuition is that, if we have a function f of type $S1 \rightarrow S2$, then we know that f accepts elements of type $S1$; clearly, f will also accept elements of any subtype $T1$ of $S1$. The type of f also tells us that it returns elements of type $S2$; we can also view these results belonging to any supertype $T2$ of $S2$. That is, any function f of type $S1 \rightarrow S2$ can also be viewed as having type $T1 \rightarrow T2$.

Records

What about subtyping for record types?

The basic intuition is that it is always safe to use a "bigger" record in place of a "smaller" one. That is, given a record type, adding extra fields will always result in a subtype. If some code is expecting a record with fields x and y, it is perfectly safe for it to receive a record with fields x, y, and z; the z field will simply be ignored. For example,

{name:String, age:Nat, gpa:Nat} <: {name:String, age:Nat} {name:String, age:Nat} <: {name:String} {name:String} <: {}

This is known as "width subtyping" for records.

We can also create a subtype of a record type by replacing the type of one of its fields with a subtype. If some code is expecting a record with a field x of type T, it will be happy with a record having a field x of type S as long as S is a subtype of T. For example,

{x:Student} <: {x:Person}

This is known as "depth subtyping".

Finally, although the fields of a record type are written in a particular order, the order does not really matter. For example,

{name:String,age:Nat} <: {age:Nat,name:String}

This is known as "permutation subtyping".

We *could* formalize these requirements in a single subtyping rule for records as follows:

forall jk in j1..jn, exists ip in i1..im, such that jk=ip and Sp <: Tk

---

(S_Rcd) {i1:S1...im:Sm} <: {j1:T1...jn:Tn}

That is, the record on the left should have all the field labels of the one on the right (and possibly more), while the types of the common fields should be in the subtype relation.

However, this rule is rather heavy and hard to read, so it is often decomposed into three simpler rules, which can be combined using S_Trans to achieve all the same effects.

First, adding fields to the end of a record type gives a subtype:

n > m

(S_RcdWidth) {i1:T1...in:Tn} <: {i1:T1...im:Tm}

We can use S_RcdWidth to drop later fields of a multi-field record while keeping earlier fields, showing for example that {*age*:Nat,*name*:String} <: {*age*:Nat}.

Second, subtyping can be applied inside the components of a compound record type:

S1 <: T1 ... Sn <: Tn

(S_RcdDepth) {i1:S1...in:Sn} <: {i1:T1...in:Tn}

For example, we can use S_RcdDepth and S_RcdWidth together to show that {y:Student, x:Nat} <: {y:Person}.

Third, subtyping can reorder fields. For example, we want {*name*:String, *gpa*:Nat, *age*:Nat} <: Person. (We haven't quite achieved this yet: using just S_RcdDepth and S_RcdWidth we can only drop fields from the *end* of a record type.) So we add:

{i1:S1...in:Sn} is a permutation of {j1:T1...jn:Tn}

(S_RcdPerm) {i1:S1...in:Sn} <: {j1:T1...jn:Tn}

It is worth noting that full-blown language designs may choose not to adopt all of these subtyping rules. For example, in Java:

- Each class member (field or method) can be assigned a single index, adding new indices "on the right" as more members are added in subclasses (i.e., no permutation for classes).

- A class may implement multiple interfaces – so-called "multiple inheritance" of interfaces (i.e., permutation is allowed for interfaces).

- In early versions of Java, a subclass could not change the argument or result types of a method of its superclass (i.e., no depth subtyping or no arrow subtyping, depending how you look at it).

Exercise: 2 stars, standard, recommended (arrow_sub_wrong) Suppose we had incorrectly defined subtyping as covariant on both the right and the left of arrow types:

S1 <: T1 S2 <: T2

(S_Arrow_wrong) S1 -> S2 <: T1 -> T2

Give a concrete example of functions f and g with the following types...

f : Student -> Nat g : (Person -> Nat) -> Nat

... such that the application g f will get stuck during execution. (Use informal syntax. No need to prove formally that the application gets stuck.)

```
Definition manual_grade_for_arrow_sub_wrong : option (nat×string) := None.
```

□

Top

Finally, it is convenient to give the subtype relation a maximum element – a type that lies above every other type and is inhabited by all (well-typed) values. We do this by adding to the language one new type constant, called Top, together with a subtyping rule that places it above every other type in the subtype relation:

(S_Top) S <: Top

The Top type is an analog of the *Object* type in Java and C

Summary

In summary, we form the STLC with subtyping by starting with the pure STLC (over some set of base types) and then...

- adding a base type Top,

- adding the rule of subsumption
  Gamma |- t \in S S <: T
  - ———————————— (T_Sub) Gamma |- t \in T

  to the typing relation, and
- defining a subtype relation as follows:
  S <: U U <: T
  - ——————— (S_Trans) S <: T
    - —— (S_Refl)

    T <: T
    - ——- (S_Top)

    S <: Top

  S1 <: T1 S2 <: T2
  - —————————- (S_Prod) S1 * S2 <: T1 * T2

  T1 <: S1 S2 <: T2
  - —————————- (S_Arrow)

  S1 -> S2 <: T1 -> T2
  n > m
  - ——————————————— (S_RcdWidth)

  {i1:T1...in:Tn} <: {i1:T1...im:Tm}
  S1 <: T1 ... Sn <: Tn
  - ———————————————— (S_RcdDepth)

  {i1:S1...in:Sn} <: {i1:T1...in:Tn}
  {i1:S1...in:Sn} is a permutation of {j1:T1...jn:Tn}
  - ———————————————————————— (S_RcdPerm) {i1:S1...in:Sn} <: {j1:T1...jn:Tn}

13.1.5 Exercises

Exercise: 1 star, standard, optional (subtype_instances_tf_1) Suppose we have types S, T, U, and V with S <: T and U <: V. Which of the following subtyping assertions are then true? Write *true* or *false* after each one. (A, B, and C here are base types like Bool, Nat, etc.)

- T→S <: T→S
- Top→U <: S→Top
- (C→C) → (A×B) <: (C→C) → (Top×B)
- T→T→U <: S→S→V
- (T→T)->U <: (S→S)->V
- ((T→S)->T)->U <: ((S→T)->S)->V
- S×V <: T×U

□

Exercise: 2 stars, standard (subtype_order) The following types happen to form a linear order with respect to subtyping:

- Top
- Top → Student
- Student → Person
- Student → Top
- Person → Student

Write these types in order from the most specific to the most general.

Where does the type Top→Top→Student fit into this order? That is, state how Top → (Top → Student) compares with each of the five types above. It may be unrelated to some of them.

```
Definition manual_grade_for_subtype_order : option (nat×string) := None.
```

□

Exercise: 1 star, standard (subtype_instances_tf_2) Which of the following statements are true? Write *true* or *false* after each one.

forall S T, S <: T -> S->S <: T->T
forall S, S <: A->A -> exists T, S = T->T /\ T <: A
forall S T1 T2, (S <: T1 -> T2) -> exists S1 S2, S = S1 -> S2 /\ T1 <: S1 /\ S2 <: T2
exists S, S <: S->S
exists S, S->S <: S
forall S T1 T2, S <: T1*T2 -> exists S1 S2, S = S1*S2 /\ S1 <: T1 /\ S2 <: T2

```
Definition manual_grade_for_subtype_instances_tf_2 : option (nat×string) := None.
```
☐

Exercise: 1 star, standard (subtype_concepts_tf) Which of the following statements are true, and which are false?

- There exists a type that is a supertype of every other type.
- There exists a type that is a subtype of every other type.
- There exists a pair type that is a supertype of every other pair type.
- There exists a pair type that is a subtype of every other pair type.
- There exists an arrow type that is a supertype of every other arrow type.
- There exists an arrow type that is a subtype of every other arrow type.
- There is an infinite descending chain of distinct types in the subtype relation—that is, an infinite sequence of types *S0*, *S1*, etc., such that all the *Si*'s are different and each S(i+1) is a subtype of *Si*.
- There is an infinite *ascending* chain of distinct types in the subtype relation—that is, an infinite sequence of types *S0*, *S1*, etc., such that all the *Si*'s are different and each S(i+1) is a supertype of *Si*.

```
Definition manual_grade_for_subtype_concepts_tf : option (nat×string) := None.
```
☐

Exercise: 2 stars, standard (proper_subtypes) Is the following statement true or false? Briefly explain your answer. (Here Base n stands for a base type, where n is a string standing for the name of the base type. See the Syntax section below.)

forall T, ~(T = Bool \/ exists n, T = Base n) -> exists S, S <: T /\ S <> T

```
Definition manual_grade_for_proper_subtypes : option (nat×string) := None.
```
☐

Exercise: 2 stars, standard (small_large_1)

- What is the *smallest* type T ("smallest" in the subtype relation) that makes the following assertion true? (Assume we have Unit among the base types and unit as a constant of this type.)

  empty |- (\p:T*Top. p.fst) ((\z:A.z), unit) \in A->A

- What is the *largest* type T that makes the same assertion true?

Definition manual_grade_for_small_large_1 : **option** (**nat**×**string**) := None.

☐

Exercise: 2 stars, standard (small_large_2)

- What is the *smallest* type T that makes the following assertion true?

  empty |- (\p:(A->A * B->B). p) ((\z:A.z), (\z:B.z)) \in T

- What is the *largest* type T that makes the same assertion true?

Definition manual_grade_for_small_large_2 : **option** (**nat**×**string**) := None.

☐

Exercise: 2 stars, standard, optional (small_large_3)

- What is the *smallest* type T that makes the following assertion true?

  a:A |- (\p:(A*T). (p.snd) (p.fst)) (a, \z:A.z) \in A

- What is the *largest* type T that makes the same assertion true?

  ☐

Exercise: 2 stars, standard (small_large_4)

- What is the *smallest* type T that makes the following assertion true?

  exists S, empty |- (\p:(A*T). (p.snd) (p.fst)) \in S

- What is the *largest* type T that makes the same assertion true?

Definition manual_grade_for_small_large_4 : **option** (**nat**×**string**) := None.

☐

Exercise: 2 stars, standard (smallest_1) What is the *smallest* type T that makes the following assertion true?

exists S t, empty |- (\x:T. x x) t \in S

Definition manual_grade_for_smallest_1 : **option** (**nat**×**string**) := None.

☐

Exercise: 2 stars, standard (smallest_2) What is the *smallest* type T that makes the following assertion true?

empty |- (\x:Top. x) ((\z:A.z) , (\z:B.z)) \in T

```
Definition manual_grade_for_smallest_2 : option (nat×string) := None.
```

☐

Exercise: 3 stars, standard, optional (count_supertypes) How many supertypes does the record type {x:A, y:C→C} have? That is, how many different types T are there such that {x:A, y:C→C} <: T? (We consider two types to be different if they are written differently, even if each is a subtype of the other. For example, {x:A,y:B} and {y:B,x:A} are different.)

☐

Exercise: 2 stars, standard (pair_permutation) The subtyping rule for product types

S1 <: T1 S2 <: T2

(S_Prod) S1*S2 <: T1*T2

intuitively corresponds to the "depth" subtyping rule for records. Extending the analogy, we might consider adding a "permutation" rule

---

T1*T2 <: T2*T1

for products. Is this a good idea? Briefly explain why or why not.

```
Definition manual_grade_for_pair_permutation : option (nat×string) := None.
```

☐

13.2 Formal Definitions

Most of the definitions needed to formalize what we've discussed above – in particular, the syntax and operational semantics of the language – are identical to what we saw in the last chapter. We just need to extend the typing relation with the subsumption rule and add a new `Inductive` definition for the subtyping relation. Let's first do the identical bits.

13.2.1 Core Definitions

Syntax

In the rest of the chapter, we formalize just base types, booleans, arrow types, Unit, and Top, omitting record types and leaving product types as an exercise. For the sake of more interesting examples, we'll add an arbitrary set of base types like String, Float, etc. (Since they are just for examples, we won't bother adding any operations over these base types, but we could easily do so.)

```
Inductive ty : Type :=
  | Top : ty
  | Bool : ty
  | Base : string → ty
  | Arrow : ty → ty → ty
  | Unit : ty
.

Inductive tm : Type :=
  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm
  | tru : tm
  | fls : tm
  | test : tm → tm → tm → tm
  | unit : tm
.
```

Substitution

The definition of substitution remains exactly the same as for the pure STLC.

```
Fixpoint subst (x:string) (s:tm) (t:tm) : tm :=
  match t with
  | var y ⇒
      if eqb_string x y then s else t
  | abs y T t1 ⇒
      abs y T (if eqb_string x y then t1 else (subst x s t1))
  | app t1 t2 ⇒
      app (subst x s t1) (subst x s t2)
  | tru ⇒
      tru
  | fls ⇒
      fls
  | test t1 t2 t3 ⇒
      test (subst x s t1) (subst x s t2) (subst x s t3)
  | unit ⇒
      unit
  end.

Notation "'[' x ':=' s ']' t" := (subst x s t) (at level 20).
```

Reduction

Likewise the definitions of the **value** property and the **step** relation.

```
Inductive value : tm → Prop :=
  | v_abs : ∀ x T t,
      value (abs x T t)
  | v_true :
      value tru
  | v_false :
      value fls
  | v_unit :
      value unit
.
```

```
Hint Constructors value.
```

```
Reserved Notation "t1 '-->' t2" (at level 40).
```

```
Inductive step : tm → tm → Prop :=
  | ST_AppAbs : ∀ x T t12 v2,
          value v2 →
          (app (abs x T t12) v2) -> [x:=v2]t12
  | ST_App1 : ∀ t1 t1' t2,
          t1 -> t1' →
          (app t1 t2) -> (app t1' t2)
  | ST_App2 : ∀ v1 t2 t2',
          value v1 →
          t2 -> t2' →
          (app v1 t2) -> (app v1 t2')
  | ST_TestTrue : ∀ t1 t2,
      (test tru t1 t2) -> t1
  | ST_TestFalse : ∀ t1 t2,
      (test fls t1 t2) -> t2
  | ST_Test : ∀ t1 t1' t2 t3,
      t1 -> t1' →
      (test t1 t2 t3) -> (test t1' t2 t3)
where "t1 '-->' t2" := (step t1 t2).
```

```
Hint Constructors step.
```

13.2.2 Subtyping

Now we come to the interesting part. We begin by defining the subtyping relation and developing some of its important technical properties.

The definition of subtyping is just what we sketched in the motivating discussion.

```
Reserved Notation "T '<:' U" (at level 40).
```

```
Inductive subtype : ty → ty → Prop :=
  | S_Refl : ∀ T,
```

```
      T <: T
  | S_Trans : ∀ S U T,
      S <: U →
      U <: T →
      S <: T
  | S_Top : ∀ S,
      S <: Top
  | S_Arrow : ∀ S1 S2 T1 T2,
      T1 <: S1 →
      S2 <: T2 →
      (Arrow S1 S2) <: (Arrow T1 T2)
where "T '<:' U" := (subtype T U).
```

Note that we don't need any special rules for base types (Bool and Base): they are automatically subtypes of themselves (by S_Refl) and Top (by S_Top), and that's all we want.

```
Hint Constructors subtype.

Module Examples.

Open Scope string_scope.
Notation x := "x".
Notation y := "y".
Notation z := "z".

Notation A := (Base "A").
Notation B := (Base "B").
Notation C := (Base "C").

Notation String := (Base "String").
Notation Float := (Base "Float").
Notation Integer := (Base "Integer").

Example subtyping_example_0 :
  (Arrow C Bool) <: (Arrow C Top).
Proof. auto. Qed.
```

Exercise: 2 stars, standard, optional (subtyping_judgements) (Leave this exercise *Admitted* until after you have finished adding product types to the language – see exercise *products* – at least up to this point in the file).

Recall that, in chapter MoreStlc, the optional section "Encoding Records" describes how records can be encoded as pairs. Using this encoding, define pair types representing the following record types:

Person := { name : String } Student := { name : String ; gpa : Float } Employee := { name : String ; ssn : Integer } Definition Person : ty
. *Admitted.*

```
Definition Student : ty
  . Admitted.
Definition Employee : ty
  . Admitted.
```

Now use the definition of the subtype relation to prove the following:

```
Example sub_student_person :
  Student <: Person.
Proof.
  Admitted.
```

```
Example sub_employee_person :
  Employee <: Person.
Proof.
  Admitted.
```

□

The following facts are mostly easy to prove in Coq. To get full benefit from the exercises, make sure you also understand how to prove them on paper!

Exercise: 1 star, standard, optional (subtyping_example_1)

```
Example subtyping_example_1
:
  (Arrow Top Student) <: (Arrow (Arrow C C) Person).
Proof with eauto.
  Admitted.
```

□

Exercise: 1 star, standard, optional (subtyping_example_2)

```
Example subtyping_example_2
:
  (Arrow Top Person) <: (Arrow Person Top).
Proof with eauto.
  Admitted.
```

□

```
End Examples.
```

13.2.3 Typing

The only change to the typing relation is the addition of the rule of subsumption, T_Sub.

```
Definition context := partial_map ty.
```

```
Reserved Notation "Gamma '|-' t '\in' T" (at level 40).
```

```
Inductive has_type : context → tm → ty → Prop :=
```

```
  | T_Var : ∀ Gamma x T,
```

```
      Gamma x = Some T →
      Gamma ⊢ var x \in T
  | T_Abs : ∀ Gamma x T11 T12 t12,
      (x |-> T11 ; Gamma) ⊢ t12 \in T12 →
      Gamma ⊢ abs x T11 t12 \in Arrow T11 T12
  | T_App : ∀ T1 T2 Gamma t1 t2,
      Gamma ⊢ t1 \in Arrow T1 T2 →
      Gamma ⊢ t2 \in T1 →
      Gamma ⊢ app t1 t2 \in T2
  | T_True : ∀ Gamma,
       Gamma ⊢ tru \in Bool
  | T_False : ∀ Gamma,
       Gamma ⊢ fls \in Bool
  | T_Test : ∀ t1 t2 t3 T Gamma,
       Gamma ⊢ t1 \in Bool →
       Gamma ⊢ t2 \in T →
       Gamma ⊢ t3 \in T →
       Gamma ⊢ test t1 t2 t3 \in T
  | T_Unit : ∀ Gamma,
      Gamma ⊢ unit \in Unit

  | T_Sub : ∀ Gamma t S T,
      Gamma ⊢ t \in S →
      S <: T →
      Gamma ⊢ t \in T

where "Gamma '|-' t '\in' T" := (has_type Gamma t T).
```

```
Hint Constructors has_type.
```

The following hints help `auto` and `eauto` construct typing derivations. They are only used in a few places, but they give a nice illustration of what `auto` can do with a bit more programming. See chapter UseAuto for more on hints.

```
Hint Extern 2 (has_type _ (app _ _) _) ⇒
  eapply T_App; auto.
Hint Extern 2 (_ = _) ⇒ compute; reflexivity.
```

```
Module Examples2.
Import Examples.
```

Do the following exercises after you have added product types to the language. For each informal typing judgement, write it as a formal statement in Coq and prove it.

Exercise: 1 star, standard, optional (typing_example_0)

Exercise: 2 stars, standard, optional (typing_example_1)

Exercise: 2 stars, standard, optional (typing_example_2) End EXAMPLES2.

13.3 Properties

The fundamental properties of the system that we want to check are the same as always: progress and preservation. Unlike the extension of the STLC with references (chapter References), we don't need to change the *statements* of these properties to take subtyping into account. However, their proofs do become a little bit more involved.

13.3.1 Inversion Lemmas for Subtyping

Before we look at the properties of the typing relation, we need to establish a couple of critical structural properties of the subtype relation:

- Bool is the only subtype of Bool, and
- every subtype of an arrow type is itself an arrow type.

These are called *inversion lemmas* because they play a similar role in proofs as the built-in `inversion` tactic: given a hypothesis that there exists a derivation of some subtyping statement S <: T and some constraints on the shape of S and/or T, each inversion lemma reasons about what this derivation must look like to tell us something further about the shapes of S and T and the existence of subtype relations between their parts.

Exercise: 2 stars, standard, optional (sub_inversion_Bool) Lemma sub_inversion_Bool : ∀ *U*,

```
    U <: Bool →
    U = Bool.
Proof with auto.
  intros U Hs.
  remember Bool as V.
   Admitted.
  □
```

Exercise: 3 stars, standard (sub_inversion_arrow) Lemma sub_inversion_arrow : ∀ *U V1 V2*,

```
    U <: Arrow V1 V2 →
    ∃ U1 U2,
    U = Arrow U1 U2 ∧ V1 <: U1 ∧ U2 <: V2.
Proof with eauto.
```

```
intros U V1 V2 Hs.
remember (Arrow V1 V2) as V.
generalize dependent V2. generalize dependent V1.
 Admitted.
```
☐

13.3.2 Canonical Forms

The proof of the progress theorem – that a well-typed non-value can always take a step – doesn't need to change too much: we just need one small refinement. When we're considering the case where the term in question is an application *t1 t2* where both *t1* and *t2* are values, we need to know that *t1* has the *form* of a lambda-abstraction, so that we can apply the ST_AppAbs reduction rule. In the ordinary STLC, this is obvious: we know that *t1* has a function type *T11*→*T12*, and there is only one rule that can be used to give a function type to a value – rule T_Abs – and the form of the conclusion of this rule forces *t1* to be an abstraction.

In the STLC with subtyping, this reasoning doesn't quite work because there's another rule that can be used to show that a value has a function type: subsumption. Fortunately, this possibility doesn't change things much: if the last rule used to show *Gamma* ⊢ *t1* \in *T11*→*T12* is subsumption, then there is some *sub*-derivation whose subject is also *t1*, and we can reason by induction until we finally bottom out at a use of T_Abs.

This bit of reasoning is packaged up in the following lemma, which tells us the possible "canonical forms" (i.e., values) of function type.

Exercise: 3 stars, standard, optional (canonical_forms_of_arrow_types)

```
Lemma canonical_forms_of_arrow_types : ∀ Gamma s T1 T2,
  Gamma ⊢ s \in Arrow T1 T2 →
  value s →
  ∃ x S1 s2,
     s = abs x S1 s2.
Proof with eauto.
 Admitted.
```
☐

Similarly, the canonical forms of type Bool are the constants tru and fls.

```
Lemma canonical_forms_of_Bool : ∀ Gamma s,
  Gamma ⊢ s \in Bool →
  value s →
  s = tru ∨ s = fls.
Proof with eauto.
  intros Gamma s Hty Hv.
  remember Bool as T.
  induction Hty; try solve_by_invert...
```

-

```
subst. apply sub_inversion_Bool in H. subst...
Qed.
```

13.3.3 Progress

The proof of progress now proceeds just like the one for the pure STLC, except that in several places we invoke canonical forms lemmas...

Theorem (Progress): For any term *t* and type T, if empty ⊢ *t* \in T then *t* is a value or *t* –> *t'* for some term *t'*.

Proof: Let *t* and T be given, with empty ⊢ *t* \in T. Proceed by induction on the typing derivation.

The cases for T_Abs, T_Unit, T_True and T_False are immediate because abstractions, unit, tru, and fls are already values. The T_Var case is vacuous because variables cannot be typed in the empty context. The remaining cases are more interesting:

- If the last step in the typing derivation uses rule T_App, then there are terms *t1 t2* and types *T1* and *T2* such that *t* = *t1 t2*, T = *T2*, empty ⊢ *t1* \in *T1* → *T2*, and empty ⊢ *t2* \in *T1*. Moreover, by the induction hypothesis, either *t1* is a value or it steps, and either *t2* is a value or it steps. There are three possibilities to consider:

    - Suppose *t1* –> *t1'* for some term *t1'*. Then *t1 t2* –> *t1' t2* by ST_App1.
    - Suppose *t1* is a value and *t2* –> *t2'* for some term *t2'*. Then *t1 t2* –> *t1 t2'* by rule ST_App2 because *t1* is a value.
    - Finally, suppose *t1* and *t2* are both values. By the canonical forms lemma for arrow types, we know that *t1* has the form \x:*S1*.*s2* for some x, *S1*, and *s2*. But then (\x:*S1*.*s2*) *t2* –> [x:=*t2*]*s2* by ST_AppAbs, since *t2* is a value.

- If the final step of the derivation uses rule T_Test, then there are terms *t1*, *t2*, and *t3* such that *t* = test *t1* then *t2* else *t3*, with empty ⊢ *t1* \in Bool and with empty ⊢ *t2* \in T and empty ⊢ *t3* \in T. Moreover, by the induction hypothesis, either *t1* is a value or it steps.

    - If *t1* is a value, then by the canonical forms lemma for booleans, either *t1* = tru or *t1* = fls. In either case, *t* can step, using rule ST_TestTrue or ST_TestFalse.
    - If *t1* can step, then so can *t*, by rule ST_Test.

- If the final step of the derivation is by T_Sub, then there is a type S such that S <: T and empty ⊢ *t* \in S. The desired result is exactly the induction hypothesis for the typing subderivation.

Formally:

```
Theorem progress : ∀ t T,
     empty ⊢ t \in T →
     value t ∨ ∃ t', t -> t'.
Proof with eauto.
  intros t T Ht.
  remember empty as Gamma.
  revert HeqGamma.
  induction Ht;
    intros HeqGamma; subst...
  -
    inversion H.
  -
    right.
    destruct IHHt1; subst...
    +
      destruct IHHt2; subst...
      ×
        destruct (canonical_forms_of_arrow_types empty t1 T1 T2)
          as [x [S1 [t12 Heqt1]]]...
        subst. ∃ ([x:=t2]t12)...
      ×
        inversion H0 as [t2' Hstp]. ∃ (app t1 t2')...
    +
      inversion H as [t1' Hstp]. ∃ (app t1' t2)...
  -
    right.
    destruct IHHt1.
    + eauto.
    + assert (t1 = tru ∨ t1 = fls)
        by (eapply canonical_forms_of_Bool; eauto).
      inversion H0; subst...
    + inversion H. rename x into t1'. eauto.
Qed.
```

13.3.4 Inversion Lemmas for Typing

The proof of the preservation theorem also becomes a little more complex with the addition of subtyping. The reason is that, as with the "inversion lemmas for subtyping" above, there are a number of facts about the typing relation that are immediate from the definition in the pure STLC (formally: that can be obtained directly from the `inversion` tactic) but that require real proofs in the presence of subtyping because there are multiple ways to derive the same **has_type** statement.

The following inversion lemma tells us that, if we have a derivation of some typing statement *Gamma* ⊢ \x:*S1*.*t2* \in T whose subject is an abstraction, then there must be some subderivation giving a type to the body *t2*.

Lemma: If *Gamma* ⊢ \x:*S1*.*t2* \in T, then there is a type *S2* such that x|->*S1*; *Gamma* ⊢ *t2* \in *S2* and *S1* → *S2* <: T.

(Notice that the lemma does *not* say, "then T itself is an arrow type" – this is tempting, but false!)

Proof: Let *Gamma*, x, *S1*, *t2* and T be given as described. Proceed by induction on the derivation of *Gamma* ⊢ \x:*S1*.*t2* \in T. Cases T_Var, T_App, are vacuous as those rules cannot be used to give a type to a syntactic abstraction.

- If the last step of the derivation is a use of T_Abs then there is a type *T12* such that T = *S1* → *T12* and x:*S1*; *Gamma* ⊢ *t2* \in *T12*. Picking *T12* for *S2* gives us what we need: *S1* → *T12* <: *S1* → *T12* follows from S_Refl.
- If the last step of the derivation is a use of T_Sub then there is a type S such that S <: T and *Gamma* ⊢ \x:*S1*.*t2* \in S. The IH for the typing subderivation tells us that there is some type *S2* with *S1* → *S2* <: S and x:*S1*; *Gamma* ⊢ *t2* \in *S2*. Picking type *S2* gives us what we need, since *S1* → *S2* <: T then follows by S_Trans.

Formally:

```
Lemma typing_inversion_abs : ∀ Gamma x S1 t2 T,
     Gamma ⊢ (abs x S1 t2) \in T →
     ∃ S2,
       Arrow S1 S2 <: T
       ∧ (x |-> S1 ; Gamma) ⊢ t2 \in S2.
Proof with eauto.
  intros Gamma x S1 t2 T H.
  remember (abs x S1 t2) as t.
  induction H;
    inversion Heqt; subst; intros; try solve_by_invert.
  -
    ∃ T12...
  -
    destruct IHhas_type as [S2 [Hsub Hty]]...
  Qed.
```

Similarly...

```
Lemma typing_inversion_var : ∀ Gamma x T,
  Gamma ⊢ (var x) \in T →
  ∃ S,
    Gamma x = Some S ∧ S <: T.
```

```
Proof with eauto.
  intros Gamma x T Hty.
  remember (var x) as t.
  induction Hty; intros;
    inversion Heqt; subst; try solve_by_invert.
  -
    ∃ T...
  -
    destruct IHHty as [U [Hctx HsubU]]... Qed.
Lemma typing_inversion_app : ∀ Gamma t1 t2 T2,
  Gamma ⊢ (app t1 t2) \in T2 →
  ∃ T1 ,
    Gamma ⊢ t1 \in (Arrow T1 T2) ∧
    Gamma ⊢ t2 \in T1.
Proof with eauto.
  intros Gamma t1 t2 T2 Hty.
  remember (app t1 t2) as t.
  induction Hty; intros;
    inversion Heqt; subst; try solve_by_invert.
  -
    ∃ T1...
  -
    destruct IHHty as [U1 [Hty1 Hty2]]...
Qed.
Lemma typing_inversion_true : ∀ Gamma T,
  Gamma ⊢ tru \in T →
  Bool <: T.
Proof with eauto.
  intros Gamma T Htyp. remember tru as tu.
  induction Htyp;
    inversion Heqtu; subst; intros...
Qed.
Lemma typing_inversion_false : ∀ Gamma T,
  Gamma ⊢ fls \in T →
  Bool <: T.
Proof with eauto.
  intros Gamma T Htyp. remember fls as tu.
  induction Htyp;
    inversion Heqtu; subst; intros...
Qed.
Lemma typing_inversion_if : ∀ Gamma t1 t2 t3 T,
```

```
  Gamma ⊢ (test t1 t2 t3) \in T →
  Gamma ⊢ t1 \in Bool
  ∧ Gamma ⊢ t2 \in T
  ∧ Gamma ⊢ t3 \in T.
Proof with eauto.
  intros Gamma t1 t2 t3 T Hty.
  remember (test t1 t2 t3) as t.
  induction Hty; intros;
    inversion Heqt; subst; try solve_by_invert.
  -
    auto.
  -
    destruct (IHHty H0) as [H1 [H2 H3]]...
Qed.

Lemma typing_inversion_unit : ∀ Gamma T,
  Gamma ⊢ unit \in T →
    Unit <: T.
Proof with eauto.
  intros Gamma T Htyp. remember unit as tu.
  induction Htyp;
    inversion Heqtu; subst; intros...
Qed.
```

The inversion lemmas for typing and for subtyping between arrow types can be packaged up as a useful "combination lemma" telling us exactly what we'll actually require below.

```
Lemma abs_arrow : ∀ x S1 s2 T1 T2,
  empty ⊢ (abs x S1 s2) \in (Arrow T1 T2) →
     T1 <: S1
  ∧ (x |-> S1 ; empty) ⊢ s2 \in T2.
Proof with eauto.
  intros x S1 s2 T1 T2 Hty.
  apply typing_inversion_abs in Hty.
  inversion Hty as [S2 [Hsub Hty1]].
  apply sub_inversion_arrow in Hsub.
  inversion Hsub as [U1 [U2 [Heq [Hsub1 Hsub2]]]].
  inversion Heq; subst... Qed.
```

13.3.5 Context Invariance

The context invariance lemma follows the same pattern as in the pure STLC.

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
```

```
        appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
        appears_free_in x t1 → appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
        appears_free_in x t2 → appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
          y ≠ x →
          appears_free_in x t12 →
          appears_free_in x (abs y T11 t12)
  | afi_test1 : ∀ x t1 t2 t3,
        appears_free_in x t1 →
        appears_free_in x (test t1 t2 t3)
  | afi_test2 : ∀ x t1 t2 t3,
        appears_free_in x t2 →
        appears_free_in x (test t1 t2 t3)
  | afi_test3 : ∀ x t1 t2 t3,
        appears_free_in x t3 →
        appears_free_in x (test t1 t2 t3)
.

Hint Constructors appears_free_in.

Lemma context_invariance : ∀ Gamma Gamma' t S,
      Gamma ⊢ t \in S →
      (∀ x, appears_free_in x t → Gamma x = Gamma' x) →
      Gamma' ⊢ t \in S.
Proof with eauto.
  intros. generalize dependent Gamma'.
  induction H;
    intros Gamma' Heqv...
  -
    apply T_Var... rewrite ← Heqv...
  -
    apply T_Abs... apply IHhas_type. intros x0 Hafi.
    unfold update, t_update. destruct (eqb_stringP x x0)...
  -
    apply T_Test...
Qed.

Lemma free_in_context : ∀ x t T Gamma,
    appears_free_in x t →
    Gamma ⊢ t \in T →
    ∃ T', Gamma x = Some T'.
Proof with eauto.
  intros x t T Gamma Hafi Htyp.
```

```
  induction Htyp;
      subst; inversion Hafi; subst...
  -
    destruct (IHHtyp H4) as [T Hctx]. ∃ T.
    unfold update, t_update in Hctx.
    rewrite ← eqb_string_false_iff in H2.
    rewrite H2 in Hctx... Qed.
```

13.3.6 Substitution

The *substitution lemma* is proved along the same lines as for the pure STLC. The only significant change is that there are several places where, instead of the built-in `inversion` tactic, we need to use the inversion lemmas that we proved above to extract structural information from assumptions about the well-typedness of subterms.

```
Lemma substitution_preserves_typing : ∀ Gamma x U v t S,
      (x |-> U ; Gamma) ⊢ t \in S →
      empty ⊢ v \in U →
      Gamma ⊢ [x:=v]t \in S.
Proof with eauto.
  intros Gamma x U v t S Htypt Htypv.
  generalize dependent S. generalize dependent Gamma.
  induction t; intros; simpl.
  -
    rename s into y.
    destruct (typing_inversion_var _ _ _ Htypt)
        as [T [Hctx Hsub]].
    unfold update, t_update in Hctx.
    destruct (eqb_stringP x y) as [Hxy|Hxy]; eauto;
    subst.
    inversion Hctx; subst. clear Hctx.
    apply context_invariance with empty...
    intros x Hcontra.
    destruct (free_in_context _ _ S empty Hcontra)
        as [T' HT']...
    inversion HT'.
  -
    destruct (typing_inversion_app _ _ _ _ Htypt)
        as [T1 [Htypt1 Htypt2]].
    eapply T_App...
  -
    rename s into y. rename t into T1.
    destruct (typing_inversion_abs _ _ _ _ _ Htypt)
```

```
      as [T2 [Hsub Htypt2]].
    apply T_Sub with (Arrow T1 T2)... apply T_Abs...
    destruct (eqb_stringP x y) as [Hxy|Hxy].
    +
      eapply context_invariance...
      subst.
      intros x Hafi. unfold update, t_update.
      destruct (eqb_string y x)...
    +
      apply IHt. eapply context_invariance...
      intros z Hafi. unfold update, t_update.
      destruct (eqb_stringP y z)...
      subst.
      rewrite <- eqb_string_false_iff in Hxy. rewrite Hxy...
  -
      assert (Bool <: S)
        by apply (typing_inversion_true _ _ Htypt)...
  -
      assert (Bool <: S)
        by apply (typing_inversion_false _ _ Htypt)...
  -
    assert ((x |-> U ; Gamma) |- t1 \in Bool
          /\ (x |-> U ; Gamma) |- t2 \in S
          /\ (x |-> U ; Gamma) |- t3 \in S)
      by apply (typing_inversion_if _ _ _ _ _ Htypt).
    inversion H as [H1 [H2 H3]].
    apply IHt1 in H1. apply IHt2 in H2. apply IHt3 in H3.
    auto.
  -
    assert (Unit <: S)
      by apply (typing_inversion_unit _ _ Htypt)...
Qed.
```

13.3.7 Preservation

The proof of preservation now proceeds pretty much as in earlier chapters, using the substitution lemma at the appropriate point and again using inversion lemmas from above to extract structural information from typing assumptions.

Theorem (Preservation): If *t*, *t'* are terms and T is a type such that empty ⊢ *t* \in T and *t* –> *t'*, then empty ⊢ *t'* \in T.

Proof: Let *t* and T be given such that empty ⊢ *t* \in T. We proceed by induction on the structure of this typing derivation, leaving *t'* general. The cases T_Abs, T_Unit, T_True, and T_False cases are vacuous because abstractions and constants don't step. Case T_Var is vacuous as well, since the context is empty.

- If the final step of the derivation is by T_App, then there are terms *t1* and *t2* and types *T1* and *T2* such that t = *t1 t2*, T = *T2*, empty ⊢ *t1* \in *T1* → *T2*, and empty ⊢ *t2* \in *T1*.

  By the definition of the step relation, there are three ways *t1 t2* can step. Cases ST_App1 and ST_App2 follow immediately by the induction hypotheses for the typing subderivations and a use of T_App.

  Suppose instead *t1 t2* steps by ST_AppAbs. Then *t1* = \x:*S*.*t12* for some type S and term *t12*, and *t'* = [x:=*t2*]*t12*.

  By lemma abs_arrow, we have *T1* <: S and x:*S1* ⊢ *s2* \in *T2*. It then follows by the substitution lemma (substitution_preserves_typing) that empty ⊢ [x:=*t2*] *t12* \in *T2* as desired.

  - If the final step of the derivation uses rule T_Test, then there are terms *t1*, *t2*, and *t3* such that t = test *t1* then *t2* else *t3*, with empty ⊢ *t1* \in Bool and with empty ⊢ *t2* \in T and empty ⊢ *t3* \in T. Moreover, by the induction hypothesis, if *t1* steps to *t1'* then empty ⊢ *t1'* : Bool. There are three cases to consider, depending on which rule was used to show t –> *t'*.

    - If t –> *t'* by rule ST_Test, then *t'* = test *t1'* then *t2* else *t3* with *t1* –> *t1'*. By the induction hypothesis, empty ⊢ *t1'* \in Bool, and so empty ⊢ *t'* \in T by T_Test.
    - If t –> *t'* by rule ST_TestTrue or ST_TestFalse, then either *t'* = *t2* or *t'* = *t3*, and empty ⊢ *t'* \in T follows by assumption.

- If the final step of the derivation is by T_Sub, then there is a type S such that S <: T and empty ⊢ t \in S. The result is immediate by the induction hypothesis for the typing subderivation and an application of T_Sub. □

```
Theorem preservation : ∀ t t' T,
     empty ⊢ t \in T →
     t -> t' →
     empty ⊢ t' \in T.
Proof with eauto.
  intros t t' T HT.
  remember empty as Gamma. generalize dependent HeqGamma.
  generalize dependent t'.
  induction HT;
    intros t' HeqGamma HE; subst; inversion HE; subst...
  -
    inversion HE; subst...
```

+

```
destruct (abs_arrow _ _ _ _ _ HT1) as [HA1 HA2].
apply substitution_preserves_typing with T...
```

Qed.

13.3.8 Records, via Products and Top

This formalization of the STLC with subtyping omits record types for brevity. If we want to deal with them more seriously, we have two choices.

First, we can treat them as part of the core language, writing down proper syntax, typing, and subtyping rules for them. Chapter RecordSub shows how this extension works.

On the other hand, if we are treating them as a derived form that is desugared in the parser, then we shouldn't need any new rules: we should just check that the existing rules for subtyping product and Unit types give rise to reasonable rules for record subtyping via this encoding. To do this, we just need to make one small change to the encoding described earlier: instead of using Unit as the base case in the encoding of tuples and the "don't care" placeholder in the encoding of records, we use Top. So:

{a:Nat, b:Nat} —-> {Nat,Nat} i.e., (Nat,(Nat,Top)) {c:Nat, a:Nat} —-> {Nat,Top,Nat} i.e., (Nat,(Top,(Nat,Top)))

The encoding of record values doesn't change at all. It is easy (and instructive) to check that the subtyping rules above are validated by the encoding.

13.3.9 Exercises

Exercise: 2 stars, standard (variations) Each part of this problem suggests a different way of changing the definition of the STLC with Unit and subtyping. (These changes are not cumulative: each part starts from the original language.) In each part, list which properties (Progress, Preservation, both, or neither) become false. If a property becomes false, give a counterexample.

- Suppose we add the following typing rule:

  Gamma |- t \in S1->S2 S1 <: T1 T1 <: S1 S2 <: T2

  - ————————————————————- (T_Funny1) Gamma |- t \in T1->T2

- Suppose we add the following reduction rule:

  - ———————————- (ST_Funny21)

  unit –> (\x:Top. x)

- Suppose we add the following subtyping rule:

  - —————————— (S_Funny3)

Unit <: Top->Top

- Suppose we add the following subtyping rule:

  - ———————— (S_Funny4)

  Top->Top <: Unit

- Suppose we add the following reduction rule:

  - —————————— (ST_Funny5)

  (unit t) –> (t unit)

- Suppose we add the same reduction rule *and* a new typing rule:

  - —————————— (ST_Funny5)

  (unit t) –> (t unit)

  - —————————————- (T_Funny6)

  empty |- unit \in Top->Top

- Suppose we *change* the arrow subtyping rule to:

  S1 <: T1 S2 <: T2

  - ————————- (S_Arrow')

  S1->S2 <: T1->T2

```
Definition manual_grade_for_variations : option (nat×string) := None.
```
□

13.4 Exercise: Adding Products

Exercise: 5 stars, standard (products) Adding pairs, projections, and product types to the system we have defined is a relatively straightforward matter. Carry out this extension by modifying the definitions and proofs above:

- Add constructors for pairs, first and second projections, and product types to the definitions of **ty** and **tm**, and extend the surrounding definitions accordingly (refer to chapter *MoreSTLC*):

  - value relation

    - substitution
    - operational semantics
    - typing relation

- Extend the subtyping relation with this rule:

  S1 <: T1 S2 <: T2

    - ———————————- (S_Prod) S1 * S2 <: T1 * T2

- Extend the proofs of progress, preservation, and all their supporting lemmas to deal with the new constructs. (You'll also need to add a couple of completely new lemmas.)

```
Definition manual_grade_for_products : option (nat×string) := None.
```

☐

Chapter 14

Typechecking: A Typechecker for STLC

The **has_type** relation of the STLC defines what it means for a term to belong to a type (in some context). But it doesn't, by itself, give us an algorithm for *checking* whether or not a term is well typed.

Fortunately, the rules defining **has_type** are *syntax directed* – that is, for every syntactic form of the language, there is just one rule that can be used to give a type to terms of that form. This makes it straightforward to translate the typing rules into clauses of a typechecking *function* that takes a term and a context and either returns the term's type or else signals that the term is not typable.

This short chapter constructs such a function and proves it correct.

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Bool.Bool.
From PLF Require Import Maps.
From PLF Require Import Smallstep.
From PLF Require Import Stlc.
From PLF Require MoreStlc.
```

```
Module STLCTypes.
Export STLC.
```

14.1 Comparing Types

First, we need a function to compare two types for equality...

```
Fixpoint eqb_ty (T1 T2:ty) : bool :=
  match T1,T2 with
  | Bool, Bool ⇒
      true
  | Arrow T11 T12, Arrow T21 T22 ⇒
      andb (eqb_ty T11 T21) (eqb_ty T12 T22)
  | _,_ ⇒
```

```
      false
  end.
```

... and we need to establish the usual two-way connection between the boolean result returned by eqb_ty and the logical proposition that its inputs are equal.

```
Lemma eqb_ty_refl : ∀ T1,
  eqb_ty T1 T1 = true.
Proof.
  intros T1. induction T1; simpl.
    reflexivity.
    rewrite IHT1_1. rewrite IHT1_2. reflexivity. Qed.

Lemma eqb_ty__eq : ∀ T1 T2,
  eqb_ty T1 T2 = true → T1 = T2.
Proof with auto.
  intros T1. induction T1; intros T2 Hbeq; destruct T2; inversion Hbeq.
  -
    reflexivity.
  -
    rewrite andb_true_iff in H0. inversion H0 as [Hbeq1 Hbeq2].
    apply IHT1_1 in Hbeq1. apply IHT1_2 in Hbeq2. subst... Qed.
End STLCTypes.
```

14.2 The Typechecker

The typechecker works by walking over the structure of the given term, returning either Some T or None. Each time we make a recursive call to find out the types of the subterms, we need to pattern-match on the results to make sure that they are not None. Also, in the app case, we use pattern matching to extract the left- and right-hand sides of the function's arrow type (and fail if the type of the function is not Arrow *T11* *T12* for some *T11* and *T12*).

```
Module FirstTry.
Import STLCTypes.

Fixpoint type_check (Gamma : context) (t : tm) : option ty :=
  match t with
  | var x ⇒
      Gamma x
  | abs x T11 t12 ⇒
      match type_check (update Gamma x T11) t12 with
      | Some T12 ⇒ Some (Arrow T11 T12)
      | _ ⇒ None
      end
  | app t1 t2 ⇒
```

```
      match type_check Gamma t1, type_check Gamma t2 with
      | Some (Arrow T11 T12),Some T2 ⇒
          if eqb_ty T11 T2 then Some T12 else None
      | _,_ ⇒ None
      end
  | tru ⇒
      Some Bool
  | fls ⇒
      Some Bool
  | test guard t f ⇒
      match type_check Gamma guard with
      | Some Bool ⇒
          match type_check Gamma t, type_check Gamma f with
          | Some T1, Some T2 ⇒
              if eqb_ty T1 T2 then Some T1 else None
          | _,_ ⇒ None
          end
      | _ ⇒ None
      end
  end.
End FirstTry.
```

14.3 Digression: Improving the Notation

Before we consider the properties of this algorithm, let's write it out again in a cleaner way, using "monadic" notations in the style of Haskell to streamline the plumbing of options. First, we define a notation for composing two potentially failing (i.e., option-returning) computations:

```
Notation " x <- e1 ;; e2" := (match e1 with
                               | Some x ⇒ e2
                               | None ⇒ None
                               end)
         (right associativity, at level 60).
```

Second, we define `return` and `fail` as synonyms for Some and None:

```
Notation " 'return' e "
  := (Some e) (at level 60).

Notation " 'fail' "
  := None.

Module STLCChecker.
Import STLCTypes.
```

Now we can write the same type-checking function in a more imperative-looking style using these notations.

```
Fixpoint type_check (Gamma : context) (t : tm) : option ty :=
  match t with
  | var x ⇒
      match Gamma x with
      | Some T ⇒ return T
      | None ⇒ fail
      end
  | abs x T11 t12 ⇒
      T12 ← type_check (update Gamma x T11) t12 ;;
      return (Arrow T11 T12)
  | app t1 t2 ⇒
      T1 ← type_check Gamma t1 ;;
      T2 ← type_check Gamma t2 ;;
      match T1 with
      | Arrow T11 T12 ⇒
          if eqb_ty T11 T2 then return T12 else fail
      | _ ⇒ fail
      end
  | tru ⇒
      return Bool
  | fls ⇒
      return Bool
  | test guard t1 t2 ⇒
      Tguard ← type_check Gamma guard ;;
      T1 ← type_check Gamma t1 ;;
      T2 ← type_check Gamma t2 ;;
      match Tguard with
      | Bool ⇒
          if eqb_ty T1 T2 then return T1 else fail
      | _ ⇒ fail
      end
  end.
```

14.4 Properties

To verify that the typechecking algorithm is correct, we show that it is *sound* and *complete* for the original **has_type** relation – that is, type_check and **has_type** define the same partial function.

```
Theorem type_checking_sound : ∀ Gamma t T,
```

```
  type_check Gamma t = Some T → has_type Gamma t T.
Proof with eauto.
  intros Gamma t. generalize dependent Gamma.
  induction t; intros Gamma T Htc; inversion Htc.
  - rename s into x. destruct (Gamma x) eqn:H.
    rename t into T'. inversion H0. subst. eauto. solve_by_invert.
  -
    remember (type_check Gamma t1) as TO1.
    destruct TO1 as [T1|]; try solve_by_invert;
    destruct T1 as [|T11 T12]; try solve_by_invert;
    remember (type_check Gamma t2) as TO2;
    destruct TO2 as [T2|]; try solve_by_invert.
    destruct (eqb_ty T11 T2) eqn: Heqb.
    apply eqb_ty__eq in Heqb.
    inversion H0; subst...
    inversion H0.
  -
    rename s into x. rename t into T1.
    remember (update Gamma x T1) as G'.
    remember (type_check G' t0) as TO2.
    destruct TO2; try solve_by_invert.
    inversion H0; subst...
  - eauto.
  - eauto.
  -
    remember (type_check Gamma t1) as TOc.
    remember (type_check Gamma t2) as TO1.
    remember (type_check Gamma t3) as TO2.
    destruct TOc as [Tc|]; try solve_by_invert.
    destruct Tc; try solve_by_invert;
    destruct TO1 as [T1|]; try solve_by_invert;
    destruct TO2 as [T2|]; try solve_by_invert.
    destruct (eqb_ty T1 T2) eqn:Heqb;
    try solve_by_invert.
    apply eqb_ty__eq in Heqb.
    inversion H0. subst. subst...
Qed.

Theorem type_checking_complete : ∀ Gamma t T,
  has_type Gamma t T → type_check Gamma t = Some T.
Proof with auto.
  intros Gamma t T Hty.
  induction Hty; simpl.
```

```
  - destruct (Gamma x0) eqn:H0; assumption.
  - rewrite IHHty...
  -
    rewrite IHHty1. rewrite IHHty2.
    rewrite (eqb_ty_refl T11)...
  - eauto.
  - eauto.
  - rewrite IHHty1. rewrite IHHty2.
    rewrite IHHty3. rewrite (eqb_ty_refl T)...
Qed.
End STLCChecker.
```

14.5 Exercises

Exercise: 5 stars, standard (typechecker_extensions) In this exercise we'll extend the typechecker to deal with the extended features discussed in chapter MoreStlc. Your job is to fill in the omitted cases in the following.

```
Module TypecheckerExtensions.
Definition manual_grade_for_type_checking_sound : option (nat×string) := None.
Definition manual_grade_for_type_checking_complete : option (nat×string) := None.
Import MoreStlc.
Import STLCExtended.

Fixpoint eqb_ty (T1 T2 : ty) : bool :=
  match T1,T2 with
  | Nat, Nat ⇒
      true
  | Unit, Unit ⇒
      true
  | Arrow T11 T12, Arrow T21 T22 ⇒
      andb (eqb_ty T11 T21) (eqb_ty T12 T22)
  | Prod T11 T12, Prod T21 T22 ⇒
      andb (eqb_ty T11 T21) (eqb_ty T12 T22)
  | Sum T11 T12, Sum T21 T22 ⇒
      andb (eqb_ty T11 T21) (eqb_ty T12 T22)
  | List T11, List T21 ⇒
      eqb_ty T11 T21
  | _,_ ⇒
      false
  end.

Lemma eqb_ty_refl : ∀ T1,
  eqb_ty T1 T1 = true.
```

```
Proof.
  intros T1.
  induction T1; simpl;
    try reflexivity;
    try (rewrite IHT1_1; rewrite IHT1_2; reflexivity);
    try (rewrite IHT1; reflexivity). Qed.

Lemma eqb_ty__eq : ∀ T1 T2,
  eqb_ty T1 T2 = true → T1 = T2.
Proof.
  intros T1.
  induction T1; intros T2 Hbeq; destruct T2; inversion Hbeq;
    try reflexivity;
    try (rewrite andb_true_iff in H0; inversion H0 as [Hbeq1 Hbeq2];
         apply IHT1_1 in Hbeq1; apply IHT1_2 in Hbeq2; subst; auto);
    try (apply IHT1 in Hbeq; subst; auto).
 Qed.

Fixpoint type_check (Gamma : context) (t : tm) : option ty :=
  match t with
  | var x ⇒
      match Gamma x with
      | Some T ⇒ return T
      | None ⇒ fail
      end
  | abs x1 T1 t2 ⇒
      T2 ← type_check (update Gamma x1 T1) t2 ;;
      return (Arrow T1 T2)
  | app t1 t2 ⇒
      T1 ← type_check Gamma t1 ;;
      T2 ← type_check Gamma t2 ;;
      match T1 with
      | Arrow T11 T12 ⇒
          if eqb_ty T11 T2 then return T12 else fail
      | _ ⇒ fail
      end
  | const _ ⇒
      return Nat
  | scc t1 ⇒
      T1 ← type_check Gamma t1 ;;
      match T1 with
      | Nat ⇒ return Nat
      | _ ⇒ fail
      end
```

```
| prd t1 ⇒
    T1 ← type_check Gamma t1 ;;
    match T1 with
    | Nat ⇒ return Nat
    | _ ⇒ fail
    end
| mlt t1 t2 ⇒
    T1 ← type_check Gamma t1 ;;
    T2 ← type_check Gamma t2 ;;
    match T1, T2 with
    | Nat, Nat ⇒ return Nat
    | _,_ ⇒ fail
    end
| test0 guard t f ⇒
    Tguard ← type_check Gamma guard ;;
    T1 ← type_check Gamma t ;;
    T2 ← type_check Gamma f ;;
    match Tguard with
    | Nat ⇒ if eqb_ty T1 T2 then return T1 else fail
    | _ ⇒ fail
    end
```

```
| tlcase t0 t1 x21 x22 t2 ⇒
    match type_check Gamma t0 with
    | Some (List T) ⇒
        match type_check Gamma t1,
              type_check (update (update Gamma x22 (List T)) x21 T) t2 with
        | Some T1', Some T2' ⇒
            if eqb_ty T1' T2' then Some T1' else None
        | _,_ ⇒ None
        end
    | _ ⇒ None
    end
```

```
  | _ ⇒ None
  end.
```

Just for fun, we'll do the soundness proof with just a bit more automation than above, using these "mega-tactics":

```
Ltac invert_typecheck Gamma t T :=
  remember (type_check Gamma t) as TO;
  destruct TO as [T|];
  try solve_by_invert; try (inversion H0; eauto); try (subst; eauto).

Ltac analyze T T1 T2 :=
  destruct T as [T1 T2| |T1 T2|T1| |T1 T2]; try solve_by_invert.

Ltac fully_invert_typecheck Gamma t T T1 T2 :=
  let TX := fresh T in
  remember (type_check Gamma t) as TO;
  destruct TO as [TX|]; try solve_by_invert;
  destruct TX as [T1 T2| |T1 T2|T1| |T1 T2];
  try solve_by_invert; try (inversion H0; eauto); try (subst; eauto).

Ltac case_equality S T :=
  destruct (eqb_ty S T) eqn: Heqb;
  inversion H0; apply eqb_ty__eq in Heqb; subst; subst; eauto.

Theorem type_checking_sound : ∀ Gamma t T,
  type_check Gamma t = Some T → has_type Gamma t T.
Proof with eauto.
  intros Gamma t. generalize dependent Gamma.
  induction t; intros Gamma T Htc; inversion Htc.
  - rename s into x. destruct (Gamma x) eqn:H.
    rename t into T'. inversion H0. subst. eauto. solve_by_invert.
  -
    invert_typecheck Gamma t1 T1.
    invert_typecheck Gamma t2 T2.
    analyze T1 T11 T12.
    case_equality T11 T2.
  -
    rename s into x. rename t into T1.
    remember (update Gamma x T1) as Gamma'.
    invert_typecheck Gamma' t0 T0.
  - eauto.
```

```
  -
    rename t into t1.
    fully_invert_typecheck Gamma t1 T1 T11 T12.
  -
    rename t into t1.
    fully_invert_typecheck Gamma t1 T1 T11 T12.
  -
    invert_typecheck Gamma t1 T1.
    invert_typecheck Gamma t2 T2.
    analyze T1 T11 T12; analyze T2 T21 T22.
    inversion H0. subst. eauto.
  -
    invert_typecheck Gamma t1 T1.
    invert_typecheck Gamma t2 T2.
    invert_typecheck Gamma t3 T3.
    destruct T1; try solve_by_invert.
    case_equality T2 T3.
  -
    rename s into x31. rename s0 into x32.
    fully_invert_typecheck Gamma t1 T1 T11 T12.
    invert_typecheck Gamma t2 T2.
    remember (update (update Gamma x32 (List T11)) x31 T11) as Gamma'2.
    invert_typecheck Gamma'2 t3 T3.
    case_equality T2 T3.
Qed.

Theorem type_checking_complete : ∀ Gamma t T,
  has_type Gamma t T → type_check Gamma t = Some T.
Proof.
  intros Gamma t T Hty.
  induction Hty; simpl;
    try (rewrite IHHty);
    try (rewrite IHHty1);
    try (rewrite IHHty2);
    try (rewrite IHHty3);
    try (rewrite (eqb_ty_refl T));
    try (rewrite (eqb_ty_refl T1));
    try (rewrite (eqb_ty_refl T2));
    eauto.
  - destruct (Gamma x); try solve_by_invert. eauto.
  Admitted. End TypecheckerExtensions.
```
□

Exercise: 5 stars, standard, optional (stlc_step_function) Above, we showed how to write a typechecking function and prove it sound and complete for the typing relation. Do the same for the operational semantics – i.e., write a function stepf of type **tm** → **option tm** and prove that it is sound and complete with respect to **step** from chapter MoreStlc.

```
Module StepFunction.
Import MoreStlc.
Import STLCExtended.

Fixpoint stepf (t : tm) : option tm
  . Admitted.

Theorem sound_stepf : ∀ t t',
    stepf t = Some t' → t -> t'.
Proof. Admitted.

Theorem complete_stepf : ∀ t t',
    t -> t' → stepf t = Some t'.
Proof. Admitted.

End StepFunction.
```

□

Exercise: 5 stars, standard, optional (stlc_impl) Using the Imp parser described in the *ImpParser* chapter of *Logical Foundations* as a guide, build a parser for extended STLC programs. Combine it with the typechecking and stepping functions from the above exercises to yield a complete typechecker and interpreter for this language.

```
Module StlcImpl.
Import StepFunction.

End StlcImpl.
```

□

Chapter 15

Records: Adding Records to STLC

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Strings.String.
From PLF Require Import Maps.
From PLF Require Import Imp.
From PLF Require Import Smallstep.
From PLF Require Import Stlc.
```

15.1 Adding Records

We saw in chapter MoreStlc how records can be treated as just syntactic sugar for nested uses of products. This is OK for simple examples, but the encoding is informal (in reality, if we actually treated records this way, it would be carried out in the parser, which we are eliding here), and anyway it is not very efficient. So it is also interesting to see how records can be treated as first-class citizens of the language. This chapter shows how.

Recall the informal definitions we gave before:

Syntax:

t ::= Terms: | {i1=t1, ..., in=tn} record | t.i projection | ...

v ::= Values: | {i1=v1, ..., in=vn} record value | ...

T ::= Types: | {i1:T1, ..., in:Tn} record type | ...

Reduction:

ti ==> ti'

(ST_Rcd) {i1=v1, ..., im=vm, in=tn, ...} ==> {i1=v1, ..., im=vm, in=tn', ...}

t1 ==> t1'

(ST_Proj1) t1.i ==> t1'.i

(ST_ProjRcd) {..., i=vi, ...}.i ==> vi

Typing:

Gamma |- t1 : T1 ... Gamma |- tn : Tn

(T_Rcd) Gamma |- {i1=t1, ..., in=tn} : {i1:T1, ..., in:Tn}
Gamma |- t : {..., i:Ti, ...}

(T_Proj) Gamma |- t.i : Ti

15.2 Formalizing Records

```
Module STLCExtendedRecords.
```

Syntax and Operational Semantics

The most obvious way to formalize the syntax of record types would be this:

```
Module FirstTry.

Definition alist (X : Type) := list (string × X).

Inductive ty : Type :=
  | Base : string → ty
  | Arrow : ty → ty → ty
  | TRcd : (alist ty) → ty.
```

Unfortunately, we encounter here a limitation in Coq: this type does not automatically give us the induction principle we expect: the induction hypothesis in the TRcd case doesn't give us any information about the **ty** elements of the list, making it useless for the proofs we want to do.

```
End FirstTry.
```

It is possible to get a better induction principle out of Coq, but the details of how this is done are not very pretty, and the principle we obtain is not as intuitive to use as the ones Coq generates automatically for simple `Inductive` definitions.

Fortunately, there is a different way of formalizing records that is, in some ways, even simpler and more natural: instead of using the standard Coq **list** type, we can essentially incorporate its constructors ("nil" and "cons") in the syntax of our types.

```
Inductive ty : Type :=
  | Base : string → ty
  | Arrow : ty → ty → ty
  | RNil : ty
  | RCons : string → ty → ty → ty.
```

Similarly, at the level of terms, we have constructors trnil, for the empty record, and rcons, which adds a single field to the front of a list of fields.

```
Inductive tm : Type :=
```

```
  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm

  | rproj : tm → string → tm
  | trnil : tm
  | rcons : string → tm → tm → tm.
```

Some examples... `Open Scope` *string_scope*.

```
Notation a := "a".
Notation f := "f".
Notation g := "g".
Notation l := "l".
Notation A := (Base "A").
Notation B := (Base "B").
Notation k := "k".
Notation i1 := "i1".
Notation i2 := "i2".
```

{ i1:A }

{ i1:A→B, i2:A }

Well-Formedness

One issue with generalizing the abstract syntax for records from lists to the nil/cons presentation is that it introduces the possibility of writing strange types like this...

```
Definition weird_type := RCons X A B.
```

where the "tail" of a record type is not actually a record type!

We'll structure our typing judgement so that no ill-formed types like weird_type are ever assigned to terms. To support this, we define predicates **record_ty** and **record_tm**, which identify record types and terms, and **well_formed_ty** which rules out the ill-formed types.

First, a type is a record type if it is built with just RNil and RCons at the outermost level.

```
Inductive record_ty : ty → Prop :=
  | RTnil :
        record_ty RNil
  | RTcons : ∀ i T1 T2,
        record_ty (RCons i T1 T2).
```

With this, we can define well-formed types.

```
Inductive well_formed_ty : ty → Prop :=
  | wfBase : ∀ i,
        well_formed_ty (Base i)
```

```
  | wfArrow : ∀ T1 T2,
        well_formed_ty T1 →
        well_formed_ty T2 →
        well_formed_ty (Arrow T1 T2)
  | wfRNil :
        well_formed_ty RNil
  | wfRCons : ∀ i T1 T2,
        well_formed_ty T1 →
        well_formed_ty T2 →
        record_ty T2 →
        well_formed_ty (RCons i T1 T2).
Hint Constructors record_ty well_formed_ty.
```

Note that **record_ty** is not recursive – it just checks the outermost constructor. The **well_formed_ty** property, on the other hand, verifies that the whole type is well formed in the sense that the tail of every record (the second argument to RCons) is a record.

Of course, we should also be concerned about ill-formed terms, not just types; but type-checking can rule those out without the help of an extra *well_formed_tm* definition because it already examines the structure of terms. All we need is an analog of **record_ty** saying that a term is a record term if it is built with trnil and rcons.

```
Inductive record_tm : tm → Prop :=
  | rtnil :
        record_tm trnil
  | rtcons : ∀ i t1 t2,
        record_tm (rcons i t1 t2).
Hint Constructors record_tm.
```

Substitution

Substitution extends easily.

```
Fixpoint subst (x:string) (s:tm) (t:tm) : tm :=
  match t with
  | var y ⇒ if eqb_string x y then s else t
  | abs y T t1 ⇒ abs y T
                        (if eqb_string x y then t1 else (subst x s t1))
  | app t1 t2 ⇒ app (subst x s t1) (subst x s t2)
  | rproj t1 i ⇒ rproj (subst x s t1) i
  | trnil ⇒ trnil
  | rcons i t1 tr1 ⇒ rcons i (subst x s t1) (subst x s tr1)
  end.
Notation "'[' x ':=' s ']' t" := (subst x s t) (at level 20).
```

Reduction

A record is a value if all of its fields are.

```
Inductive value : tm → Prop :=
  | v_abs : ∀ x T11 t12,
      value (abs x T11 t12)
  | v_rnil : value trnil
  | v_rcons : ∀ i v1 vr,
      value v1 →
      value vr →
      value (rcons i v1 vr).
Hint Constructors value.
```

To define reduction, we'll need a utility function for extracting one field from record term:

```
Fixpoint tlookup (i:string) (tr:tm) : option tm :=
  match tr with
  | rcons i' t tr' ⇒ if eqb_string i i' then Some t else tlookup i tr'
  | _ ⇒ None
  end.
```

The **step** function uses this term-level lookup function in the projection rule.

```
Reserved Notation "t1 '-->' t2" (at level 40).
Inductive step : tm → tm → Prop :=
  | ST_AppAbs : ∀ x T11 t12 v2,
        value v2 →
        (app (abs x T11 t12) v2) -> ([x:=v2]t12)
  | ST_App1 : ∀ t1 t1' t2,
        t1 -> t1' →
        (app t1 t2) -> (app t1' t2)
  | ST_App2 : ∀ v1 t2 t2',
        value v1 →
        t2 -> t2' →
        (app v1 t2) -> (app v1 t2')
  | ST_Proj1 : ∀ t1 t1' i,
      t1 -> t1' →
      (rproj t1 i) -> (rproj t1' i)
  | ST_ProjRcd : ∀ tr i vi,
      value tr →
      tlookup i tr = Some vi →
      (rproj tr i) -> vi
  | ST_Rcd_Head : ∀ i t1 t1' tr2,
      t1 -> t1' →
```

```
        (rcons i t1 tr2) -> (rcons i t1' tr2)
  | ST_Rcd_Tail : ∀ i v1 tr2 tr2',
        value v1 →
        tr2 -> tr2' →
        (rcons i v1 tr2) -> (rcons i v1 tr2')

where "t1 '->' t2" := (step t1 t2).

Notation multistep := (multi step).
Notation "t1 '->*' t2" := (multistep t1 t2) (at level 40).

Hint Constructors step.
```

Typing

Next we define the typing rules. These are nearly direct transcriptions of the inference rules shown above: the only significant difference is the use of **well_formed_ty**. In the informal presentation we used a grammar that only allowed well-formed record types, so we didn't have to add a separate check.

One sanity condition that we'd like to maintain is that, whenever **has_type** *Gamma t* T holds, will also be the case that **well_formed_ty** T, so that **has_type** never assigns ill-formed types to terms. In fact, we prove this theorem below. However, we don't want to clutter the definition of **has_type** with unnecessary uses of **well_formed_ty**. Instead, we place **well_formed_ty** checks only where needed: where an inductive call to **has_type** won't already be checking the well-formedness of a type. For example, we check **well_formed_ty** T in the T_Var case, because there is no inductive **has_type** call that would enforce this. Similarly, in the T_Abs case, we require a proof of **well_formed_ty** *T11* because the inductive call to **has_type** only guarantees that *T12* is well-formed.

```
Fixpoint Tlookup (i:string) (Tr:ty) : option ty :=
  match Tr with
  | RCons i' T Tr' ⇒
      if eqb_string i i' then Some T else Tlookup i Tr'
  | _ ⇒ None
  end.

Definition context := partial_map ty.

Reserved Notation "Gamma '|-' t '\in' T" (at level 40).

Inductive has_type : context → tm → ty → Prop :=
  | T_Var : ∀ Gamma x T,
      Gamma x = Some T →
      well_formed_ty T →
      Gamma ⊢ (var x) \in T
  | T_Abs : ∀ Gamma x T11 T12 t12,
      well_formed_ty T11 →
```

```
      (update Gamma x T11) ⊢ t12 \in T12 →
      Gamma ⊢ (abs x T11 t12) \in (Arrow T11 T12)
  | T_App : ∀ T1 T2 Gamma t1 t2,
      Gamma ⊢ t1 \in (Arrow T1 T2) →
      Gamma ⊢ t2 \in T1 →
      Gamma ⊢ (app t1 t2) \in T2

  | T_Proj : ∀ Gamma i t Ti Tr,
      Gamma ⊢ t \in Tr →
      Tlookup i Tr = Some Ti →
      Gamma ⊢ (rproj t i) \in Ti
  | T_RNil : ∀ Gamma,
      Gamma ⊢ trnil \in RNil
  | T_RCons : ∀ Gamma i t T tr Tr,
      Gamma ⊢ t \in T →
      Gamma ⊢ tr \in Tr →
      record_ty Tr →
      record_tm tr →
      Gamma ⊢ (rcons i t tr) \in (RCons i T Tr)

where "Gamma '|-' t '\in' T" := (has_type Gamma t T).

Hint Constructors has_type.
```

15.2.1 Examples

Exercise: 2 stars, standard (examples) Finish the proofs below. Feel free to use Coq's automation features in this proof. However, if you are not confident about how the type system works, you may want to carry out the proofs first using the basic features (`apply` instead of `eapply`, in particular) and then perhaps compress it using automation. Before starting to prove anything, make sure you understand what it is saying.

```
Lemma typing_example_2 :
  empty ⊢
    (app (abs a (RCons i1 (Arrow A A)
                      (RCons i2 (Arrow B B)
                       RNil))
              (rproj (var a) i2))
            (rcons i1 (abs a A (var a))
            (rcons i2 (abs a B (var a))
             trnil))) \in
    (Arrow B B).
Proof.
   Admitted.
```

```
Example typing_nonexample :
  ¬ ∃ T,
      (update empty a (RCons i2 (Arrow A A)
                                         RNil)) ⊢
                   (rcons i1 (abs a B (var a)) (var a)) \in
                   T.
Proof.
    Admitted.

Example typing_nonexample_2 : ∀ y,
  ¬ ∃ T,
    (update empty y A) ⊢
             (app (abs a (RCons i1 A RNil)
                         (rproj (var a) i1))
                       (rcons i1 (var y) (rcons i2 (var y) trnil))) \in
             T.
Proof.
    Admitted.
```

15.2.2 Properties of Typing

The proofs of progress and preservation for this system are essentially the same as for the pure simply typed lambda-calculus, but we need to add some technical lemmas involving records.

Well-Formedness

```
Lemma wf_rcd_lookup : ∀ i T Ti,
  well_formed_ty T →
  Tlookup i T = Some Ti →
  well_formed_ty Ti.
Proof with eauto.
  intros i T.
  induction T; intros; try solve_by_invert.
  -
    inversion H. subst. unfold Tlookup in H0.
    destruct (eqb_string i s)...
    inversion H0. subst... Qed.

Lemma step_preserves_record_tm : ∀ tr tr',
  record_tm tr →
  tr -> tr' →
  record_tm tr'.
Proof.
```

```
  intros tr tr' Hrt Hstp.
  inversion Hrt; subst; inversion Hstp; subst; auto.
Qed.

Lemma has_type__wf : ∀ Gamma t T,
  Gamma ⊢ t \in T → well_formed_ty T.
Proof with eauto.
  intros Gamma t T Htyp.
  induction Htyp...
  -
    inversion IHHtyp1...
  -
    eapply wf_rcd_lookup...
Qed.
```

Field Lookup

Lemma: If empty ⊢ v : T and Tlookup i T returns Some Ti, then tlookup i v returns Some ti for some term ti such that empty ⊢ ti \in Ti.

Proof: By induction on the typing derivation $Htyp$. Since Tlookup i T = Some Ti, T must be a record type, this and the fact that v is a value eliminate most cases by inspection, leaving only the T_RCons case.

If the last step in the typing derivation is by T_RCons, then t = rcons $i0$ t tr and T = RCons $i0$ T Tr for some $i0$, t, tr, T and Tr.

This leaves two possiblities to consider - either $i0$ = i or not.

- If i = $i0$, then since Tlookup i (RCons $i0$ T Tr) = Some Ti we have T = Ti. It follows that t itself satisfies the theorem.
- On the other hand, suppose i ≠ $i0$. Then

  Tlookup i T = Tlookup i Tr

  and

  tlookup i t = tlookup i tr,

  so the result follows from the induction hypothesis. □

Here is the formal statement:

```
Lemma lookup_field_in_value : ∀ v T i Ti,
  value v →
  empty ⊢ v \in T →
  Tlookup i T = Some Ti →
  ∃ ti, tlookup i v = Some ti ∧ empty ⊢ ti \in Ti.
Proof with eauto.
  intros v T i Ti Hval Htyp Hget.
```

```
  remember (@empty ty) as Gamma.
  induction Htyp; subst; try solve_by_invert...
  -
    simpl in Hget. simpl. destruct (eqb_string i i0).
    +
      simpl. inversion Hget. subst.
      ∃ t...
    +
      destruct IHHtyp2 as [vi [Hgeti Htypi]]...
      inversion Hval... Qed.
```

Progress

```
Theorem progress : ∀ t T,
     empty ⊢ t \in T →
     value t ∨ ∃ t', t -> t'.
Proof with eauto.
  intros t T Ht.
  remember (@empty ty) as Gamma.
  generalize dependent HeqGamma.
  induction Ht; intros HeqGamma; subst.
  -

    inversion H.
  -

    left...
  -

    right.
    destruct IHHt1; subst...
    +
      destruct IHHt2; subst...
      ×

        inversion H; subst; try solve_by_invert.
        ∃ ([x:=t2]t12)...
      ×

        destruct H0 as [t2' Hstp]. ∃ (app t1 t2')...
    +

      destruct H as [t1' Hstp]. ∃ (app t1' t2)...
```

-

right. destruct *IHHt*...

+

destruct (lookup_field_in_value _ _ _ _ *H0* *Ht* *H*)
as [*ti* [*Hlkup* _]].
∃ *ti*...

+

destruct *H0* as [*t'* *Hstp*]. ∃ (rproj *t'* *i*)...

-

left...

-

destruct *IHHt1*...

+

destruct *IHHt2*; try reflexivity.

×

left...

×

right. destruct *H2* as [*tr'* *Hstp*].
∃ (rcons *i* *t* *tr'*)...

+

right. destruct *H1* as [*t'* *Hstp*].
∃ (rcons *i* *t'* *tr*)... Qed.

Context Invariance

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
      appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
      appears_free_in x t1 → appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
      appears_free_in x t2 → appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
        y ≠ x →
        appears_free_in x t12 →
```

```
        appears_free_in x (abs y T11 t12)
  | afi_proj : ∀ x t i,
      appears_free_in x t →
      appears_free_in x (rproj t i)
  | afi_rhead : ∀ x i ti tr,
       appears_free_in x ti →
       appears_free_in x (rcons i ti tr)
  | afi_rtail : ∀ x i ti tr,
       appears_free_in x tr →
       appears_free_in x (rcons i ti tr).

Hint Constructors appears_free_in.

Lemma context_invariance : ∀ Gamma Gamma' t S,
     Gamma ⊢ t \in S →
     (∀ x, appears_free_in x t → Gamma x = Gamma' x) →
     Gamma' ⊢ t \in S.
Proof with eauto.
  intros. generalize dependent Gamma'.
  induction H;
    intros Gamma' Heqv...
  -
    apply T_Var... rewrite ← Heqv...
  -
    apply T_Abs... apply IHhas_type. intros y Hafi.
    unfold update, t_update. destruct (eqb_stringP x y)...
  -
    apply T_App with T1...
  -
    apply T_RCons... Qed.

Lemma free_in_context : ∀ x t T Gamma,
    appears_free_in x t →
    Gamma ⊢ t \in T →
    ∃ T', Gamma x = Some T'.
Proof with eauto.
  intros x t T Gamma Hafi Htyp.
  induction Htyp; inversion Hafi; subst...
  -
    destruct IHHtyp as [T' Hctx]... ∃ T'.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx...
Qed.
```

Preservation

```
Lemma substitution_preserves_typing : ∀ Gamma x U v t S,
     (update Gamma x U) ⊢ t \in S →
     empty ⊢ v \in U →
     Gamma ⊢ ([x:=v]t) \in S.
Proof with eauto.
  intros Gamma x U v t S Htypt Htypv.
  generalize dependent Gamma. generalize dependent S.
  induction t;
    intros S Gamma Htypt; simpl; inversion Htypt; subst...
  -
    simpl. rename s into y.
    unfold update, t_update in H0.
    destruct (eqb_stringP x y) as [Hxy|Hxy].
    +

      subst.
      inversion H0; subst. clear H0.
      eapply context_invariance...
      intros x Hcontra.
      destruct (free_in_context _ _ S empty Hcontra)
        as [T' HT']...
      inversion HT'.
    +

      apply T_Var...
  -
    rename s into y. rename t into T11.
    apply T_Abs...
    destruct (eqb_stringP x y) as [Hxy|Hxy].
    +

      eapply context_invariance...
      subst.
      intros x Hafi. unfold update, t_update.
      destruct (eqb_string y x)...
    +

      apply IHt. eapply context_invariance...
      intros z Hafi. unfold update, t_update.
      destruct (eqb_stringP y z)...
      subst. rewrite false_eqb_string...
```

```
-
    apply T_RCons... inversion H7; subst; simpl...
Qed.

Theorem preservation : ∀ t t' T,
     empty ⊢ t \in T →
     t -> t' →
     empty ⊢ t' \in T.
Proof with eauto.
  intros t t' T HT.
  remember (@empty ty) as Gamma. generalize dependent HeqGamma.
  generalize dependent t'.
  induction HT;
    intros t' HeqGamma HE; subst; inversion HE; subst...
  -

    inversion HE; subst...
    +

      apply substitution_preserves_typing with T1...
      inversion HT1...
  -

    destruct (lookup_field_in_value _ _ _ _ H2 HT H)
      as [vi [Hget Htyp]].
    rewrite H4 in Hget. inversion Hget. subst...
  -

    apply T_RCons... eapply step_preserves_record_tm...
Qed.
```

☐

End STLCExtendedRecords.

Chapter 16

References: Typing Mutable References

Up to this point, we have considered a variety of *pure* language features, including functional abstraction, basic types such as numbers and booleans, and structured types such as records and variants. These features form the backbone of most programming languages – including purely functional languages such as Haskell and "mostly functional" languages such as ML, as well as imperative languages such as C and object-oriented languages such as Java, C#, and Scala.

However, most practical languages also include various *impure* features that cannot be described in the simple semantic framework we have used so far. In particular, besides just yielding results, computation in these languages may assign to mutable variables (reference cells, arrays, mutable record fields, etc.); perform input and output to files, displays, or network connections; make non-local transfers of control via exceptions, jumps, or continuations; engage in inter-process synchronization and communication; and so on. In the literature on programming languages, such "side effects" of computation are collectively referred to as *computational effects*.

In this chapter, we'll see how one sort of computational effect – mutable references – can be added to the calculi we have studied. The main extension will be dealing explicitly with a *store* (or *heap*) and *pointers* that name store locations. This extension is fairly straightforward to define; the most interesting part is the refinement we need to make to the statement of the type preservation theorem.

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Strings.String.
From Coq Require Import Arith.Arith.
From Coq Require Import omega.Omega.
From PLF Require Import Maps.
From PLF Require Import Smallstep.
From Coq Require Import Lists.List.
Import ListNotations.
```

16.1 Definitions

Pretty much every programming language provides some form of assignment operation that changes the contents of a previously allocated piece of storage. (Coq's internal language Gallina is a rare exception!)

In some languages – notably ML and its relatives – the mechanisms for name-binding and those for assignment are kept separate. We can have a variable x whose *value* is the number 5, or we can have a variable y whose value is a *reference* (or *pointer*) to a mutable cell whose current contents is 5. These are different things, and the difference is visible to the programmer. We can add x to another number, but not assign to it. We can use y to assign a new value to the cell that it points to (by writing y:=84), but we cannot use y directly as an argument to an operation like +. Instead, we must explicitly *dereference* it, writing !y to obtain its current contents.

In most other languages – in particular, in all members of the C family, including Java – *every* variable name refers to a mutable cell, and the operation of dereferencing a variable to obtain its current contents is implicit.

For purposes of formal study, it is useful to keep these mechanisms separate. The development in this chapter will closely follow ML's model. Applying the lessons learned here to C-like languages is a straightforward matter of collapsing some distinctions and rendering some operations such as dereferencing implicit instead of explicit.

16.2 Syntax

In this chapter, we study adding mutable references to the simply-typed lambda calculus with natural numbers.

`Module` STLCREF.

The basic operations on references are *allocation*, *dereferencing*, and *assignment*.

- To allocate a reference, we use the ref operator, providing an initial value for the new cell. For example, ref 5 creates a new cell containing the value 5, and reduces to a reference to that cell.

- To read the current value of this cell, we use the dereferencing operator !; for example, !(ref 5) reduces to 5.

- To change the value stored in a cell, we use the assignment operator. If r is a reference, r := 7 will store the value 7 in the cell referenced by r.

Types

We start with the simply typed lambda calculus over the natural numbers. Besides the base natural number type and arrow types, we need to add two more types to deal with references.

First, we need the *unit type*, which we will use as the result type of an assignment operation. We then add *reference types*.

If T is a type, then Ref T is the type of references to cells holding values of type T.

T ::= Nat | Unit | T -> T | Ref T

```
Inductive ty : Type :=
  | Nat : ty
  | Unit : ty
  | Arrow : ty → ty → ty
  | Ref : ty → ty.
```

Terms

Besides variables, abstractions, applications, natural-number-related terms, and unit, we need four more sorts of terms in order to handle mutable references:

t ::= ... Terms | ref t allocation | !t dereference | t := t assignment | l location

```
Inductive tm : Type :=

  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm
  | const : nat → tm
  | scc : tm → tm
  | prd : tm → tm
  | mlt : tm → tm → tm
  | test0 : tm → tm → tm → tm

  | unit : tm
  | ref : tm → tm
  | deref : tm → tm
  | assign : tm → tm → tm
  | loc : nat → tm.
```

Intuitively:

- ref *t* (formally, ref *t*) allocates a new reference cell with the value *t* and reduces to the location of the newly allocated cell;
- !*t* (formally, deref *t*) reduces to the contents of the cell referenced by *t*;
- *t1* := *t2* (formally, assign *t1* *t2*) assigns *t2* to the cell referenced by *t1*; and
- l (formally, loc l) is a reference to the cell at location l. We'll discuss locations later.

In informal examples, we'll also freely use the extensions of the STLC developed in the MoreStlc chapter; however, to keep the proofs small, we won't bother formalizing them again here. (It would be easy to do so, since there are no very interesting interactions between those features and references.)

Typing (Preview)

Informally, the typing rules for allocation, dereferencing, and assignment will look like this:

Gamma |- t1 : T1

(T_Ref) Gamma |- ref t1 : Ref T1

Gamma |- t1 : Ref T11

(T_Deref) Gamma |- !t1 : T11

Gamma |- t1 : Ref T11 Gamma |- t2 : T11

(T_Assign) Gamma |- t1 := t2 : Unit

The rule for locations will require a bit more machinery, and this will motivate some changes to the other rules; we'll come back to this later.

Values and Substitution

Besides abstractions and numbers, we have two new types of values: the unit value, and locations.

```
Inductive value : tm → Prop :=
  | v_abs : ∀ x T t,
      value (abs x T t)
  | v_nat : ∀ n,
      value (const n)
  | v_unit :
      value unit
  | v_loc : ∀ l,
      value (loc l).
Hint Constructors value.
```

Extending substitution to handle the new syntax of terms is straightforward.

```
Fixpoint subst (x:string) (s:tm) (t:tm) : tm :=
  match t with
  | var x' ⇒
      if eqb_string x x' then s else t
  | app t1 t2 ⇒
      app (subst x s t1) (subst x s t2)
  | abs x' T t1 ⇒
      if eqb_string x x' then t else abs x' T (subst x s t1)
  | const n ⇒
```

```
      t
  | scc t1 ⇒
      scc (subst x s t1)
  | prd t1 ⇒
      prd (subst x s t1)
  | mlt t1 t2 ⇒
      mlt (subst x s t1) (subst x s t2)
  | test0 t1 t2 t3 ⇒
      test0 (subst x s t1) (subst x s t2) (subst x s t3)
  | unit ⇒
      t
  | ref t1 ⇒
      ref (subst x s t1)
  | deref t1 ⇒
      deref (subst x s t1)
  | assign t1 t2 ⇒
      assign (subst x s t1) (subst x s t2)
  | loc _ ⇒
      t
  end.
Notation "'[' x ':=' s ']' t" := (subst x s t) (at level 20).
```

16.3 Pragmatics

16.3.1 Side Effects and Sequencing

The fact that we've chosen the result of an assignment expression to be the trivial value unit allows a nice abbreviation for *sequencing*. For example, we can write

r:=succ(!r); !r

as an abbreviation for

(\x:Unit. !r) (r:=succ(!r)).

This has the effect of reducing two expressions in order and returning the value of the second. Restricting the type of the first expression to Unit helps the typechecker to catch some silly errors by permitting us to throw away the first value only if it is really guaranteed to be trivial.

Notice that, if the second expression is also an assignment, then the type of the whole sequence will be Unit, so we can validly place it to the left of another ; to build longer sequences of assignments:

r:=succ(!r); r:=succ(!r); r:=succ(!r); r:=succ(!r); !r

Formally, we introduce sequencing as a *derived form* tseq that expands into an abstraction and an application.

```
Definition tseq t1 t2 :=
  app (abs "x" Unit t2) t1.
```

16.3.2 References and Aliasing

It is important to bear in mind the difference between the *reference* that is bound to some variable r and the *cell* in the store that is pointed to by this reference.

If we make a copy of r, for example by binding its value to another variable s, what gets copied is only the *reference*, not the contents of the cell itself.

For example, after reducing

let r = ref 5 in let s = r in s := 82; (!r)+1

the cell referenced by r will contain the value 82, while the result of the whole expression will be 83. The references r and s are said to be *aliases* for the same cell.

The possibility of aliasing can make programs with references quite tricky to reason about. For example, the expression

r := 5; r := !s

assigns 5 to r and then immediately overwrites it with s's current value; this has exactly the same effect as the single assignment

r := !s

unless we happen to do it in a context where r and s are aliases for the same cell!

16.3.3 Shared State

Of course, aliasing is also a large part of what makes references useful. In particular, it allows us to set up "implicit communication channels" – shared state – between different parts of a program. For example, suppose we define a reference cell and two functions that manipulate its contents:

let c = ref 0 in let incc = _:Unit. (c := succ (!c); !c) in let decc = _:Unit. (c := pred (!c); !c) in ...

Note that, since their argument types are Unit, the arguments to the abstractions in the definitions of *incc* and *decc* are not providing any useful information to the bodies of these functions (using the wildcard _ as the name of the bound variable is a reminder of this). Instead, their purpose of these abstractions is to "slow down" the execution of the function bodies. Since function abstractions are values, the two lets are executed simply by binding these functions to the names *incc* and *decc*, rather than by actually incrementing or decrementing *c*. Later, each call to one of these functions results in its body being executed once and performing the appropriate mutation on *c*. Such functions are often called *thunks*.

In the context of these declarations, calling *incc* results in changes to *c* that can be observed by calling *decc*. For example, if we replace the ... with (*incc* unit; *incc* unit; *decc* unit), the result of the whole program will be 1.

16.3.4 Objects

We can go a step further and write a *function* that creates *c*, *incc*, and *decc*, packages *incc* and *decc* together into a record, and returns this record:

newcounter = _:Unit. let c = ref 0 in let incc = _:Unit. (c := succ (!c); !c) in let decc = _:Unit. (c := pred (!c); !c) in {i=incc, d=decc}

Now, each time we call *newcounter*, we get a new record of functions that share access to the same storage cell *c*. The caller of *newcounter* can't get at this storage cell directly, but can affect it indirectly by calling the two functions. In other words, we've created a simple form of *object*.

let c1 = newcounter unit in let c2 = newcounter unit in // Note that we've allocated two separate storage cells now! let r1 = c1.i unit in let r2 = c2.i unit in r2 // yields 1, not 2!

Exercise: 1 star, standard, optional (store_draw) Draw (on paper) the contents of the store at the point in execution where the first two `let`s have finished and the third one is about to begin.

16.3.5 References to Compound Types

A reference cell need not contain just a number: the primitives we've defined above allow us to create references to values of any type, including functions. For example, we can use references to functions to give an (inefficient) implementation of arrays of numbers, as follows. Write *NatArray* for the type Ref (Nat→Nat).

Recall the equal function from the MoreStlc chapter:

equal = fix (\eq:Nat->Nat->Bool. \m:Nat. \n:Nat. if m=0 then iszero n else if n=0 then false else eq (pred m) (pred n))

To build a new array, we allocate a reference cell and fill it with a function that, when given an index, always returns 0.

newarray = _:Unit. ref (\n:Nat.0)

To look up an element of an array, we simply apply the function to the desired index.

lookup = \a:NatArray. \n:Nat. (!a) n

The interesting part of the encoding is the `update` function. It takes an array, an index, and a new value to be stored at that index, and does its job by creating (and storing in the reference) a new function that, when it is asked for the value at this very index, returns the new value that was given to `update`, while on all other indices it passes the lookup to the function that was previously stored in the reference.

update = \a:NatArray. \m:Nat. \v:Nat. let oldf = !a in a := (\n:Nat. if equal m n then v else oldf n);

References to values containing other references can also be very useful, allowing us to define data structures such as mutable lists and trees.

Exercise: 2 stars, standard, recommended (compact_update) If we defined `update` more compactly like this

update = \a:NatArray. \m:Nat. \v:Nat. a := (\n:Nat. if equal m n then v else (!a) n)
would it behave the same?

```
Definition manual_grade_for_compact_update : option (nat×string) := None.
```

□

16.3.6 Null References

There is one final significant difference between our references and C-style mutable variables: in C-like languages, variables holding pointers into the heap may sometimes have the value *NULL*. Dereferencing such a "null pointer" is an error, and results either in a clean exception (Java and C#) or in arbitrary and possibly insecure behavior (C and relatives like C++). Null pointers cause significant trouble in C-like languages: the fact that any pointer might be null means that any dereference operation in the program can potentially fail.

Even in ML-like languages, there are occasionally situations where we may or may not have a valid pointer in our hands. Fortunately, there is no need to extend the basic mechanisms of references to represent such situations: the sum types introduced in the MoreStlc chapter already give us what we need.

First, we can use sums to build an analog of the **option** types introduced in the *Lists* chapter of *Logical Foundations*. Define *Option* T to be an abbreviation for Unit + T.

Then a "nullable reference to a T" is simply an element of the type *Option* (Ref T).

16.3.7 Garbage Collection

A last issue that we should mention before we move on with formalizing references is storage *de*-allocation. We have not provided any primitives for freeing reference cells when they are no longer needed. Instead, like many modern languages (including ML and Java) we rely on the run-time system to perform *garbage collection*, automatically identifying and reusing cells that can no longer be reached by the program.

This is *not* just a question of taste in language design: it is extremely difficult to achieve type safety in the presence of an explicit deallocation operation. One reason for this is the familiar *dangling reference* problem: we allocate a cell holding a number, save a reference to it in some data structure, use it for a while, then deallocate it and allocate a new cell holding a boolean, possibly reusing the same storage. Now we can have two names for the same storage cell – one with type Ref Nat and the other with type Ref Bool.

Exercise: 2 stars, standard (type_safety_violation) Show how this can lead to a violation of type safety.

```
Definition manual_grade_for_type_safety_violation : option (nat×string) := None.
```

□

16.4 Operational Semantics

16.4.1 Locations

The most subtle aspect of the treatment of references appears when we consider how to formalize their operational behavior. One way to see why is to ask, "What should be the *values* of type Ref T?" The crucial observation that we need to take into account is that reducing a ref operator should *do* something – namely, allocate some storage – and the result of the operation should be a reference to this storage.

What, then, is a reference?

The run-time store in most programming-language implementations is essentially just a big array of bytes. The run-time system keeps track of which parts of this array are currently in use; when we need to allocate a new reference cell, we allocate a large enough segment from the free region of the store (4 bytes for integer cells, 8 bytes for cells storing Floats, etc.), record somewhere that it is being used, and return the index (typically, a 32- or 64-bit integer) of the start of the newly allocated region. These indices are references.

For present purposes, there is no need to be quite so concrete. We can think of the store as an array of *values*, rather than an array of bytes, abstracting away from the different sizes of the run-time representations of different values. A reference, then, is simply an index into the store. (If we like, we can even abstract away from the fact that these indices are numbers, but for purposes of formalization in Coq it is convenient to use numbers.) We use the word *location* instead of *reference* or *pointer* to emphasize this abstract quality.

Treating locations abstractly in this way will prevent us from modeling the *pointer arithmetic* found in low-level languages such as C. This limitation is intentional. While pointer arithmetic is occasionally very useful, especially for implementing low-level services such as garbage collectors, it cannot be tracked by most type systems: knowing that location n in the store contains a *float* doesn't tell us anything useful about the type of location n+4. In C, pointer arithmetic is a notorious source of type-safety violations.

16.4.2 Stores

Recall that, in the small-step operational semantics for IMP, the step relation needed to carry along an auxiliary state in addition to the program being executed. In the same way, once we have added reference cells to the STLC, our step relation must carry along a store to keep track of the contents of reference cells.

We could re-use the same functional representation we used for states in IMP, but for carrying out the proofs in this chapter it is actually more convenient to represent a store simply as a *list* of values. (The reason we didn't use this representation before is that, in IMP, a program could modify any location at any time, so states had to be ready to map *any* variable to a value. However, in the STLC with references, the only way to create a reference cell is with ref *t1*, which puts the value of *t1* in a new reference cell and reduces to the location of the newly created reference cell. When reducing such an expression, we can just add a new reference cell to the end of the list representing the store.)

```
Definition store := list tm.
```

We use store_lookup n *st* to retrieve the value of the reference cell at location n in the store *st*. Note that we must give a default value to nth in case we try looking up an index which is too large. (In fact, we will never actually do this, but proving that we don't will require a bit of work.)

```
Definition store_lookup (n:nat) (st:store) :=
  nth n st unit.
```

To update the store, we use the `replace` function, which replaces the contents of a cell at a particular index.

```
Fixpoint replace {A:Type} (n:nat) (x:A) (l:list A) : list A :=
  match l with
  | nil ⇒ nil
  | h :: t ⇒
    match n with
    | O ⇒ x :: t
    | S n' ⇒ h :: replace n' x t
    end
  end.
```

As might be expected, we will also need some technical lemmas about `replace`; they are straightforward to prove.

```
Lemma replace_nil : ∀ A n (x:A),
  replace n x nil = nil.
Proof.
  destruct n; auto.
Qed.

Lemma length_replace : ∀ A n x (l:list A),
  length (replace n x l) = length l.
Proof with auto.
  intros A n x l. generalize dependent n.
  induction l; intros n.
    destruct n...
    destruct n...
      simpl. rewrite IHl...
Qed.

Lemma lookup_replace_eq : ∀ l t st,
  l < length st →
  store_lookup l (replace l t st) = t.
Proof with auto.
  intros l t st.
  unfold store_lookup.
```

```
  generalize dependent l.
  induction st as [|t' st']; intros l Hlen.
  -
   inversion Hlen.
  -
    destruct l; simpl...
    apply IHst'. simpl in Hlen. omega.
Qed.
```

```
Lemma lookup_replace_neq : ∀ l1 l2 t st,
  l1 ≠ l2 →
  store_lookup l1 (replace l2 t st) = store_lookup l1 st.
Proof with auto.
  unfold store_lookup.
  induction l1 as [|l1']; intros l2 t st Hneq.
  -
    destruct st.
    + rewrite replace_nil...
    + destruct l2... contradict Hneq...
  -
    destruct st as [|t2 st2].
    + destruct l2...
    +
      destruct l2...
      simpl; apply IHl1'...
Qed.
```

16.4.3 Reduction

Next, we need to extend the operational semantics to take stores into account. Since the result of reducing an expression will in general depend on the contents of the store in which it is reduced, the evaluation rules should take not just a term but also a store as argument. Furthermore, since the reduction of a term can cause side effects on the store, and these may affect the reduction of other terms in the future, the reduction rules need to return a new store. Thus, the shape of the single-step reduction relation needs to change from *t –> t'* to *t / st –> t' / st'*, where *st* and *st'* are the starting and ending states of the store.

To carry through this change, we first need to augment all of our existing reduction rules with stores:

value v2

(ST_AppAbs) (\x:T.t12) v2 / st –> x:=*v2*t12 / st

t1 / st –> t1' / st'

(ST_App1) t1 t2 / st –> t1' t2 / st'
value v1 t2 / st –> t2' / st'

(ST_App2) v1 t2 / st –> v1 t2' / st'

Note that the first rule here returns the store unchanged, since function application, in itself, has no side effects. The other two rules simply propagate side effects from premise to conclusion.

Now, the result of reducing a ref expression will be a fresh location; this is why we included locations in the syntax of terms and in the set of values. It is crucial to note that making this extension to the syntax of terms does not mean that we intend *programmers* to write terms involving explicit, concrete locations: such terms will arise only as intermediate results during reduction. This may seem odd, but it follows naturally from our design decision to represent the result of every reduction step by a modified *term*. If we had chosen a more "machine-like" model, e.g., with an explicit stack to contain values of bound identifiers, then the idea of adding locations to the set of allowed values might seem more obvious.

In terms of this expanded syntax, we can state reduction rules for the new constructs that manipulate locations and the store. First, to reduce a dereferencing expression !*t1*, we must first reduce *t1* until it becomes a value:

t1 / st –> t1' / st'

(ST_Deref) !t1 / st –> !t1' / st'

Once *t1* has finished reducing, we should have an expression of the form !l, where l is some location. (A term that attempts to dereference any other sort of value, such as a function or unit, is erroneous, as is a term that tries to dereference a location that is larger than the size |*st*| of the currently allocated store; the reduction rules simply get stuck in this case. The type-safety properties established below assure us that well-typed terms will never misbehave in this way.)

l < |st|

(ST_DerefLoc) !(loc l) / st –> lookup l st / st

Next, to reduce an assignment expression *t1*:=*t2*, we must first reduce *t1* until it becomes a value (a location), and then reduce *t2* until it becomes a value (of any sort):

t1 / st –> t1' / st'

(ST_Assign1) t1 := t2 / st –> t1' := t2 / st'
t2 / st –> t2' / st'

(ST_Assign2) v1 := t2 / st –> v1 := t2' / st'

Once we have finished with *t1* and *t2*, we have an expression of the form l:=*v2*, which we execute by updating the store to make location l contain *v2*:

l < |st|

(ST_Assign) loc l := v2 / st -> unit / l:=*v2*st

The notation [l:=*v2*]*st* means "the store that maps l to *v2* and maps all other locations to the same thing as *st*." Note that the term resulting from this reduction step is just unit; the interesting result is the updated store.

Finally, to reduct an expression of the form ref *t1*, we first reduce *t1* until it becomes a value:

t1 / st -> t1' / st'

(ST_Ref) ref t1 / st -> ref t1' / st'

Then, to reduce the ref itself, we choose a fresh location at the end of the current store – i.e., location |*st*| – and yield a new store that extends *st* with the new value *v1*.

(ST_RefValue) ref v1 / st -> loc |st| / st,v1

The value resulting from this step is the newly allocated location itself. (Formally, *st*,*v1* means *st* ++ *v1*::nil – i.e., to add a new reference cell to the store, we append it to the end.)

Note that these reduction rules do not perform any kind of garbage collection: we simply allow the store to keep growing without bound as reduction proceeds. This does not affect the correctness of the results of reduction (after all, the definition of "garbage" is precisely parts of the store that are no longer reachable and so cannot play any further role in reduction), but it means that a naive implementation of our evaluator might run out of memory where a more sophisticated evaluator would be able to continue by reusing locations whose contents have become garbage.

Here are the rules again, formally:

```
Reserved Notation "t1 '/' st1 '->' t2 '/' st2"
  (at level 40, st1 at level 39, t2 at level 39).

Import ListNotations.

Inductive step : tm × store → tm × store → Prop :=
  | ST_AppAbs : ∀ x T t12 v2 st,
         value v2 →
         app (abs x T t12) v2 / st -> [x:=v2]t12 / st
  | ST_App1 : ∀ t1 t1' t2 st st',
         t1 / st -> t1' / st' →
         app t1 t2 / st -> app t1' t2 / st'
  | ST_App2 : ∀ v1 t2 t2' st st',
         value v1 →
         t2 / st -> t2' / st' →
         app v1 t2 / st -> app v1 t2'/ st'
  | ST_SuccNat : ∀ n st,
         scc (const n) / st -> const (S n) / st
  | ST_Succ : ∀ t1 t1' st st',
         t1 / st -> t1' / st' →
         scc t1 / st -> scc t1' / st'
```

```
| ST_PredNat : ∀ n st,
      prd (const n) / st -> const (pred n) / st
| ST_Pred : ∀ t1 t1' st st',
      t1 / st -> t1' / st' →
      prd t1 / st -> prd t1' / st'
| ST_MultNats : ∀ n1 n2 st,
      mlt (const n1) (const n2) / st -> const (mult n1 n2) / st
| ST_Mult1 : ∀ t1 t2 t1' st st',
      t1 / st -> t1' / st' →
      mlt t1 t2 / st -> mlt t1' t2 / st'
| ST_Mult2 : ∀ v1 t2 t2' st st',
      value v1 →
      t2 / st -> t2' / st' →
      mlt v1 t2 / st -> mlt v1 t2' / st'
| ST_If0 : ∀ t1 t1' t2 t3 st st',
      t1 / st -> t1' / st' →
      test0 t1 t2 t3 / st -> test0 t1' t2 t3 / st'
| ST_If0_Zero : ∀ t2 t3 st,
      test0 (const 0) t2 t3 / st -> t2 / st
| ST_If0_Nonzero : ∀ n t2 t3 st,
      test0 (const (S n)) t2 t3 / st -> t3 / st
| ST_RefValue : ∀ v1 st,
      value v1 →
      ref v1 / st -> loc (length st) / (st ++ v1::nil)
| ST_Ref : ∀ t1 t1' st st',
      t1 / st -> t1' / st' →
      ref t1 / st -> ref t1' / st'
| ST_DerefLoc : ∀ st l,
      l < length st →
      deref (loc l) / st -> store_lookup l st / st
| ST_Deref : ∀ t1 t1' st st',
      t1 / st -> t1' / st' →
      deref t1 / st -> deref t1' / st'
| ST_Assign : ∀ v2 l st,
      value v2 →
      l < length st →
      assign (loc l) v2 / st -> unit / replace l v2 st
| ST_Assign1 : ∀ t1 t1' t2 st st',
      t1 / st -> t1' / st' →
      assign t1 t2 / st -> assign t1' t2 / st'
| ST_Assign2 : ∀ v1 t2 t2' st st',
      value v1 →
```

```
t2 / st -> t2' / st' →
assign v1 t2 / st -> assign v1 t2' / st'
```

where "t1 '/' st1 '–>' t2 '/' st2" := (**step** (t1,st1) (t2,st2)).

One slightly ugly point should be noted here: In the ST_RefValue rule, we extend the state by writing *st* ++ *v1*::nil rather than the more natural *st* ++ [*v1*]. The reason for this is that the notation we've defined for substitution uses square brackets, which clash with the standard library's notation for lists.

```
Hint Constructors step.
```

```
Definition multistep := (multi step).
Notation "t1 '/' st '–>*' t2 '/' st'" :=
               (multistep (t1,st) (t2,st'))
               (at level 40, st at level 39, t2 at level 39).
```

16.5 Typing

The contexts assigning types to free variables are exactly the same as for the STLC: partial maps from identifiers to types.

```
Definition context := partial_map ty.
```

16.5.1 Store typings

Having extended our syntax and reduction rules to accommodate references, our last job is to write down typing rules for the new constructs (and, of course, to check that these rules are sound!). Naturally, the key question is, "What is the type of a location?"

First of all, notice that this question doesn't arise when typechecking terms that programmers actually write. Concrete location constants arise only in terms that are the intermediate results of reduction; they are not in the language that programmers write. So we only need to determine the type of a location when we're in the middle of a reduction sequence, e.g., trying to apply the progress or preservation lemmas. Thus, even though we normally think of typing as a *static* program property, it makes sense for the typing of locations to depend on the *dynamic* progress of the program too.

As a first try, note that when we reduce a term containing concrete locations, the type of the result depends on the contents of the store that we start with. For example, if we reduce the term !(loc 1) in the store [unit, unit], the result is unit; if we reduce the same term in the store [unit, \x:*Unit.x*], the result is \x:*Unit.x*. With respect to the former store, the location 1 has type Unit, and with respect to the latter it has type Unit→Unit. This observation leads us immediately to a first attempt at a typing rule for locations:

Gamma |- lookup l st : T1

Gamma |- loc l : Ref T1

That is, to find the type of a location l, we look up the current contents of l in the store and calculate the type *T1* of the contents. The type of the location is then Ref *T1*.

Having begun in this way, we need to go a little further to reach a consistent state. In effect, by making the type of a term depend on the store, we have changed the typing relation from a three-place relation (between contexts, terms, and types) to a four-place relation (between contexts, *stores*, terms, and types). Since the store is, intuitively, part of the context in which we calculate the type of a term, let's write this four-place relation with the store to the left of the turnstile: *Gamma*; *st* $\vdash$ *t* : T. Our rule for typing references now has the form

Gamma; st |- lookup l st : T1

Gamma; st |- loc l : Ref T1

and all the rest of the typing rules in the system are extended similarly with stores. (The other rules do not need to do anything interesting with their stores – just pass them from premise to conclusion.)

However, this rule will not quite do. For one thing, typechecking is rather inefficient, since calculating the type of a location l involves calculating the type of the current contents *v* of l. If l appears many times in a term *t*, we will re-calculate the type of *v* many times in the course of constructing a typing derivation for *t*. Worse, if *v* itself contains locations, then we will have to recalculate *their* types each time they appear. Worse yet, the proposed typing rule for locations may not allow us to derive anything at all, if the store contains a *cycle*. For example, there is no finite typing derivation for the location 0 with respect to this store:

\x:Nat. (!(loc 1)) x, \x:Nat. (!(loc 0)) x

Exercise: 2 stars, standard (cyclic_store) Can you find a term whose reduction will create this particular cyclic store?

```
Definition manual_grade_for_cyclic_store : option (nat×string) := None.
```

□

These problems arise from the fact that our proposed typing rule for locations requires us to recalculate the type of a location every time we mention it in a term. But this, intuitively, should not be necessary. After all, when a location is first created, we know the type of the initial value that we are storing into it. Suppose we are willing to enforce the invariant that the type of the value contained in a given location *never changes*; that is, although we may later store other values into this location, those other values will always have the same type as the initial one. In other words, we always have in mind a single, definite type for every location in the store, which is fixed when the location is allocated. Then these intended types can be collected together as a *store typing* – a finite function mapping locations to types.

As with the other type systems we've seen, this conservative typing restriction on allowed updates means that we will rule out as ill-typed some programs that could reduce perfectly well without getting stuck.

Just as we did for stores, we will represent a store type simply as a list of types: the type at index i records the type of the values that we expect to be stored in cell i.

```
Definition store_ty := list ty.
```

The store_Tlookup function retrieves the type at a particular index.

```
Definition store_Tlookup (n:nat) (ST:store_ty) :=
  nth n ST Unit.
```

Suppose we are given a store typing ST describing the store st in which some term t will be reduced. Then we can use ST to calculate the type of the result of t without ever looking directly at st. For example, if ST is [Unit, Unit→Unit], then we can immediately infer that !(loc 1) has type Unit→Unit. More generally, the typing rule for locations can be reformulated in terms of store typings like this:

l < |ST|

Gamma; ST |- loc l : Ref (lookup l ST)

That is, as long as l is a valid location, we can compute the type of l just by looking it up in ST. Typing is again a four-place relation, but it is parameterized on a store *typing* rather than a concrete store. The rest of the typing rules are analogously augmented with store typings.

16.5.2 The Typing Relation

We can now formalize the typing relation for the STLC with references. Here, again, are the rules we're adding to the base STLC (with numbers and Unit):

l < |ST|

(T_Loc) Gamma; ST |- loc l : Ref (lookup l ST)

Gamma; ST |- t1 : T1

(T_Ref) Gamma; ST |- ref t1 : Ref T1

Gamma; ST |- t1 : Ref T11

(T_Deref) Gamma; ST |- !t1 : T11

Gamma; ST |- t1 : Ref T11 Gamma; ST |- t2 : T11

(T_Assign) Gamma; ST |- t1 := t2 : Unit

```
Reserved Notation "Gamma ';' ST '|-' t '\in' T" (at level 40).

Inductive has_type : context → store_ty → tm → ty → Prop :=
  | T_Var : ∀ Gamma ST x T,
      Gamma x = Some T →
      Gamma; ST ⊢ (var x) \in T
  | T_Abs : ∀ Gamma ST x T11 T12 t12,
```

```
        (update Gamma x T11); ST ⊢ t12 \in T12 →
        Gamma; ST ⊢ (abs x T11 t12) \in (Arrow T11 T12)
  | T_App : ∀ T1 T2 Gamma ST t1 t2,
        Gamma; ST ⊢ t1 \in (Arrow T1 T2) →
        Gamma; ST ⊢ t2 \in T1 →
        Gamma; ST ⊢ (app t1 t2) \in T2
  | T_Nat : ∀ Gamma ST n,
        Gamma; ST ⊢ (const n) \in Nat
  | T_Succ : ∀ Gamma ST t1,
        Gamma; ST ⊢ t1 \in Nat →
        Gamma; ST ⊢ (scc t1) \in Nat
  | T_Pred : ∀ Gamma ST t1,
        Gamma; ST ⊢ t1 \in Nat →
        Gamma; ST ⊢ (prd t1) \in Nat
  | T_Mult : ∀ Gamma ST t1 t2,
        Gamma; ST ⊢ t1 \in Nat →
        Gamma; ST ⊢ t2 \in Nat →
        Gamma; ST ⊢ (mlt t1 t2) \in Nat
  | T_If0 : ∀ Gamma ST t1 t2 t3 T,
        Gamma; ST ⊢ t1 \in Nat →
        Gamma; ST ⊢ t2 \in T →
        Gamma; ST ⊢ t3 \in T →
        Gamma; ST ⊢ (test0 t1 t2 t3) \in T
  | T_Unit : ∀ Gamma ST,
        Gamma; ST ⊢ unit \in Unit
  | T_Loc : ∀ Gamma ST l,
        l < length ST →
        Gamma; ST ⊢ (loc l) \in (Ref (store_Tlookup l ST))
  | T_Ref : ∀ Gamma ST t1 T1,
        Gamma; ST ⊢ t1 \in T1 →
        Gamma; ST ⊢ (ref t1) \in (Ref T1)
  | T_Deref : ∀ Gamma ST t1 T11,
        Gamma; ST ⊢ t1 \in (Ref T11) →
        Gamma; ST ⊢ (deref t1) \in T11
  | T_Assign : ∀ Gamma ST t1 t2 T11,
        Gamma; ST ⊢ t1 \in (Ref T11) →
        Gamma; ST ⊢ t2 \in T11 →
        Gamma; ST ⊢ (assign t1 t2) \in Unit

where "Gamma ';' ST '|-' t '\in' T" := (has_type Gamma ST t T).
Hint Constructors has_type.
```

Of course, these typing rules will accurately predict the results of reduction only if the concrete store used during reduction actually conforms to the store typing that we assume for purposes of typechecking. This proviso exactly parallels the situation with free variables in the basic STLC: the substitution lemma promises that, if *Gamma* $\vdash$ *t* : T, then we can replace the free variables in *t* with values of the types listed in *Gamma* to obtain a closed term of type T, which, by the type preservation theorem will reduce to a final result of type T if it yields any result at all. We will see below how to formalize an analogous intuition for stores and store typings.

However, for purposes of typechecking the terms that programmers actually write, we do not need to do anything tricky to guess what store typing we should use. Concrete locations arise only in terms that are the intermediate results of reduction; they are not in the language that programmers write. Thus, we can simply typecheck the programmer's terms with respect to the *empty* store typing. As reduction proceeds and new locations are created, we will always be able to see how to extend the store typing by looking at the type of the initial values being placed in newly allocated cells; this intuition is formalized in the statement of the type preservation theorem below.

16.6 Properties

Our final task is to check that standard type safety properties continue to hold for the STLC with references. The progress theorem ("well-typed terms are not stuck") can be stated and proved almost as for the STLC; we just need to add a few straightforward cases to the proof to deal with the new constructs. The preservation theorem is a bit more interesting, so let's look at it first.

16.6.1 Well-Typed Stores

Since we have extended both the reduction relation (with initial and final stores) and the typing relation (with a store typing), we need to change the statement of preservation to include these parameters. But clearly we cannot just add stores and store typings without saying anything about how they are related – i.e., this is wrong:

```
Theorem preservation_wrong1 : ∀ ST T t st t' st',
  empty; ST ⊢ t \in T →
  t / st -> t' / st' →
  empty; ST ⊢ t' \in T.
Abort.
```

If we typecheck with respect to some set of assumptions about the types of the values in the store and then reduce with respect to a store that violates these assumptions, the result will be disaster. We say that a store *st* is *well typed* with respect a store typing *ST* if the term at each location l in *st* has the type at location l in *ST*. Since only closed terms ever get stored in locations (why?), it suffices to type them in the empty context. The following definition of store_well_typed formalizes this.

```
Definition store_well_typed (ST:store_ty) (st:store) :=
  length ST = length st ∧
  (∀ l, l < length st →
     empty; ST ⊢ (store_lookup l st) \in (store_Tlookup l ST)).
```

Informally, we will write $ST \vdash st$ for store_well_typed ST st.

Intuitively, a store *st* is consistent with a store typing *ST* if every value in the store has the type predicted by the store typing. The only subtle point is the fact that, when typing the values in the store, we supply the very same store typing to the typing relation. This allows us to type circular stores like the one we saw above.

Exercise: 2 stars, standard (store_not_unique) Can you find a store *st*, and two different store typings *ST1* and *ST2* such that both $ST1 \vdash st$ and $ST2 \vdash st$?

```
Definition manual_grade_for_store_not_unique : option (nat×string) := None.
```

□

We can now state something closer to the desired preservation property:

```
Theorem preservation_wrong2 : ∀ ST T t st t' st',
  empty; ST ⊢ t \in T →
  t / st -> t' / st' →
  store_well_typed ST st →
  empty; ST ⊢ t' \in T.
Abort.
```

This statement is fine for all of the reduction rules except the allocation rule ST_RefValue. The problem is that this rule yields a store with a larger domain than the initial store, which falsifies the conclusion of the above statement: if *st'* includes a binding for a fresh location l, then l cannot be in the domain of *ST*, and it will not be the case that *t'* (which definitely mentions l) is typable under *ST*.

16.6.2 Extending Store Typings

Evidently, since the store can increase in size during reduction, we need to allow the store typing to grow as well. This motivates the following definition. We say that the store type *ST' extends ST* if *ST'* is just *ST* with some new types added to the end.

```
Inductive extends : store_ty → store_ty → Prop :=
  | extends_nil : ∀ ST',
      extends ST' nil
  | extends_cons : ∀ x ST' ST,
      extends ST' ST →
      extends (x::ST') (x::ST).
Hint Constructors extends.
```

We'll need a few technical lemmas about extended contexts.

First, looking up a type in an extended store typing yields the same result as in the original:

```
Lemma extends_lookup : ∀ l ST ST',
  l < length ST →
  extends ST' ST →
  store_Tlookup l ST' = store_Tlookup l ST.
Proof with auto.
  intros l ST ST' Hlen H.
  generalize dependent ST'. generalize dependent l.
  induction ST as [|a ST2]; intros l Hlen ST' HST'.
  - inversion Hlen.
  - unfold store_Tlookup in *.
    destruct ST'.
    + inversion HST'.
    +
      inversion HST'; subst.
      destruct l as [|l'].
      × auto.
      × simpl. apply IHST2...
        simpl in Hlen; omega.
Qed.
```

Next, if ST' extends ST, the length of ST' is at least that of ST.

```
Lemma length_extends : ∀ l ST ST',
  l < length ST →
  extends ST' ST →
  l < length ST'.
Proof with eauto.
  intros. generalize dependent l. induction H0; intros l Hlen.
    inversion Hlen.
    simpl in *.
    destruct l; try omega.
      apply lt_n_S. apply IHextends. omega.
Qed.
```

Finally, ST ++ T extends ST, and **extends** is reflexive.

```
Lemma extends_app : ∀ ST T,
  extends (ST ++ T) ST.
Proof with auto.
  induction ST; intros T...
  simpl...
Qed.
```

```
Lemma extends_refl : ∀ ST,
```

```
  extends ST ST.
Proof.
  induction ST; auto.
Qed.
```

16.6.3 Preservation, Finally

We can now give the final, correct statement of the type preservation property:

```
Definition preservation_theorem := ∀ ST t t' T st st',
  empty; ST ⊢ t \in T →
  store_well_typed ST st →
  t / st -> t' / st' →
  ∃ ST',
    (extends ST' ST ∧
     empty; ST' ⊢ t' \in T ∧
     store_well_typed ST' st').
```

Note that the preservation theorem merely asserts that there is *some* store typing *ST'* extending *ST* (i.e., agreeing with *ST* on the values of all the old locations) such that the new term *t'* is well typed with respect to *ST'*; it does not tell us exactly what *ST'* is. It is intuitively clear, of course, that *ST'* is either *ST* or else exactly *ST* ++ *T1*::nil, where *T1* is the type of the value *v1* in the extended store *st* ++ *v1*::nil, but stating this explicitly would complicate the statement of the theorem without actually making it any more useful: the weaker version above is already in the right form (because its conclusion implies its hypothesis) to "turn the crank" repeatedly and conclude that every *sequence* of reduction steps preserves well-typedness. Combining this with the progress property, we obtain the usual guarantee that "well-typed programs never go wrong."

In order to prove this, we'll need a few lemmas, as usual.

16.6.4 Substitution Lemma

First, we need an easy extension of the standard substitution lemma, along with the same machinery about context invariance that we used in the proof of the substitution lemma for the STLC.

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
      appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
      appears_free_in x t1 → appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
      appears_free_in x t2 → appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
      y ≠ x →
```

```
      appears_free_in x t12 →
      appears_free_in x (abs y T11 t12)
  | afi_succ : ∀ x t1,
      appears_free_in x t1 →
      appears_free_in x (scc t1)
  | afi_pred : ∀ x t1,
      appears_free_in x t1 →
      appears_free_in x (prd t1)
  | afi_mult1 : ∀ x t1 t2,
      appears_free_in x t1 →
      appears_free_in x (mlt t1 t2)
  | afi_mult2 : ∀ x t1 t2,
      appears_free_in x t2 →
      appears_free_in x (mlt t1 t2)
  | afi_if0_1 : ∀ x t1 t2 t3,
      appears_free_in x t1 →
      appears_free_in x (test0 t1 t2 t3)
  | afi_if0_2 : ∀ x t1 t2 t3,
      appears_free_in x t2 →
      appears_free_in x (test0 t1 t2 t3)
  | afi_if0_3 : ∀ x t1 t2 t3,
      appears_free_in x t3 →
      appears_free_in x (test0 t1 t2 t3)
  | afi_ref : ∀ x t1,
      appears_free_in x t1 → appears_free_in x (ref t1)
  | afi_deref : ∀ x t1,
      appears_free_in x t1 → appears_free_in x (deref t1)
  | afi_assign1 : ∀ x t1 t2,
      appears_free_in x t1 → appears_free_in x (assign t1 t2)
  | afi_assign2 : ∀ x t1 t2,
      appears_free_in x t2 → appears_free_in x (assign t1 t2).
Hint Constructors appears_free_in.

Lemma free_in_context : ∀ x t T Gamma ST,
   appears_free_in x t →
   Gamma; ST ⊢ t \in T →
   ∃ T', Gamma x = Some T'.
Proof with eauto.
  intros. generalize dependent Gamma. generalize dependent T.
  induction H;
        intros; (try solve [ inversion H0; subst; eauto ]).
  -
    inversion H1; subst.
```

```
    apply IHappears_free_in in H8.
    rewrite update_neq in H8; assumption.
Qed.

Lemma context_invariance : ∀ Gamma Gamma' ST t T,
  Gamma; ST ⊢ t \in T →
  (∀ x, appears_free_in x t → Gamma x = Gamma' x) →
  Gamma'; ST ⊢ t \in T.
Proof with eauto.
  intros.
  generalize dependent Gamma'.
  induction H; intros...
  -
    apply T_Var. symmetry. rewrite ← H...
  -
    apply T_Abs. apply IHhas_type; intros.
    unfold update, t_update.
    destruct (eqb_stringP x x0)...
  -
    eapply T_App.
      apply IHhas_type1...
      apply IHhas_type2...
  -
    eapply T_Mult.
      apply IHhas_type1...
      apply IHhas_type2...
  -
    eapply T_If0.
      apply IHhas_type1...
      apply IHhas_type2...
      apply IHhas_type3...
  -
    eapply T_Assign.
      apply IHhas_type1...
      apply IHhas_type2...
Qed.

Lemma substitution_preserves_typing : ∀ Gamma ST x s S t T,
  empty; ST ⊢ s \in S →
  (update Gamma x S); ST ⊢ t \in T →
  Gamma; ST ⊢ ([x:=s]t) \in T.
Proof with eauto.
  intros Gamma ST x s S t T Hs Ht.
  generalize dependent Gamma. generalize dependent T.
```

```
  induction t; intros T Gamma H;
    inversion H; subst; simpl...
  -
    rename s0 into y.
    destruct (eqb_stringP x y).
    +
      subst.
      rewrite update_eq in H3.
      inversion H3; subst.
      eapply context_invariance...
      intros x Hcontra.
      destruct (free_in_context _ _ _ _ _ _ Hcontra Hs)
        as [T' HT'].
      inversion HT'.
    +
      apply T_Var.
      rewrite update_neq in H3...
  - subst.
    rename s0 into y.
    destruct (eqb_stringP x y).
    +
      subst.
      apply T_Abs. eapply context_invariance...
      intros. rewrite update_shadow. reflexivity.
    +
      apply T_Abs. apply IHt.
      eapply context_invariance...
      intros. unfold update, t_update.
      destruct (eqb_stringP y x0)...
      subst.
      rewrite false_eqb_string...
Qed.
```

16.6.5 Assignment Preserves Store Typing

Next, we must show that replacing the contents of a cell in the store with a new value of appropriate type does not change the overall type of the store. (This is needed for the ST_Assign rule.)

```
Lemma assign_pres_store_typing : ∀ ST st l t,
  l < length st →
  store_well_typed ST st →
  empty; ST ⊢ t \in (store_Tlookup l ST) →
```

```
  store_well_typed ST (replace l t st).
Proof with auto.
  intros ST st l t Hlen HST Ht.
  inversion HST; subst.
  split. rewrite length_replace...
  intros l' Hl'.
  destruct (l' =? l) eqn: Heqll'.
  -
    apply Nat.eqb_eq in Heqll'; subst.
    rewrite lookup_replace_eq...
  -
    apply Nat.eqb_neq in Heqll'.
    rewrite lookup_replace_neq...
    rewrite length_replace in Hl'.
    apply H0...
Qed.
```

16.6.6 Weakening for Stores

Finally, we need a lemma on store typings, stating that, if a store typing is extended with a new location, the extended one still allows us to assign the same types to the same terms as the original.

(The lemma is called store_weakening because it resembles the "weakening" lemmas found in proof theory, which show that adding a new assumption to some logical theory does not decrease the set of provable theorems.)

```
Lemma store_weakening : ∀ Gamma ST ST' t T,
  extends ST' ST →
  Gamma; ST ⊢ t \in T →
  Gamma; ST' ⊢ t \in T.
Proof with eauto.
  intros. induction H0; eauto.
  -
    erewrite ← extends_lookup...
    apply T_Loc.
    eapply length_extends...
Qed.
```

We can use the store_weakening lemma to prove that if a store is well typed with respect to a store typing, then the store extended with a new term t will still be well typed with respect to the store typing extended with t's type.

```
Lemma store_well_typed_app : ∀ ST st t1 T1,
  store_well_typed ST st →
  empty; ST ⊢ t1 \in T1 →
```

```
  store_well_typed (ST ++ T1::nil) (st ++ t1::nil).
Proof with auto.
  intros.
  unfold store_well_typed in *.
  inversion H as [Hlen Hmatch]; clear H.
  rewrite app_length, plus_comm. simpl.
  rewrite app_length, plus_comm. simpl.
  split...
  -
    intros l Hl.
    unfold store_lookup, store_Tlookup.
    apply le_lt_eq_dec in Hl; inversion Hl as [Hlt | Heq].
    +
      apply lt_S_n in Hlt.
      rewrite !app_nth1...
      × apply store_weakening with ST. apply extends_app.
        apply Hmatch...
      × rewrite Hlen...
    +
      inversion Heq.
      rewrite app_nth2; try omega.
      rewrite ← Hlen.
      rewrite minus_diag. simpl.
      apply store_weakening with ST...
      { apply extends_app. }
        rewrite app_nth2; try omega.
      rewrite minus_diag. simpl. trivial.
Qed.
```

16.6.7 Preservation!

Now that we've got everything set up right, the proof of preservation is actually quite straightforward.

Begin with one technical lemma:

```
Lemma nth_eq_last : ∀ A (l:list A) x d,
  nth (length l) (l ++ x::nil) d = x.
Proof.
  induction l; intros; [ auto | simpl; rewrite IHl; auto ].
Qed.
```

And here, at last, is the preservation theorem and proof:

```
Theorem preservation : ∀ ST t t' T st st',
  empty; ST ⊢ t \in T →
```

```
  store_well_typed ST st →
  t / st -> t' / st' →
  ∃ ST',
    (extends ST' ST ∧
     empty; ST' ⊢ t' \in T ∧
     store_well_typed ST' st').
Proof with eauto using store_weakening, extends_refl.
  remember (@empty ty) as Gamma.
  intros ST t t' T st st' Ht.
  generalize dependent t'.
  induction Ht; intros t' HST Hstep;
    subst; try solve_by_invert; inversion Hstep; subst;
    try (eauto using store_weakening, extends_refl).
  - ∃ ST.
    inversion Ht1; subst.
    split; try split... eapply substitution_preserves_typing...
  -
    eapply IHHt1 in H0...
    inversion H0 as [ST' [Hext [Hty Hsty]]].
    ∃ ST'...
  -
    eapply IHHt2 in H5...
    inversion H5 as [ST' [Hext [Hty Hsty]]].
    ∃ ST'...
  -
    +
      eapply IHHt in H0...
      inversion H0 as [ST' [Hext [Hty Hsty]]].
      ∃ ST'...
  -
    +
      eapply IHHt in H0...
      inversion H0 as [ST' [Hext [Hty Hsty]]].
      ∃ ST'...
  -
    eapply IHHt1 in H0...
    inversion H0 as [ST' [Hext [Hty Hsty]]].
    ∃ ST'...
  -
    eapply IHHt2 in H5...
    inversion H5 as [ST' [Hext [Hty Hsty]]].
    ∃ ST'...
```

```
-
  +
    eapply IHHt1 in H0...
    inversion H0 as [ST' [Hext [Hty Hsty]]].
    ∃ ST'... split...
-
  ∃ (ST ++ T1::nil).
  inversion HST; subst.
  split.
    apply extends_app.
  split.
    replace (Ref T1)
      with (Ref (store_Tlookup (length st) (ST ++ T1::nil))).
    apply T_Loc.
    rewrite ← H. rewrite app_length, plus_comm. simpl. omega.
    unfold store_Tlookup. rewrite ← H. rewrite nth_eq_last.
    reflexivity.
    apply store_well_typed_app; assumption.
-
  eapply IHHt in H0...
  inversion H0 as [ST' [Hext [Hty Hsty]]].
  ∃ ST'...
-
  ∃ ST. split; try split...
  inversion HST as [_ Hsty].
  replace T11 with (store_Tlookup l ST).
  apply Hsty...
  inversion Ht; subst...
-
  eapply IHHt in H0...
  inversion H0 as [ST' [Hext [Hty Hsty]]].
  ∃ ST'...
-
  ∃ ST. split; try split...
  eapply assign_pres_store_typing...
  inversion Ht1; subst...
-
  eapply IHHt1 in H0...
  inversion H0 as [ST' [Hext [Hty Hsty]]].
  ∃ ST'...
-
  eapply IHHt2 in H5...
```

```
    inversion H5 as [ST' [Hext [Hty Hsty]]].
    ∃ ST'...
Qed.
```

Exercise: 3 stars, standard (preservation_informal) Write a careful informal proof of the preservation theorem, concentrating on the T_App, T_Deref, T_Assign, and T_Ref cases.

```
Definition manual_grade_for_preservation_informal : option (nat×string) := None.
```

□

16.6.8 Progress

As we've said, progress for this system is pretty easy to prove; the proof is very similar to the proof of progress for the STLC, with a few new cases for the new syntactic constructs.

```
Theorem progress : ∀ ST t T st,
  empty; ST ⊢ t \in T →
  store_well_typed ST st →
  (value t ∨ ∃ t' st', t / st -> t' / st').
Proof with eauto.
  intros ST t T st Ht HST. remember (@empty ty) as Gamma.
  induction Ht; subst; try solve_by_invert...
  -
    right. destruct IHHt1 as [Ht1p | Ht1p]...
    +
      inversion Ht1p; subst; try solve_by_invert.
      destruct IHHt2 as [Ht2p | Ht2p]...
      ×
        inversion Ht2p as [t2' [st' Hstep]].
        ∃ (app (abs x T t) t2'), st'...
    +
      inversion Ht1p as [t1' [st' Hstep]].
      ∃ (app t1' t2), st'...
  -
    right. destruct IHHt as [Ht1p | Ht1p]...
    +
      inversion Ht1p; subst; try solve [ inversion Ht ].
      ×
        ∃ (const (S n)), st...
    +
      inversion Ht1p as [t1' [st' Hstep]].
      ∃ (scc t1'), st'...
  -
```

```
  right. destruct IHHt as [Ht1p | Ht1p]...
  +
    inversion Ht1p; subst; try solve [inversion Ht ].
    ×
      ∃ (const (pred n)), st...
  +
    inversion Ht1p as [t1' [st' Hstep]].
    ∃ (prd t1'), st'...
-
  right. destruct IHHt1 as [Ht1p | Ht1p]...
  +
    inversion Ht1p; subst; try solve [inversion Ht1].
    destruct IHHt2 as [Ht2p | Ht2p]...
    ×
      inversion Ht2p; subst; try solve [inversion Ht2].
      ∃ (const (mult n n0)), st...
    ×
      inversion Ht2p as [t2' [st' Hstep]].
      ∃ (mlt (const n) t2'), st'...
  +
    inversion Ht1p as [t1' [st' Hstep]].
    ∃ (mlt t1' t2), st'...
-
  right. destruct IHHt1 as [Ht1p | Ht1p]...
  +
    inversion Ht1p; subst; try solve [inversion Ht1].
    destruct n.
    × ∃ t2, st...
    × ∃ t3, st...
  +
    inversion Ht1p as [t1' [st' Hstep]].
    ∃ (test0 t1' t2 t3), st'...
-
  right. destruct IHHt as [Ht1p | Ht1p]...
  +
    inversion Ht1p as [t1' [st' Hstep]].
    ∃ (ref t1'), st'...
-
  right. destruct IHHt as [Ht1p | Ht1p]...
  +
    inversion Ht1p; subst; try solve_by_invert.
    eexists. eexists. apply ST_DerefLoc...
```

```
      inversion Ht; subst. inversion HST; subst.
      rewrite ← H...
    +
      inversion Ht1p as [t1' [st' Hstep]].
      ∃ (deref t1'), st'...
  -
    right. destruct IHHt1 as [Ht1p|Ht1p]...
    +
      destruct IHHt2 as [Ht2p|Ht2p]...
      ×
        inversion Ht1p; subst; try solve_by_invert.
        eexists. eexists. apply ST_Assign...
        inversion HST; subst. inversion Ht1; subst.
        rewrite H in H5...
      ×
        inversion Ht2p as [t2' [st' Hstep]].
        ∃ (assign t1 t2'), st'...
    +
      inversion Ht1p as [t1' [st' Hstep]].
      ∃ (assign t1' t2), st'...
Qed.
```

16.7 References and Nontermination

An important fact about the STLC (proved in chapter Norm) is that it is is *normalizing* – that is, every well-typed term can be reduced to a value in a finite number of steps.

What about STLC + references? Surprisingly, adding references causes us to lose the normalization property: there exist well-typed terms in the STLC + references which can continue to reduce forever, without ever reaching a normal form!

How can we construct such a term? The main idea is to make a function which calls itself. We first make a function which calls another function stored in a reference cell; the trick is that we then smuggle in a reference to itself!

(\r:Ref (Unit -> Unit). r := (\x:Unit.(!r) unit); (!r) unit) (ref (\x:Unit.unit))

First, ref (\x:*Unit.unit*) creates a reference to a cell of type Unit → Unit. We then pass this reference as the argument to a function which binds it to the name r, and assigns to it the function \x:Unit.(!r) unit – that is, the function which ignores its argument and calls the function stored in r on the argument unit; but of course, that function is itself! To start the divergent loop, we execute the function stored in the cell by evaluating (!r) unit.

Here is the divergent term in Coq:

```
Module ExampleVariables.

Open Scope string_scope.
```

```
Definition x := "x".
Definition y := "y".
Definition r := "r".
Definition s := "s".

End ExampleVariables.

Module RefsAndNontermination.
Import ExampleVariables.

Definition loop_fun :=
  abs x Unit (app (deref (var r)) unit).

Definition loop :=
  app
    (abs r (Ref (Arrow Unit Unit))
      (tseq (assign (var r) loop_fun)
              (app (deref (var r)) unit)))
    (ref (abs x Unit unit)).
```

This term is well typed:

```
Lemma loop_typeable : ∃ T, empty; nil ⊢ loop \in T.
Proof with eauto.
  eexists. unfold loop. unfold loop_fun.
  eapply T_App...
  eapply T_Abs...
  eapply T_App...
    eapply T_Abs. eapply T_App. eapply T_Deref. eapply T_Var.
    unfold update, t_update. simpl. reflexivity. auto.
  eapply T_Assign.
    eapply T_Var. unfold update, t_update. simpl. reflexivity.
  eapply T_Abs.
    eapply T_App...
      eapply T_Deref. eapply T_Var. reflexivity.
Qed.
```

To show formally that the term diverges, we first define the **step_closure** of the single-step reduction relation, written –>+. This is just like the reflexive step closure of single-step reduction (which we're been writing –>*), except that it is not reflexive: *t* –>+ *t'* means that *t* can reach *t'* by *one or more* steps of reduction.

```
Inductive step_closure {X:Type} (R: relation X) : X → X → Prop :=
  | sc_one : ∀ (x y : X),
                R x y → step_closure R x y
  | sc_step : ∀ (x y z : X),
                R x y →
                step_closure R y z →
```

step_closure R x z.

```
Definition multistep1 := (step_closure step).
Notation "t1 '/' st '->+' t2 '/' st'" :=
        (multistep1 (t1,st) (t2,st'))
        (at level 40, st at level 39, t2 at level 39).
```

Now, we can show that the expression loop reduces to the expression !(loc 0) unit and the size-one store [r:=(loc 0)]loop_fun.

As a convenience, we introduce a slight variant of the *normalize* tactic, called *reduce*, which tries solving the goal with multi_refl at each step, instead of waiting until the goal can't be reduced any more. Of course, the whole point is that loop doesn't normalize, so the old *normalize* tactic would just go into an infinite loop reducing it forever!

```
Ltac print_goal := match goal with ⊢ ?x ⇒ idtac x end.
Ltac reduce :=
    repeat (print_goal; eapply multi_step ;
            [ (eauto 10; fail) | (instantiate; compute)];
            try solve [apply multi_refl]).
```

Next, we use *reduce* to show that loop steps to !(loc 0) unit, starting from the empty store.

```
Lemma loop_steps_to_loop_fun :
  loop / nil ->*
  app (deref (loc 0)) unit / cons ([r:=loc 0]loop_fun) nil.
Proof.
  unfold loop.
  reduce.
Qed.
```

Finally, we show that the latter expression reduces in two steps to itself!

```
Lemma loop_fun_step_self :
  app (deref (loc 0)) unit / cons ([r:=loc 0]loop_fun) nil ->+
  app (deref (loc 0)) unit / cons ([r:=loc 0]loop_fun) nil.
Proof with eauto.
  unfold loop_fun; simpl.
  eapply sc_step. apply ST_App1...
  eapply sc_one. compute. apply ST_AppAbs...
Qed.
```

Exercise: 4 stars, standard (factorial_ref) Use the above ideas to implement a factorial function in STLC with references. (There is no need to prove formally that it really behaves like the factorial. Just uncomment the example below to make sure it gives the correct result when applied to the argument 4.)

```
Definition factorial : tm
```

. *Admitted.*

Lemma factorial_type : empty; nil ⊢ *factorial* \in (Arrow Nat Nat).
Proof with eauto.
Admitted.

If your definition is correct, you should be able to just uncomment the example below; the proof should be fully automatic using the *reduce* tactic.

16.8 Additional Exercises

Exercise: 5 stars, standard, optional (garabage_collector) Challenge problem: modify our formalization to include an account of garbage collection, and prove that it satisfies whatever nice properties you can think to prove about it.

□

End RefsAndNontermination.
End STLCRef.

Chapter 17

RecordSub: Subtyping with Records

In this chapter, we combine two significant extensions of the pure STLC – records (from chapter Records) and subtyping (from chapter Sub) – and explore their interactions. Most of the concepts have already been discussed in those chapters, so the presentation here is somewhat terse. We just comment where things are nonstandard.

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Strings.String.
From PLF Require Import Maps.
From PLF Require Import Smallstep.
From PLF Require Import MoreStlc.
```

17.1 Core Definitions

Syntax

```
Inductive ty : Type :=

  | Top : ty
  | Base : string → ty
  | Arrow : ty → ty → ty

  | RNil : ty
  | RCons : string → ty → ty → ty.
Inductive tm : Type :=

  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm
  | rproj : tm → string → tm
```

```
  | rnil : tm
  | rcons : string → tm → tm → tm.
```

Well-Formedness

The syntax of terms and types is a bit too loose, in the sense that it admits things like a record type whose final "tail" is Top or some arrow type rather than *Nil*. To avoid such cases, it is useful to assume that all the record types and terms that we see will obey some simple well-formedness conditions.

An interesting technical question is *whether the basic properties* of *the system* – progress *and* preservation – *remain* true if *we* drop *these conditions.* I *believe they* do, *and* I *would encourage motivated readers to* try *to check this* by *dropping the conditions from the definitions* of *typing and subtyping and adjusting the proofs* in *the rest* of *the chapter accordingly. This* is not a trivial *exercise* (*or I'd have* done *it*!), *but it should* not *involve changing the basic structure* of *the proofs.* If *someone does* do *it*, *please* let *me know.* –*BCP* 5/16.

```
Inductive record_ty : ty → Prop :=
  | RTnil :
        record_ty RNil
  | RTcons : ∀ i T1 T2,
        record_ty (RCons i T1 T2).

Inductive record_tm : tm → Prop :=
  | rtnil :
        record_tm rnil
  | rtcons : ∀ i t1 t2,
        record_tm (rcons i t1 t2).

Inductive well_formed_ty : ty → Prop :=
  | wfTop :
        well_formed_ty Top
  | wfBase : ∀ i,
        well_formed_ty (Base i)
  | wfArrow : ∀ T1 T2,
        well_formed_ty T1 →
        well_formed_ty T2 →
        well_formed_ty (Arrow T1 T2)
  | wfRNil :
        well_formed_ty RNil
  | wfRCons : ∀ i T1 T2,
        well_formed_ty T1 →
        well_formed_ty T2 →
        record_ty T2 →
        well_formed_ty (RCons i T1 T2).

Hint Constructors record_ty record_tm well_formed_ty.
```

Substitution

Substitution and reduction are as before.

```
Fixpoint subst (x:string) (s:tm) (t:tm) : tm :=
  match t with
  | var y ⇒ if eqb_string x y then s else t
  | abs y T t1 ⇒ abs y T (if eqb_string x y then t1
                          else (subst x s t1))
  | app t1 t2 ⇒ app (subst x s t1) (subst x s t2)
  | rproj t1 i ⇒ rproj (subst x s t1) i
  | rnil ⇒ rnil
  | rcons i t1 tr2 ⇒ rcons i (subst x s t1) (subst x s tr2)
  end.
Notation "'[' x ':=' s ']' t" := (subst x s t) (at level 20).
```

Reduction

```
Inductive value : tm → Prop :=
  | v_abs : ∀ x T t,
      value (abs x T t)
  | v_rnil : value rnil
  | v_rcons : ∀ i v vr,
      value v →
      value vr →
      value (rcons i v vr).
Hint Constructors value.
Fixpoint Tlookup (i:string) (Tr:ty) : option ty :=
  match Tr with
  | RCons i' T Tr' ⇒
      if eqb_string i i' then Some T else Tlookup i Tr'
  | _ ⇒ None
  end.
Fixpoint tlookup (i:string) (tr:tm) : option tm :=
  match tr with
  | rcons i' t tr' ⇒
      if eqb_string i i' then Some t else tlookup i tr'
  | _ ⇒ None
  end.
Reserved Notation "t1 '-->' t2" (at level 40).
Inductive step : tm → tm → Prop :=
  | ST_AppAbs : ∀ x T t12 v2,
```

```
        value v2 →
        (app (abs x T t12) v2) -> [x:=v2]t12
  | ST_App1 : ∀ t1 t1' t2,
        t1 -> t1' →
        (app t1 t2) -> (app t1' t2)
  | ST_App2 : ∀ v1 t2 t2',
        value v1 →
        t2 -> t2' →
        (app v1 t2) -> (app v1 t2')
  | ST_Proj1 : ∀ tr tr' i,
       tr -> tr' →
       (rproj tr i) -> (rproj tr' i)
  | ST_ProjRcd : ∀ tr i vi,
       value tr →
       tlookup i tr = Some vi →
      (rproj tr i) -> vi
  | ST_Rcd_Head : ∀ i t1 t1' tr2,
       t1 -> t1' →
       (rcons i t1 tr2) -> (rcons i t1' tr2)
  | ST_Rcd_Tail : ∀ i v1 tr2 tr2',
       value v1 →
       tr2 -> tr2' →
       (rcons i v1 tr2) -> (rcons i v1 tr2')

where "t1 '->' t2" := (step t1 t2).

Hint Constructors step.
```

17.2 Subtyping

Now we come to the interesting part, where the features we've added start to interact. We begin by defining the subtyping relation and developing some of its important technical properties.

17.2.1 Definition

The definition of subtyping is essentially just what we sketched in the discussion of record subtyping in chapter Sub, but we need to add well-formedness side conditions to some of the rules. Also, we replace the "n-ary" width, depth, and permutation subtyping rules by binary rules that deal with just the first field.

```
Reserved Notation "T '<:' U" (at level 40).

Inductive subtype : ty → ty → Prop :=
```

```
  | S_Refl : ∀ T,
    well_formed_ty T →
    T <: T
  | S_Trans : ∀ S U T,
    S <: U →
    U <: T →
    S <: T
  | S_Top : ∀ S,
    well_formed_ty S →
    S <: Top
  | S_Arrow : ∀ S1 S2 T1 T2,
    T1 <: S1 →
    S2 <: T2 →
    Arrow S1 S2 <: Arrow T1 T2

  | S_RcdWidth : ∀ i T1 T2,
    well_formed_ty (RCons i T1 T2) →
    RCons i T1 T2 <: RNil
  | S_RcdDepth : ∀ i S1 T1 Sr2 Tr2,
    S1 <: T1 →
    Sr2 <: Tr2 →
    record_ty Sr2 →
    record_ty Tr2 →
    RCons i S1 Sr2 <: RCons i T1 Tr2
  | S_RcdPerm : ∀ i1 i2 T1 T2 Tr3,
    well_formed_ty (RCons i1 T1 (RCons i2 T2 Tr3)) →
    i1 ≠ i2 →
       RCons i1 T1 (RCons i2 T2 Tr3)
    <: RCons i2 T2 (RCons i1 T1 Tr3)

where "T '<:' U" := (subtype T U).

Hint Constructors subtype.
```

17.2.2 Examples

```
Module Examples.
Open Scope string_scope.

Notation x := "x".
Notation y := "y".
Notation z := "z".
Notation j := "j".
```

```
Notation k := "k".
Notation i := "i".
Notation A := (Base "A").
Notation B := (Base "B").
Notation C := (Base "C").

Definition TRcd_j :=
  (RCons j (Arrow B B) RNil). Definition TRcd_kj :=
  RCons k (Arrow A A) TRcd_j.

Example subtyping_example_0 :
  subtype (Arrow C TRcd_kj)
          (Arrow C RNil).
Proof.
  apply S_Arrow.
    apply S_Refl. auto.
    unfold TRcd_kj, TRcd_j. apply S_RcdWidth; auto.
Qed.
```

The following facts are mostly easy to prove in Coq. To get full benefit, make sure you also understand how to prove them on paper!

Exercise: 2 stars, standard (subtyping_example_1) Example subtyping_example_1 :

```
  subtype TRcd_kj TRcd_j.
Proof with eauto.
```

Admitted.

☐

Exercise: 1 star, standard (subtyping_example_2) Example subtyping_example_2 :

```
  subtype (Arrow Top TRcd_kj)
          (Arrow (Arrow C C) TRcd_j).
Proof with eauto.
```

Admitted.

☐

Exercise: 1 star, standard (subtyping_example_3) Example subtyping_example_3 :

```
  subtype (Arrow RNil (RCons j A RNil))
          (Arrow (RCons k B RNil) RNil).
Proof with eauto.
```

Admitted.

☐

Exercise: 2 stars, standard (subtyping_example_4) Example subtyping_example_4 :

```
  subtype (RCons x A (RCons y B (RCons z C RNil)))
```

(RCons z C (RCons y B (RCons x A RNil))).
Proof with eauto.
Admitted.
□

End Examples.

17.2.3 Properties of Subtyping

Well-Formedness

To get started proving things about subtyping, we need a couple of technical lemmas that intuitively (1) allow us to extract the well-formedness assumptions embedded in subtyping derivations and (2) record the fact that fields of well-formed record types are themselves well-formed types.

```
Lemma subtype__wf : ∀ S T,
  subtype S T →
  well_formed_ty T ∧ well_formed_ty S.
Proof with eauto.
  intros S T Hsub.
  induction Hsub;
    intros; try (destruct IHHsub1; destruct IHHsub2)...
  -
    split... inversion H. subst. inversion H5... Qed.
```

```
Lemma wf_rcd_lookup : ∀ i T Ti,
  well_formed_ty T →
  Tlookup i T = Some Ti →
  well_formed_ty Ti.
Proof with eauto.
  intros i T.
  induction T; intros; try solve_by_invert.
  -
    inversion H. subst. unfold Tlookup in H0.
    destruct (eqb_string i s)... inversion H0; subst... Qed.
```

Field Lookup

The record matching lemmas get a little more complicated in the presence of subtyping, for two reasons. First, record types no longer necessarily describe the exact structure of the corresponding terms. And second, reasoning by induction on typing derivations becomes harder in general, because typing is no longer syntax directed.

```
Lemma rcd_types_match : ∀ S T i Ti,
  subtype S T →
```

```
  Tlookup i T = Some Ti →
  ∃ Si, Tlookup i S = Some Si ∧ subtype Si Ti.
Proof with (eauto using wf_rcd_lookup).
  intros S T i Ti Hsub Hget. generalize dependent Ti.
  induction Hsub; intros Ti Hget;
    try solve_by_invert.
  -
    ∃ Ti...
  -
    destruct (IHHsub2 Ti) as [Ui Hui]... destruct Hui.
    destruct (IHHsub1 Ui) as [Si Hsi]... destruct Hsi.
    ∃ Si...
  -
    rename i0 into k.
    unfold Tlookup. unfold Tlookup in Hget.
    destruct (eqb_string i k)...
    +
      inversion Hget. subst. ∃ S1...
  -
    ∃ Ti. split.
    +
      unfold Tlookup. unfold Tlookup in Hget.
      destruct (eqb_stringP i i1)...
      ×
        destruct (eqb_stringP i i2)...
        destruct H0.
        subst...
    +
      inversion H. subst. inversion H5. subst... Qed.
```

Exercise: 3 stars, standard (rcd_types_match_informal) Write a careful informal proof of the rcd_types_match lemma.

```
Definition manual_grade_for_rcd_types_match_informal : option (nat×string) := None.
```

□

Inversion Lemmas

Exercise: 3 stars, standard, optional (sub_inversion_arrow) Lemma sub_inversion_arrow : ∀ U V1 V2,

```
     subtype U (Arrow V1 V2) →
     ∃ U1 U2,
       (U=(Arrow U1 U2)) ∧ (subtype V1 U1) ∧ (subtype U2 V2).
```

```
Proof with eauto.
  intros U V1 V2 Hs.
  remember (Arrow V1 V2) as V.
  generalize dependent V2. generalize dependent V1.
   Admitted.
```

□

17.3 Typing

```
Definition context := partial_map ty.

Reserved Notation "Gamma '|-' t '\in' T" (at level 40).

Inductive has_type : context → tm → ty → Prop :=
  | T_Var : ∀ Gamma x T,
      Gamma x = Some T →
      well_formed_ty T →
      Gamma ⊢ var x \in T
  | T_Abs : ∀ Gamma x T11 T12 t12,
      well_formed_ty T11 →
      update Gamma x T11 ⊢ t12 \in T12 →
      Gamma ⊢ abs x T11 t12 \in Arrow T11 T12
  | T_App : ∀ T1 T2 Gamma t1 t2,
      Gamma ⊢ t1 \in Arrow T1 T2 →
      Gamma ⊢ t2 \in T1 →
      Gamma ⊢ app t1 t2 \in T2
  | T_Proj : ∀ Gamma i t T Ti,
      Gamma ⊢ t \in T →
      Tlookup i T = Some Ti →
      Gamma ⊢ rproj t i \in Ti

  | T_Sub : ∀ Gamma t S T,
      Gamma ⊢ t \in S →
      subtype S T →
      Gamma ⊢ t \in T

  | T_RNil : ∀ Gamma,
      Gamma ⊢ rnil \in RNil
  | T_RCons : ∀ Gamma i t T tr Tr,
      Gamma ⊢ t \in T →
      Gamma ⊢ tr \in Tr →
      record_ty Tr →
      record_tm tr →
```

$Gamma \vdash$ rcons i t tr \in RCons i T Tr

where "Gamma '|-' t '\in' T" := (**has_type** $Gamma$ t T).

Hint Constructors **has_type**.

17.3.1 Typing Examples

Module EXAMPLES2.
Import *Examples*.

Exercise: 1 star, standard (typing_example_0) Definition trcd_kj :=

```
  (rcons k (abs z A (var z))
           (rcons j (abs z B (var z))
                      rnil)).
```

```
Example typing_example_0 :
  has_type empty
           (rcons k (abs z A (var z))
                     (rcons j (abs z B (var z))
                               rnil))
           TRcd_kj.
Proof.
   Admitted.
```

□

Exercise: 2 stars, standard (typing_example_1) Example typing_example_1 :

```
  has_type empty
           (app (abs x TRcd_j (rproj (var x) j))
                  (trcd_kj))
           (Arrow B B).
Proof with eauto.
   Admitted.
```

□

Exercise: 2 stars, standard, optional (typing_example_2) Example typing_example_2
:

```
  has_type empty
           (app (abs z (Arrow (Arrow C C) TRcd_j)
                          (rproj (app (var z)
                                          (abs x C (var x)))
                                   j))
                  (abs z (Arrow C C) trcd_kj))
```

```
            (Arrow B B).
Proof with eauto.
  Admitted.
  □
End Examples2.
```

17.3.2 Properties of Typing

Well-Formedness

```
Lemma has_type__wf : ∀ Gamma t T,
  has_type Gamma t T → well_formed_ty T.
Proof with eauto.
  intros Gamma t T Htyp.
  induction Htyp...
  -
    inversion IHHtyp1...
  -
    eapply wf_rcd_lookup...
  -
    apply subtype__wf in H.
    destruct H...
Qed.

Lemma step_preserves_record_tm : ∀ tr tr',
  record_tm tr →
  tr -> tr' →
  record_tm tr'.
Proof.
  intros tr tr' Hrt Hstp.
  inversion Hrt; subst; inversion Hstp; subst; eauto.
Qed.
```

Field Lookup

```
Lemma lookup_field_in_value : ∀ v T i Ti,
  value v →
  has_type empty v T →
  Tlookup i T = Some Ti →
  ∃ vi, tlookup i v = Some vi ∧ has_type empty vi Ti.
Proof with eauto.
  remember empty as Gamma.
  intros t T i Ti Hval Htyp. revert Ti HeqGamma Hval.
```

```
  induction Htyp; intros; subst; try solve_by_invert.
  -
    apply (rcd_types_match S) in H0...
    destruct H0 as [Si [HgetSi Hsub]].
    destruct (IHHtyp Si) as [vi [Hget Htyvi]]...
  -
    simpl in H0. simpl. simpl in H1.
    destruct (eqb_string i i0).
    +
      inversion H1. subst. ∃ t...
    +
      destruct (IHHtyp2 Ti) as [vi [get Htyvi]]...
      inversion Hval... Qed.
```

Progress

Exercise: 3 stars, standard (canonical_forms_of_arrow_types) Lemma canonical_forms_of_arrow_ty

```
: ∀ Gamma s T1 T2,
     has_type Gamma s (Arrow T1 T2) →
     value s →
     ∃ x S1 s2,
        s = abs x S1 s2.
Proof with eauto.
   Admitted.
   □

Theorem progress : ∀ t T,
     has_type empty t T →
     value t ∨ ∃ t', t -> t'.
Proof with eauto.
  intros t T Ht.
  remember empty as Gamma.
  revert HeqGamma.
  induction Ht;
    intros HeqGamma; subst...
  -
    inversion H.
  -
    right.
    destruct IHHt1; subst...
    +
      destruct IHHt2; subst...
      ×
        destruct (canonical_forms_of_arrow_types empty t1 T1 T2)
```

```
      as [x [S1 [t12 Heqt1]]]...
    subst. ∃ ([x:=t2]t12)...
  ×
    destruct H0 as [t2' Hstp]. ∃ (app t1 t2')...
  +
    destruct H as [t1' Hstp]. ∃ (app t1' t2)...
-
  right. destruct IHHt...
  +
    destruct (lookup_field_in_value t T i Ti)
      as [t' [Hget Ht']]...
  +
    destruct H0 as [t' Hstp]. ∃ (rproj t' i)...
-
  destruct IHHt1...
  +
    destruct IHHt2...
    ×
      right. destruct H2 as [tr' Hstp].
      ∃ (rcons i t tr')...
  +
    right. destruct H1 as [t' Hstp].
    ∃ (rcons i t' tr)... Qed.
```

Theorem : For any term *t* and type T, if empty ⊢ *t* : T then *t* is a value or *t* –> *t'* for some term *t'*.

Proof: Let *t* and T be given such that empty ⊢ *t* : T. We proceed by induction on the given typing derivation.

- The cases where the last step in the typing derivation is T_Abs or T_RNil are immediate because abstractions and {} are always values. The case for T_Var is vacuous because variables cannot be typed in the empty context.

- If the last step in the typing derivation is by T_App, then there are terms *t1 t2* and types *T1 T2* such that *t* = *t1 t2*, T = *T2*, empty ⊢ *t1* : *T1* → *T2* and empty ⊢ *t2* : *T1*.

  The induction hypotheses for these typing derivations yield that *t1* is a value or steps, and that *t2* is a value or steps.

  - Suppose *t1* –> *t1'* for some term *t1'*. Then *t1 t2* –> *t1' t2* by ST_App1.
  - Otherwise *t1* is a value.
    - Suppose *t2* –> *t2'* for some term *t2'*. Then *t1 t2* –> *t1 t2'* by rule ST_App2 because *t1* is a value.

    - Otherwise, *t2* is a value. By Lemma *canonical_forms_for_arrow_types*, *t1* = \x:*S1*.*s2* for some x, *S1*, and *s2*. But then (\x:*S1*.*s2*) *t2* –> [x:=*t2*]*s2* by ST_AppAbs, since *t2* is a value.

- If the last step of the derivation is by T_Proj, then there are a term *tr*, a type *Tr*, and a label i such that *t* = *tr*.*i*, empty ⊢ *tr* : *Tr*, and Tlookup i *Tr* = Some T.

  By the IH, either *tr* is a value or it steps. If *tr* –> *tr'* for some term *tr'*, then *tr*.*i* –> *tr'*.*i* by rule ST_Proj1.

  If *tr* is a value, then Lemma lookup_field_in_value yields that there is a term *ti* such that tlookup i *tr* = Some *ti*. It follows that *tr*.*i* –> *ti* by rule ST_ProjRcd.

- If the final step of the derivation is by T_Sub, then there is a type S such that S <: T and empty ⊢ *t* : S. The desired result is exactly the induction hypothesis for the typing subderivation.

- If the final step of the derivation is by T_RCons, then there exist some terms *t1* *tr*, types *T1* *Tr* and a label *t* such that *t* = {i=*t1*, *tr*}, T = {i:*T1*, *Tr*}, **record_ty** *tr*, **record_tm** *Tr*, empty ⊢ *t1* : *T1* and empty ⊢ *tr* : *Tr*.

  The induction hypotheses for these typing derivations yield that *t1* is a value or steps, and that *tr* is a value or steps. We consider each case:

  - Suppose *t1* –> *t1'* for some term *t1'*. Then {i=*t1*, *tr*} –> {i=*t1'*, *tr*} by rule ST_Rcd_Head.
  - Otherwise *t1* is a value.
    - Suppose *tr* –> *tr'* for some term *tr'*. Then {i=*t1*, *tr*} –> {i=*t1*, *tr'*} by rule ST_Rcd_Tail, since *t1* is a value.
    - Otherwise, *tr* is also a value. So, {i=*t1*, *tr*} is a value by v_rcons.

Inversion Lemmas

Lemma typing_inversion_var : ∀ *Gamma* *x* *T*,
  has_type *Gamma* (var *x*) *T* →
  ∃ *S* ,
    Gamma *x* = Some *S* ∧ **subtype** *S* *T*.
Proof with eauto.
  intros *Gamma* *x* *T* *Hty*.
  remember (var *x*) as *t*.
  induction *Hty*; intros;
    inversion *Heqt*; subst; try *solve_by_invert*.
  -
    ∃ *T*...

```
  -
    destruct IHHty as [U [Hctx HsubU]]... Qed.
Lemma typing_inversion_app : ∀ Gamma t1 t2 T2,
  has_type Gamma (app t1 t2) T2 →
  ∃ T1,
    has_type Gamma t1 (Arrow T1 T2) ∧
    has_type Gamma t2 T1.
Proof with eauto.
  intros Gamma t1 t2 T2 Hty.
  remember (app t1 t2) as t.
  induction Hty; intros;
    inversion Heqt; subst; try solve_by_invert.
  -
    ∃ T1...
  -
    destruct IHHty as [U1 [Hty1 Hty2]]...
    assert (Hwf := has_type__wf _ _ _ Hty2).
    ∃ U1... Qed.
Lemma typing_inversion_abs : ∀ Gamma x S1 t2 T,
     has_type Gamma (abs x S1 t2) T →
     (∃ S2, subtype (Arrow S1 S2) T
               ∧ has_type (update Gamma x S1) t2 S2).
Proof with eauto.
  intros Gamma x S1 t2 T H.
  remember (abs x S1 t2) as t.
  induction H;
    inversion Heqt; subst; intros; try solve_by_invert.
  -
    assert (Hwf := has_type__wf _ _ _ H0).
    ∃ T12...
  -
    destruct IHhas_type as [S2 [Hsub Hty]]...
    Qed.
Lemma typing_inversion_proj : ∀ Gamma i t1 Ti,
  has_type Gamma (rproj t1 i) Ti →
  ∃ T Si,
    Tlookup i T = Some Si ∧ subtype Si Ti ∧ has_type Gamma t1 T.
Proof with eauto.
  intros Gamma i t1 Ti H.
  remember (rproj t1 i) as t.
  induction H;
    inversion Heqt; subst; intros; try solve_by_invert.
```

```
  -
    assert (well_formed_ty Ti) as Hwf.
    {
      apply (wf_rcd_lookup i T Ti)...
      apply has_type__wf in H... }
    ∃ T, Ti...
  -
    destruct IHhas_type as [U [Ui [Hget [Hsub Hty]]]]...
    ∃ U, Ui... Qed.
Lemma typing_inversion_rcons : ∀ Gamma i ti tr T,
  has_type Gamma (rcons i ti tr) T →
  ∃ Si Sr,
    subtype (RCons i Si Sr) T ∧ has_type Gamma ti Si ∧
    record_tm tr ∧ has_type Gamma tr Sr.
Proof with eauto.
  intros Gamma i ti tr T Hty.
  remember (rcons i ti tr) as t.
  induction Hty;
    inversion Heqt; subst...
  -
    apply IHHty in H0.
    destruct H0 as [Ri [Rr [HsubRS [HtypRi HtypRr]]]].
    ∃ Ri, Rr...
  -
    assert (well_formed_ty (RCons i T Tr)) as Hwf.
    {
      apply has_type__wf in Hty1.
      apply has_type__wf in Hty2... }
    ∃ T, Tr... Qed.
Lemma abs_arrow : ∀ x S1 s2 T1 T2,
  has_type empty (abs x S1 s2) (Arrow T1 T2) →
     subtype T1 S1
  ∧ has_type (update empty x S1) s2 T2.
Proof with eauto.
  intros x S1 s2 T1 T2 Hty.
  apply typing_inversion_abs in Hty.
  destruct Hty as [S2 [Hsub Hty]].
  apply sub_inversion_arrow in Hsub.
  destruct Hsub as [U1 [U2 [Heq [Hsub1 Hsub2]]]].
  inversion Heq; subst... Qed.
```

Context Invariance

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
      appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
      appears_free_in x t1 → appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
      appears_free_in x t2 → appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
        y ≠ x →
        appears_free_in x t12 →
        appears_free_in x (abs y T11 t12)
  | afi_proj : ∀ x t i,
      appears_free_in x t →
      appears_free_in x (rproj t i)
  | afi_rhead : ∀ x i t tr,
      appears_free_in x t →
      appears_free_in x (rcons i t tr)
  | afi_rtail : ∀ x i t tr,
      appears_free_in x tr →
      appears_free_in x (rcons i t tr).

Hint Constructors appears_free_in.

Lemma context_invariance : ∀ Gamma Gamma' t S,
     has_type Gamma t S →
     (∀ x, appears_free_in x t → Gamma x = Gamma' x) →
     has_type Gamma' t S.
Proof with eauto.
  intros. generalize dependent Gamma'.
  induction H;
    intros Gamma' Heqv...
  -
    apply T_Var... rewrite ← Heqv...
  -
    apply T_Abs... apply IHhas_type. intros x0 Hafi.
    unfold update, t_update. destruct (eqb_stringP x x0)...
  -
    apply T_App with T1...
  -
    apply T_RCons... Qed.

Lemma free_in_context : ∀ x t T Gamma,
   appears_free_in x t →
```

```
   has_type Gamma t T →
   ∃ T', Gamma x = Some T'.
Proof with eauto.
  intros x t T Gamma Hafi Htyp.
  induction Htyp; subst; inversion Hafi; subst...
  -
    destruct (IHHtyp H5) as [T Hctx]. ∃ T.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx... Qed.
```

Preservation

```
Lemma substitution_preserves_typing : ∀ Gamma x U v t S,
     has_type (update Gamma x U) t S →
     has_type empty v U →
     has_type Gamma ([x:=v]t) S.
Proof with eauto.
  intros Gamma x U v t S Htypt Htypv.
  generalize dependent S. generalize dependent Gamma.
  induction t; intros; simpl.
  -
    rename s into y.
    destruct (typing_inversion_var _ _ _ Htypt) as [T [Hctx Hsub]].
    unfold update, t_update in Hctx.
    destruct (eqb_stringP x y)...
    +
      subst.
      inversion Hctx; subst. clear Hctx.
      apply context_invariance with empty...
      intros x Hcontra.
      destruct (free_in_context _ _ S empty Hcontra) as [T' HT']...
      inversion HT'.
    +
      destruct (subtype__wf _ _ Hsub)...
  -
    destruct (typing_inversion_app _ _ _ _ Htypt)
      as [T1 [Htypt1 Htypt2]].
    eapply T_App...
  -
    rename s into y. rename t into T1.
    destruct (typing_inversion_abs _ _ _ _ _ Htypt)
      as [T2 [Hsub Htypt2]].
    destruct (subtype__wf _ _ Hsub) as [Hwf1 Hwf2].
```

```
    inversion Hwf2. subst.
    apply T_Sub with (Arrow T1 T2)... apply T_Abs...
    destruct (eqb_stringP x y).
    +
      eapply context_invariance...
      subst.
      intros x Hafi. unfold update, t_update.
      destruct (eqb_string y x)...
    +
      apply IHt. eapply context_invariance...
      intros z Hafi. unfold update, t_update.
      destruct (eqb_stringP y z)...
      subst. rewrite false_eqb_string...
  -
    destruct (typing_inversion_proj _ _ _ _ Htypt)
      as [T [Ti [Hget [Hsub Htypt1]]]]...
  -
    eapply context_invariance...
    intros y Hcontra. inversion Hcontra.
  -
    destruct (typing_inversion_rcons _ _ _ _ _ Htypt) as
      [Ti [Tr [Hsub [HtypTi [Hrcdt2 HtypTr]]]]].
    apply T_Sub with (RCons s Ti Tr)...
    apply T_RCons...
    +
      apply subtype__wf in Hsub. destruct Hsub. inversion H0...
    +
      inversion Hrcdt2; subst; simpl... Qed.
Theorem preservation : ∀ t t' T,
     has_type empty t T →
     t -> t' →
     has_type empty t' T.
Proof with eauto.
  intros t t' T HT.
  remember empty as Gamma. generalize dependent HeqGamma.
  generalize dependent t'.
  induction HT;
    intros t' HeqGamma HE; subst; inversion HE; subst...
  -
    inversion HE; subst...
    +
      destruct (abs_arrow _ _ _ _ _ HT1) as [HA1 HA2].
```

```
      apply substitution_preserves_typing with T...
  -
    destruct (lookup_field_in_value _ _ _ _ H2 HT H)
      as [vi [Hget Hty]].
    rewrite H4 in Hget. inversion Hget. subst...
  -
    eauto using step_preserves_record_tm. Qed.
```

Theorem: If t, t' are terms and T is a type such that empty $\vdash t$: T and t –> t', then empty $\vdash t'$: T.

Proof: Let t and T be given such that empty $\vdash t$: T. We go by induction on the structure of this typing derivation, leaving t' general. Cases T_Abs and T_RNil are vacuous because abstractions and {} don't step. Case T_Var is vacuous as well, since the context is empty.

- If the final step of the derivation is by T_App, then there are terms $t1$ $t2$ and types $T1$ $T2$ such that $t = t1\ t2$, T $= T2$, empty $\vdash t1 : T1 \rightarrow T2$ and empty $\vdash t2 : T1$.

  By inspection of the definition of the step relation, there are three ways $t1$ $t2$ can step. Cases ST_App1 and ST_App2 follow immediately by the induction hypotheses for the typing subderivations and a use of T_App.

  Suppose instead $t1$ $t2$ steps by ST_AppAbs. Then $t1 = \backslash$x:$S.t12$ for some type S and term $t12$, and $t' =$ [x:=$t2$]$t12$.

  By Lemma abs_arrow, we have $T1$ <: S and x:$S1 \vdash s2 : T2$. It then follows by lemma substitution_preserves_typing that empty $\vdash$ [x:=$t2$] $t12 : T2$ as desired.

- If the final step of the derivation is by T_Proj, then there is a term tr, type Tr and label i such that $t = tr.i$, empty $\vdash tr : Tr$, and Tlookup i Tr = Some T.

  The IH for the typing derivation gives us that, for any term tr', if tr –> tr' then empty $\vdash tr'\ Tr$. Inspection of the definition of the step relation reveals that there are two ways a projection can step. Case ST_Proj1 follows immediately by the IH.

  Instead suppose $tr.i$ steps by ST_ProjRcd. Then tr is a value and there is some term vi such that tlookup i tr = Some vi and $t' = vi$. But by lemma lookup_field_in_value, empty $\vdash vi : Ti$ as desired.

- If the final step of the derivation is by T_Sub, then there is a type S such that S <: T and empty $\vdash t$: S. The result is immediate by the induction hypothesis for the typing subderivation and an application of T_Sub.

- If the final step of the derivation is by T_RCons, then there exist some terms $t1$ tr, types $T1$ Tr and a label t such that t = {i=$t1$, tr}, T = {i:$T1$, Tr}, **record_ty** tr, **record_tm** Tr, empty $\vdash t1 : T1$ and empty $\vdash tr : Tr$.

  By the definition of the step relation, t must have stepped by ST_Rcd_Head or ST_Rcd_Tail. In the first case, the result follows by the IH for $t1$'s typing derivation and T_RCons.

In the second case, the result follows by the IH for tr's typing derivation, T_RCons, and a use of the step_preserves_record_tm lemma.

Chapter 18

Norm: Normalization of STLC

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Lists.List. Import ListNotations.
From Coq Require Import Strings.String.
From PLF Require Import Maps.
From PLF Require Import Smallstep.
```

```
Hint Constructors multi.
```

This optional chapter is based on chapter 12 of *Types and Programming Languages* (Pierce). It may be useful to look at the two together, as that chapter includes explanations and informal proofs that are not repeated here.

In this chapter, we consider another fundamental theoretical property of the simply typed lambda-calculus: the fact that the evaluation of a well-typed program is guaranteed to halt in a finite number of steps—i.e., every well-typed term is *normalizable*.

Unlike the type-safety properties we have considered so far, the normalization property does not extend to full-blown programming languages, because these languages nearly always extend the simply typed lambda-calculus with constructs, such as general recursion (see the MoreStlc chapter) or recursive types, that can be used to write nonterminating programs. However, the issue of normalization reappears at the level of *types* when we consider the metatheory of polymorphic versions of the lambda calculus such as System F-omega: in this system, the language of types effectively contains a copy of the simply typed lambda-calculus, and the termination of the typechecking algorithm will hinge on the fact that a "normalization" operation on type expressions is guaranteed to terminate.

Another reason for studying normalization proofs is that they are some of the most beautiful—and mind-blowing—mathematics to be found in the type theory literature, often (as here) involving the fundamental proof technique of *logical relations*.

The calculus we shall consider here is the simply typed lambda-calculus over a single base type **bool** and with pairs. We'll give most details of the development for the basic lambda-calculus terms treating **bool** as an uninterpreted base type, and leave the extension to the boolean operators and pairs to the reader. Even for the base calculus, normalization is not entirely trivial to prove, since each reduction of a term can duplicate redexes in subterms.

Exercise: 2 stars, standard (norm_fail) Where do we fail if we attempt to prove normalization by a straightforward induction on the size of a well-typed term?

```
Definition manual_grade_for_norm_fail : option (nat×string) := None.
```
□

Exercise: 5 stars, standard, recommended (norm) The best ways to understand an intricate proof like this is are (1) to help fill it in and (2) to extend it. We've left out some parts of the following development, including some proofs of lemmas and the all the cases involving products and conditionals. Fill them in.

```
Definition manual_grade_for_norm : option (nat×string) := None.
```
□

18.1 Language

We begin by repeating the relevant language definition, which is similar to those in the MoreStlc chapter, plus supporting results including type preservation and step determinism. (We won't need progress.) You may just wish to skip down to the Normalization section...

Syntax and Operational Semantics

```
Inductive ty : Type :=
  | Bool : ty
  | Arrow : ty → ty → ty
  | Prod : ty → ty → ty
.
Inductive tm : Type :=

  | var : string → tm
  | app : tm → tm → tm
  | abs : string → ty → tm → tm

  | pair : tm → tm → tm
  | fst : tm → tm
  | snd : tm → tm

  | tru : tm
  | fls : tm
  | test : tm → tm → tm → tm.
```

Substitution

```
Fixpoint subst (x:string) (s:tm) (t:tm) : tm :=
  match t with
  | var y ⇒ if eqb_string x y then s else t
  | abs y T t1 ⇒
      abs y T (if eqb_string x y then t1 else (subst x s t1))
  | app t1 t2 ⇒ app (subst x s t1) (subst x s t2)
  | pair t1 t2 ⇒ pair (subst x s t1) (subst x s t2)
  | fst t1 ⇒ fst (subst x s t1)
  | snd t1 ⇒ snd (subst x s t1)
  | tru ⇒ tru
  | fls ⇒ fls
  | test t0 t1 t2 ⇒
      test (subst x s t0) (subst x s t1) (subst x s t2)
  end.
Notation "'[' x ':=' s ']' t" := (subst x s t) (at level 20).
```

Reduction

```
Inductive value : tm → Prop :=
  | v_abs : ∀ x T11 t12,
      value (abs x T11 t12)
  | v_pair : ∀ v1 v2,
      value v1 →
      value v2 →
      value (pair v1 v2)
  | v_tru : value tru
  | v_fls : value fls
.
Hint Constructors value.
Reserved Notation "t1 '-->' t2" (at level 40).
Inductive step : tm → tm → Prop :=
  | ST_AppAbs : ∀ x T11 t12 v2,
          value v2 →
          (app (abs x T11 t12) v2) -> [x:=v2]t12
  | ST_App1 : ∀ t1 t1' t2,
          t1 -> t1' →
          (app t1 t2) -> (app t1' t2)
  | ST_App2 : ∀ v1 t2 t2',
          value v1 →
          t2 -> t2' →
```

```
        (app v1 t2) -> (app v1 t2')

  | ST_Pair1 : ∀ t1 t1' t2,
        t1 -> t1' →
        (pair t1 t2) -> (pair t1' t2)
  | ST_Pair2 : ∀ v1 t2 t2',
        value v1 →
        t2 -> t2' →
        (pair v1 t2) -> (pair v1 t2')
  | ST_Fst : ∀ t1 t1',
        t1 -> t1' →
        (fst t1) -> (fst t1')
  | ST_FstPair : ∀ v1 v2,
        value v1 →
        value v2 →
        (fst (pair v1 v2)) -> v1
  | ST_Snd : ∀ t1 t1',
        t1 -> t1' →
        (snd t1) -> (snd t1')
  | ST_SndPair : ∀ v1 v2,
        value v1 →
        value v2 →
        (snd (pair v1 v2)) -> v2

  | ST_TestTrue : ∀ t1 t2,
        (test tru t1 t2) -> t1
  | ST_TestFalse : ∀ t1 t2,
        (test fls t1 t2) -> t2
  | ST_Test : ∀ t0 t0' t1 t2,
        t0 -> t0' →
        (test t0 t1 t2) -> (test t0' t1 t2)

where "t1 '->' t2" := (step t1 t2).

Notation multistep := (multi step).
Notation "t1 '->*' t2" := (multistep t1 t2) (at level 40).

Hint Constructors step.

Notation step_normal_form := (normal_form step).

Lemma value__normal : ∀ t, value t → step_normal_form t.
Proof with eauto.
  intros t H; induction H; intros [t' ST]; inversion ST...
Qed.
```

Typing

```
Definition context := partial_map ty.
Inductive has_type : context → tm → ty → Prop :=

  | T_Var : ∀ Gamma x T,
      Gamma x = Some T →
      has_type Gamma (var x) T
  | T_Abs : ∀ Gamma x T11 T12 t12,
      has_type (update Gamma x T11) t12 T12 →
      has_type Gamma (abs x T11 t12) (Arrow T11 T12)
  | T_App : ∀ T1 T2 Gamma t1 t2,
      has_type Gamma t1 (Arrow T1 T2) →
      has_type Gamma t2 T1 →
      has_type Gamma (app t1 t2) T2

  | T_Pair : ∀ Gamma t1 t2 T1 T2,
      has_type Gamma t1 T1 →
      has_type Gamma t2 T2 →
      has_type Gamma (pair t1 t2) (Prod T1 T2)
  | T_Fst : ∀ Gamma t T1 T2,
      has_type Gamma t (Prod T1 T2) →
      has_type Gamma (fst t) T1
  | T_Snd : ∀ Gamma t T1 T2,
      has_type Gamma t (Prod T1 T2) →
      has_type Gamma (snd t) T2

  | T_True : ∀ Gamma,
      has_type Gamma tru Bool
  | T_False : ∀ Gamma,
      has_type Gamma fls Bool
  | T_Test : ∀ Gamma t0 t1 t2 T,
      has_type Gamma t0 Bool →
      has_type Gamma t1 T →
      has_type Gamma t2 T →
      has_type Gamma (test t0 t1 t2) T
.

Hint Constructors has_type.

Hint Extern 2 (has_type _ (app _ _) _) ⇒ eapply T_App; auto.
Hint Extern 2 (_ = _) ⇒ compute; reflexivity.
```

Context Invariance

```
Inductive appears_free_in : string → tm → Prop :=
  | afi_var : ∀ x,
      appears_free_in x (var x)
  | afi_app1 : ∀ x t1 t2,
      appears_free_in x t1 → appears_free_in x (app t1 t2)
  | afi_app2 : ∀ x t1 t2,
      appears_free_in x t2 → appears_free_in x (app t1 t2)
  | afi_abs : ∀ x y T11 t12,
        y ≠ x →
        appears_free_in x t12 →
        appears_free_in x (abs y T11 t12)

  | afi_pair1 : ∀ x t1 t2,
      appears_free_in x t1 →
      appears_free_in x (pair t1 t2)
  | afi_pair2 : ∀ x t1 t2,
      appears_free_in x t2 →
      appears_free_in x (pair t1 t2)
  | afi_fst : ∀ x t,
      appears_free_in x t →
      appears_free_in x (fst t)
  | afi_snd : ∀ x t,
      appears_free_in x t →
      appears_free_in x (snd t)

  | afi_test0 : ∀ x t0 t1 t2,
      appears_free_in x t0 →
      appears_free_in x (test t0 t1 t2)
  | afi_test1 : ∀ x t0 t1 t2,
      appears_free_in x t1 →
      appears_free_in x (test t0 t1 t2)
  | afi_test2 : ∀ x t0 t1 t2,
      appears_free_in x t2 →
      appears_free_in x (test t0 t1 t2)
.

Hint Constructors appears_free_in.

Definition closed (t:tm) :=
  ∀ x, ¬ appears_free_in x t.

Lemma context_invariance : ∀ Gamma Gamma' t S,
     has_type Gamma t S →
```

```
     (∀ x, appears_free_in x t → Gamma x = Gamma' x) →
     has_type Gamma' t S.
Proof with eauto.
  intros. generalize dependent Gamma'.
  induction H;
    intros Gamma' Heqv...
  -
    apply T_Var... rewrite ← Heqv...
  -
    apply T_Abs... apply IHhas_type. intros y Hafi.
    unfold update, t_update. destruct (eqb_stringP x y)...
  -
    apply T_Pair...
  -
    eapply T_Test...
Qed.

Lemma free_in_context : ∀ x t T Gamma,
   appears_free_in x t →
   has_type Gamma t T →
   ∃ T', Gamma x = Some T'.
Proof with eauto.
  intros x t T Gamma Hafi Htyp.
  induction Htyp; inversion Hafi; subst...
  -
    destruct IHHtyp as [T' Hctx]... ∃ T'.
    unfold update, t_update in Hctx.
    rewrite false_eqb_string in Hctx...
Qed.

Corollary typable_empty__closed : ∀ t T,
    has_type empty t T →
    closed t.
Proof.
  intros. unfold closed. intros x H1.
  destruct (free_in_context _ _ _ _ H1 H) as [T' C].
  inversion C. Qed.
```

Preservation

```
Lemma substitution_preserves_typing : ∀ Gamma x U v t S,
     has_type (update Gamma x U) t S →
     has_type empty v U →
     has_type Gamma ([x:=v]t) S.
```

```
Proof with eauto.
  intros Gamma x U v t S Htypt Htypv.
  generalize dependent Gamma. generalize dependent S.
  induction t;
    intros S Gamma Htypt; simpl; inversion Htypt; subst...
  -
    simpl. rename s into y.
    unfold update, t_update in H1.
    destruct (eqb_stringP x y).
    +

      subst.
      inversion H1; subst. clear H1.
      eapply context_invariance...
      intros x Hcontra.
      destruct (free_in_context _ _ S empty Hcontra) as [T' HT']...
      inversion HT'.
    +

      apply T_Var...
  -
    rename s into y. rename t into T11.
    apply T_Abs...
    destruct (eqb_stringP x y).
    +

      eapply context_invariance...
      subst.
      intros x Hafi. unfold update, t_update.
      destruct (eqb_string y x)...
    +

      apply IHt. eapply context_invariance...
      intros z Hafi. unfold update, t_update.
      destruct (eqb_stringP y z)...
      subst. rewrite false_eqb_string...
Qed.

Theorem preservation : ∀ t t' T,
     has_type empty t T →
     t -> t' →
     has_type empty t' T.
Proof with eauto.
```

```
  intros t t' T HT.
  remember (@empty ty) as Gamma. generalize dependent HeqGamma.
  generalize dependent t'.
  induction HT;
    intros t' HeqGamma HE; subst; inversion HE; subst...
  -

    inversion HE; subst...
    +

      apply substitution_preserves_typing with T1...
      inversion HT1...
  -
    inversion HT...
  -
    inversion HT...
Qed.
```

Determinism

```
Lemma step_deterministic :
   deterministic step.
Proof with eauto.
   unfold deterministic.
   intros t t' t'' E1 E2.
   generalize dependent t''.
   induction E1; intros t'' E2; inversion E2; subst; clear E2...
   - inversion H3.
   - exfalso; apply value__normal in H...
   - inversion E1.
   - f_equal...
   - exfalso; apply value__normal in H1...
   - exfalso; apply value__normal in H3...
   - exfalso; apply value__normal in H...
   - f_equal...
   - f_equal...
   - exfalso; apply value__normal in H1...
   - exfalso; apply value__normal in H...
   - f_equal...
   - f_equal...
   - exfalso.
     inversion E1; subst.
     + apply value__normal in H0...
```

```
    + apply value__normal in H1...
  - exfalso.
    inversion H2; subst.
    + apply value__normal in H...
    + apply value__normal in H0...
  - f_equal...
  - exfalso.
    inversion E1; subst.
    + apply value__normal in H0...
    + apply value__normal in H1...
  - exfalso.
    inversion H2; subst.
    + apply value__normal in H...
    + apply value__normal in H0...
  -
      inversion H3.
  -
      inversion H3.
  - inversion E1.
  - inversion E1.
  - f_equal...
Qed.
```

18.2 Normalization

Now for the actual normalization proof.

Our goal is to prove that every well-typed term reduces to a normal form. In fact, it turns out to be convenient to prove something slightly stronger, namely that every well-typed term reduces to a *value*. This follows from the weaker property anyway via Progress (why?) but otherwise we don't need Progress, and we didn't bother re-proving it above.

Here's the key definition:

```
Definition halts (t:tm) : Prop := ∃ t', t ->* t' ∧ value t'.
```

A trivial fact:

```
Lemma value_halts : ∀ v, value v → halts v.
Proof.
  intros v H. unfold halts.
  ∃ v. split.
  apply multi_refl.
  assumption.
Qed.
```

The key issue in the normalization proof (as in many proofs by induction) is finding a strong enough induction hypothesis. To this end, we begin by defining, for each type T, a set *R_T* of closed terms of type T. We will specify these sets using a relation R and write R T t when t is in *R_T*. (The sets *R_T* are sometimes called *saturated sets* or *reducibility candidates*.)

Here is the definition of R for the base language:

- R **bool** t iff t is a closed term of type **bool** and t halts in a value
- R (*T1* → *T2*) t iff t is a closed term of type *T1* → *T2* and t halts in a value *and* for any term s such that R *T1* s, we have R *T2* (t s).

This definition gives us the strengthened induction hypothesis that we need. Our primary goal is to show that all *programs* —i.e., all closed terms of base type—halt. But closed terms of base type can contain subterms of functional type, so we need to know something about these as well. Moreover, it is not enough to know that these subterms halt, because the application of a normalized function to a normalized argument involves a substitution, which may enable more reduction steps. So we need a stronger condition for terms of functional type: not only should they halt themselves, but, when applied to halting arguments, they should yield halting results.

The form of R is characteristic of the *logical relations* proof technique. (Since we are just dealing with unary relations here, we could perhaps more properly say *logical properties*.) If we want to prove some property *P* of all closed terms of type A, we proceed by proving, by induction on types, that all terms of type A *possess* property *P*, all terms of type A→A *preserve* property *P*, all terms of type (A→A)->(A→A) *preserve the property of preserving* property *P*, and so on. We do this by defining a family of properties, indexed by types. For the base type A, the property is just *P*. For functional types, it says that the function should map values satisfying the property at the input type to values satisfying the property at the output type.

When we come to formalize the definition of R in Coq, we hit a problem. The most obvious formulation would be as a parameterized Inductive proposition like this:

Inductive R : ty -> tm -> Prop := | R_bool : forall b t, has_type empty t Bool -> halts t -> R Bool t | R_arrow : forall T1 T2 t, has_type empty t (Arrow T1 T2) -> halts t -> (forall s, R T1 s -> R T2 (app t s)) -> R (Arrow T1 T2) t.

Unfortunately, Coq rejects this definition because it violates the *strict positivity requirement* for inductive definitions, which says that the type being defined must not occur to the left of an arrow in the type of a constructor argument. Here, it is the third argument to *R_arrow*, namely (∀ s, R *T1* s → R *TS* (app t s)), and specifically the R *T1* s part, that violates this rule. (The outermost arrows separating the constructor arguments don't count when applying this rule; otherwise we could never have genuinely inductive properties at all!) The reason for the rule is that types defined with non-positive recursion can be used to build non-terminating functions, which as we know would be a disaster for Coq's logical soundness. Even though the relation we want in this case might be perfectly innocent, Coq still rejects it because it fails the positivity test.

Fortunately, it turns out that we *can* define R using a Fixpoint:

```
Fixpoint R (T:ty) (t:tm) {struct T} : Prop :=
  has_type empty t T ∧ halts t ∧
  (match T with
   | Bool ⇒ True
   | Arrow T1 T2 ⇒ (∀ s, R T1 s → R T2 (app t s))


   | Prod T1 T2 ⇒ False
   end).
```

As immediate consequences of this definition, we have that every element of every set R_T halts in a value and is closed with type t :

```
Lemma R_halts : ∀ {T} {t}, R T t → halts t.
Proof.
  intros. destruct T; unfold R in H; inversion H; inversion H1; assumption.
Qed.

Lemma R_typable_empty : ∀ {T} {t}, R T t → has_type empty t T.
Proof.
  intros. destruct T; unfold R in H; inversion H; inversion H1; assumption.
Qed.
```

Now we proceed to show the main result, which is that every well-typed term of type T is an element of R_T. Together with R_halts, that will show that every well-typed term halts in a value.

18.2.1 Membership in R_T Is Invariant Under Reduction

We start with a preliminary lemma that shows a kind of strong preservation property, namely that membership in R_T is *invariant* under reduction. We will need this property in both directions, i.e., both to show that a term in R_T stays in R_T when it takes a forward step, and to show that any term that ends up in R_T after a step must have been in R_T to begin with.

First of all, an easy preliminary lemma. Note that in the forward direction the proof depends on the fact that our language is determinstic. This lemma might still be true for nondeterministic languages, but the proof would be harder!

```
Lemma step_preserves_halting : ∀ t t', (t -> t') → (halts t ↔ halts t').
Proof.
 intros t t' ST. unfold halts.
 split.
 -
  intros [t'' [STM V]].
  inversion STM; subst.
```

```
    exfalso. apply value__normal in V. unfold normal_form in V. apply V. ∃ t'. auto.
    rewrite (step_deterministic _ _ _ ST H). ∃ t''. split; assumption.
  -
   intros [t'0 [STM V]].
   ∃ t'0. split; eauto.
Qed.
```

Now the main lemma, which comes in two parts, one for each direction. Each proceeds by induction on the structure of the type T. In fact, this is where we make fundamental use of the structure of types.

One requirement for staying in R_T is to stay in type T. In the forward direction, we get this from ordinary type Preservation.

```
Lemma step_preserves_R : ∀ T t t', (t -> t') → R T t → R T t'.
Proof.
 induction T; intros t t' E Rt; unfold R; fold R; unfold R in Rt; fold R in Rt;
               destruct Rt as [typable_empty_t [halts_t RRt]].
  split. eapply preservation; eauto.
  split. apply (step_preserves_halting _ _ E); eauto.
  auto.
  split. eapply preservation; eauto.
  split. apply (step_preserves_halting _ _ E); eauto.
  intros.
  eapply IHT2.
  apply ST_App1. apply E.
  apply RRt; auto.
   Admitted.
```

The generalization to multiple steps is trivial:

```
Lemma multistep_preserves_R : ∀ T t t',
  (t ->* t') → R T t → R T t'.
Proof.
  intros T t t' STM; induction STM; intros.
  assumption.
  apply IHSTM. eapply step_preserves_R. apply H. assumption.
Qed.
```

In the reverse direction, we must add the fact that t has type T before stepping as an additional hypothesis.

```
Lemma step_preserves_R' : ∀ T t t',
  has_type empty t T → (t -> t') → R T t' → R T t.
Proof.
   Admitted.

Lemma multistep_preserves_R' : ∀ T t t',
  has_type empty t T → (t ->* t') → R T t' → R T t.
```

```
Proof.
  intros T t t' HT STM.
  induction STM; intros.
    assumption.
    eapply step_preserves_R'. assumption. apply H. apply IHSTM.
    eapply preservation; eauto. auto.
Qed.
```

18.2.2 Closed Instances of Terms of Type t Belong to R_T

Now we proceed to show that every term of type T belongs to R_T. Here, the induction will be on typing derivations (it would be surprising to see a proof about well-typed terms that did not somewhere involve induction on typing derivations!). The only technical difficulty here is in dealing with the abstraction case. Since we are arguing by induction, the demonstration that a term abs x *T1 t2* belongs to $R\_(T1 \to T2)$ should involve applying the induction hypothesis to show that *t2* belongs to $R\_(T2)$. But $R\_(T2)$ is defined to be a set of *closed* terms, while *t2* may contain x free, so this does not make sense.

This problem is resolved by using a standard trick to suitably generalize the induction hypothesis: instead of proving a statement involving a closed term, we generalize it to cover all closed *instances* of an open term *t*. Informally, the statement of the lemma will look like this:

If $x1{:}T1,..xn{:}Tn \vdash t$: T and *v1*,...,*vn* are values such that R *T1 v1*, R *T2 v2*, ..., R *Tn vn*, then R T ([$x1{:=}v1$][$x2{:=}v2$]...[$xn{:=}vn$]t).

The proof will proceed by induction on the typing derivation $x1{:}T1,..xn{:}Tn \vdash t$: T; the most interesting case will be the one for abstraction.

Multisubstitutions, Multi-Extensions, and Instantiations

However, before we can proceed to formalize the statement and proof of the lemma, we'll need to build some (rather tedious) machinery to deal with the fact that we are performing *multiple* substitutions on term *t* and *multiple* extensions of the typing context. In particular, we must be precise about the order in which the substitutions occur and how they act on each other. Often these details are simply elided in informal paper proofs, but of course Coq won't let us do that. Since here we are substituting closed terms, we don't need to worry about how one substitution might affect the term put in place by another. But we still do need to worry about the *order* of substitutions, because it is quite possible for the same identifier to appear multiple times among the *x1*,...*xn* with different associated *vi* and *Ti*.

To make everything precise, we will assume that environments are extended from left to right, and multiple substitutions are performed from right to left. To see that this is consistent, suppose we have an environment written as ...,y:**bool**,...,y:**nat**,... and a corresponding term substitution written as ...[y:=(*tbool* true)]...[y:=(const 3)]...*t*. Since environments are extended from left to right, the binding y:**nat** hides the binding y:**bool**; since substitutions are performed right to left, we do the substitution y:=(const 3) first, so that the substitution y:=(*tbool* true) has no effect. Substitution thus correctly preserves the type of the term.

With these points in mind, the following definitions should make sense.

A *multisubstitution* is the result of applying a list of substitutions, which we call an *environment*.

```
Definition env := list (string × tm).

Fixpoint msubst (ss:env) (t:tm) {struct ss} : tm :=
match ss with
| nil ⇒ t
| ((x,s)::ss') ⇒ msubst ss' ([x:=s]t)
end.
```

We need similar machinery to talk about repeated extension of a typing context using a list of (identifier, type) pairs, which we call a *type assignment*.

```
Definition tass := list (string × ty).

Fixpoint mupdate (Gamma : context) (xts : tass) :=
  match xts with
  | nil ⇒ Gamma
  | ((x,v)::xts') ⇒ update (mupdate Gamma xts') x v
  end.
```

We will need some simple operations that work uniformly on environments and type assigments

```
Fixpoint lookup {X:Set} (k : string) (l : list (string × X)) {struct l}
              : option X :=
  match l with
    | nil ⇒ None
    | (j,x) :: l' ⇒
      if eqb_string j k then Some x else lookup k l'
  end.

Fixpoint drop {X:Set} (n:string) (nxs:list (string × X)) {struct nxs}
            : list (string × X) :=
  match nxs with
    | nil ⇒ nil
    | ((n',x)::nxs') ⇒
        if eqb_string n' n then drop n nxs'
        else (n',x)::(drop n nxs')
  end.
```

An *instantiation* combines a type assignment and a value environment with the same domains, where corresponding elements are in R.

```
Inductive instantiation : tass → env → Prop :=
| V_nil :
```

```
    instantiation nil nil
| V_cons : ∀ x T v c e,
    value v → R T v →
    instantiation c e →
    instantiation ((x,T)::c) ((x,v)::e).
```

We now proceed to prove various properties of these definitions.

More Substitution Facts

First we need some additional lemmas on (ordinary) substitution.

```
Lemma vacuous_substitution : ∀ t x,
     ¬ appears_free_in x t →
     ∀ t', [x:=t']t = t.
Proof with eauto.
  Admitted.

Lemma subst_closed: ∀ t,
     closed t →
     ∀ x t', [x:=t']t = t.
Proof.
  intros. apply vacuous_substitution. apply H. Qed.

Lemma subst_not_afi : ∀ t x v,
    closed v → ¬ appears_free_in x ([x:=v]t).
Proof with eauto.   unfold closed, not.
  induction t; intros x v P A; simpl in A.
    -
     destruct (eqb_stringP x s)...
     inversion A; subst. auto.
    -
     inversion A; subst...
    -
     destruct (eqb_stringP x s)...
     + inversion A; subst...
     + inversion A; subst...
    -
     inversion A; subst...
    -
     inversion A; subst...
    -
     inversion A; subst...
    -
     inversion A.
    -
```

```
    inversion A.
  -
    inversion A; subst...
Qed.
```

```
Lemma duplicate_subst : ∀ t' x t v,
  closed v → [x:=t]([x:=v]t') = [x:=v]t'.
Proof.
  intros. eapply vacuous_substitution. apply subst_not_afi. auto.
Qed.
```

```
Lemma swap_subst : ∀ t x x1 v v1,
    x ≠ x1 →
    closed v → closed v1 →
    [x1:=v1]([x:=v]t) = [x:=v]([x1:=v1]t).
Proof with eauto.
 induction t; intros; simpl.
  -
   destruct (eqb_stringP x s); destruct (eqb_stringP x1 s).
   + subst. exfalso...
   + subst. simpl. rewrite ← eqb_string_refl. apply subst_closed...
   + subst. simpl. rewrite ← eqb_string_refl. rewrite subst_closed...
   + simpl. rewrite false_eqb_string... rewrite false_eqb_string...
   Admitted.
```

Properties of Multi-Substitutions

```
Lemma msubst_closed: ∀ t, closed t → ∀ ss, msubst ss t = t.
Proof.
  induction ss.
    reflexivity.
    destruct a. simpl. rewrite subst_closed; assumption.
Qed.
```

Closed environments are those that contain only closed terms.

```
Fixpoint closed_env (env:env) {struct env} :=
  match env with
  | nil ⇒ True
  | (x,t)::env' ⇒ closed t ∧ closed_env env'
  end.
```

Next come a series of lemmas charcterizing how msubst of closed terms distributes over subst and over each term form

```
Lemma subst_msubst: ∀ env x v t, closed v → closed_env env →
    msubst env ([x:=v]t) = [x:=v](msubst (drop x env) t).
```

```
Proof.
  induction env0; intros; auto.
  destruct a. simpl.
  inversion H0. fold closed_env in H2.
  destruct (eqb_stringP s x).
  - subst. rewrite duplicate_subst; auto.
  - simpl. rewrite swap_subst; eauto.
Qed.

Lemma msubst_var: ∀ ss x, closed_env ss →
   msubst ss (var x) =
   match lookup x ss with
   | Some t ⇒ t
   | None ⇒ var x
  end.
Proof.
  induction ss; intros.
    reflexivity.
    destruct a.
     simpl. destruct (eqb_string s x).
      apply msubst_closed. inversion H; auto.
      apply IHss. inversion H; auto.
Qed.

Lemma msubst_abs: ∀ ss x T t,
  msubst ss (abs x T t) = abs x T (msubst (drop x ss) t).
Proof.
  induction ss; intros.
    reflexivity.
    destruct a.
      simpl. destruct (eqb_string s x); simpl; auto.
Qed.

Lemma msubst_app : ∀ ss t1 t2, msubst ss (app t1 t2) = app (msubst ss t1) (msubst ss t2).
Proof.
 induction ss; intros.
   reflexivity.
   destruct a.
    simpl. rewrite ← IHss. auto.
Qed.
```

You'll need similar functions for the other term constructors.

Properties of Multi-Extensions

We need to connect the behavior of type assignments with that of their corresponding contexts.

```
Lemma mupdate_lookup : ∀ (c : tass) (x:string),
    lookup x c = (mupdate empty c) x.
Proof.
  induction c; intros.
    auto.
    destruct a. unfold lookup, mupdate, update, t_update. destruct (eqb_string s x);
auto.
Qed.
```

```
Lemma mupdate_drop : ∀ (c: tass) Gamma x x',
      mupdate Gamma (drop x c) x'
    = if eqb_string x x' then Gamma x' else mupdate Gamma c x'.
Proof.
  induction c; intros.
  - destruct (eqb_stringP x x'); auto.
  - destruct a. simpl.
    destruct (eqb_stringP s x).
    + subst. rewrite IHc.
      unfold update, t_update. destruct (eqb_stringP x x'); auto.
    + simpl. unfold update, t_update. destruct (eqb_stringP s x'); auto.
      subst. rewrite false_eqb_string; congruence.
Qed.
```

Properties of Instantiations

These are strightforward.

```
Lemma instantiation_domains_match: ∀ {c} {e},
    instantiation c e →
    ∀ {x} {T},
      lookup x c = Some T → ∃ t, lookup x e = Some t.
Proof.
  intros c e V. induction V; intros x0 T0 C.
    solve_by_invert.
    simpl in *.
    destruct (eqb_string x x0); eauto.
Qed.
```

```
Lemma instantiation_env_closed : ∀ c e,
  instantiation c e → closed_env e.
Proof.
```

```
  intros c e V; induction V; intros.
    econstructor.
    unfold closed_env. fold closed_env.
    split. eapply typable_empty__closed. eapply R_typable_empty. eauto.
        auto.
Qed.

Lemma instantiation_R : ∀ c e,
    instantiation c e →
    ∀ x t T,
      lookup x c = Some T →
      lookup x e = Some t → R T t.
Proof.
  intros c e V. induction V; intros x' t' T' G E.
    solve_by_invert.
    unfold lookup in *. destruct (eqb_string x x').
      inversion G; inversion E; subst. auto.
      eauto.
Qed.

Lemma instantiation_drop : ∀ c env,
    instantiation c env →
    ∀ x, instantiation (drop x c) (drop x env).
Proof.
  intros c e V. induction V.
    intros. simpl. constructor.
    intros. unfold drop. destruct (eqb_string x x0); auto. constructor; eauto.
Qed.
```

Congruence Lemmas on Multistep

We'll need just a few of these; add them as the demand arises.

```
Lemma multistep_App2 : ∀ v t t',
  value v → (t ->* t') → (app v t) ->* (app v t').
Proof.
  intros v t t' V STM. induction STM.
   apply multi_refl.
   eapply multi_step.
     apply ST_App2; eauto. auto.
Qed.
```

The R Lemma.

We can finally put everything together.

The key lemma about preservation of typing under substitution can be lifted to multi-substitutions:

```
Lemma msubst_preserves_typing : ∀ c e,
     instantiation c e →
     ∀ Gamma t S, has_type (mupdate Gamma c) t S →
     has_type Gamma (msubst e t) S.
Proof.
  induction 1; intros.
    simpl in H. simpl. auto.
    simpl in H2. simpl.
    apply IHinstantiation.
    eapply substitution_preserves_typing; eauto.
    apply (R_typable_empty H0).
Qed.
```

And at long last, the main lemma.

```
Lemma msubst_R : ∀ c env t T,
    has_type (mupdate empty c) t T →
    instantiation c env →
    R T (msubst env t).
Proof.
  intros c env0 t T HT V.
  generalize dependent env0.
  remember (mupdate empty c) as Gamma.
  assert (∀ x, Gamma x = lookup x c).
    intros. rewrite HeqGamma. rewrite mupdate_lookup. auto.
  clear HeqGamma.
  generalize dependent c.
  induction HT; intros.

  -
   rewrite H0 in H. destruct (instantiation_domains_match V H) as [t P].
   eapply instantiation_R; eauto.
   rewrite msubst_var. rewrite P. auto. eapply instantiation_env_closed; eauto.

  -
    rewrite msubst_abs.
    assert (WT: has_type empty (abs x T11 (msubst (drop x env0) t12)) (Arrow T11
T12)).
    { eapply T_Abs. eapply msubst_preserves_typing.
      { eapply instantiation_drop; eauto. }
      eapply context_invariance.
      { apply HT. }
      intros.
```

```
      unfold update, t_update. rewrite mupdate_drop. destruct (eqb_stringP x x0).
      + auto.
      + rewrite H.
        clear - c n. induction c.
        simpl. rewrite false_eqb_string; auto.
        simpl. destruct a. unfold update, t_update.
        destruct (eqb_string s x0); auto. }
    unfold R. fold R. split.
      auto.
     split. apply value_halts. apply v_abs.
     intros.
     destruct (R_halts H0) as [v [P Q]].
     pose proof (multistep_preserves_R _ _ _ P H0).
     apply multistep_preserves_R' with (msubst ((x,v)::env0) t12).
       eapply T_App. eauto.
       apply R_typable_empty; auto.
       eapply multi_trans. eapply multistep_App2; eauto.
       eapply multi_R.
       simpl. rewrite subst_msubst.
       eapply ST_AppAbs; eauto.
       eapply typable_empty__closed.
       apply (R_typable_empty H1).
       eapply instantiation_env_closed; eauto.
       eapply (IHHT ((x,T11)::c)).
          intros. unfold update, t_update, lookup. destruct (eqb_string x x0); auto.
       constructor; auto.
  -
    rewrite msubst_app.
    destruct (IHHT1 c H env0 V) as [_ [_ P1]].
    pose proof (IHHT2 c H env0 V) as P2. fold R in P1. auto.
  Admitted.
```

Normalization Theorem

And the final theorem:

```
Theorem normalization : ∀ t T, has_type empty t T → halts t.
Proof.
  intros.
  replace t with (msubst nil t) by reflexivity.
  apply (@R_halts T).
  apply (msubst_R nil); eauto.
  eapply V_nil.
```

```
Qed.
```

Chapter 19

LibTactics: A Collection of Handy General-Purpose Tactics

This file contains a set of tactics that extends the set of builtin tactics provided with the standard distribution of Coq. It intends to overcome a number of limitations of the standard set of tactics, and thereby to help user to write shorter and more robust scripts.

Hopefully, Coq tactics will be improved as time goes by, and this file should ultimately be useless. In the meanwhile, serious Coq users will probably find it very useful.

The present file contains the implementation and the detailed documentation of those tactics. The SF reader need not read this file; instead, he/she is encouraged to read the chapter named UseTactics.v, which is gentle introduction to the most useful tactics from the LibTactic library.

The main features offered are:

- More convenient syntax for naming hypotheses, with tactics for introduction and inversion that take as input only the name of hypotheses of type `Prop`, rather than the name of all variables.

- Tactics providing true support for manipulating N-ary conjunctions, disjunctions and existentials, hidding the fact that the underlying implementation is based on binary propositions.

- Convenient support for automation: tactics followed with the symbol "~" or "*" will call automation on the generated subgoals. The symbol "~" stands for `auto` and "*" for `intuition eauto`. These bindings can be customized.

- Forward-chaining tactics are provided to instantiate lemmas either with variable or hypotheses or a mix of both.

- A more powerful implementation of `apply` is provided (it is based on `refine` and thus behaves better with respect to conversion).

- An improved inversion tactic which substitutes equalities on variables generated by the standard inversion mecanism. Moreover, it supports the elimination of dependently-typed equalities (requires axiom K, which is a weak form of Proof Irrelevance).

- Tactics for saving time when writing proofs, with tactics to asserts hypotheses or sub-goals, and improved tactics for clearing, renaming, and sorting hypotheses.

External credits:

- thanks to Xavier Leroy for providing the idea of tactic *forward*

- thanks to Georges Gonthier for the implementation trick in *rapply*

```
Set Implicit Arguments.
Require Import List.
```

Remove Hints Bool.trans_eq_bool.

19.1 Tools for Programming with Ltac

19.1.1 Identity Continuation

```
Ltac idcont tt :=
  idtac.
```

19.1.2 Untyped Arguments for Tactics

Any Coq value can be boxed into the type **Boxer**. This is useful to use Coq computations for implementing tactics.

```
Inductive Boxer : Type :=
  | boxer : ∀ (A:Type), A → Boxer.
```

19.1.3 Optional Arguments for Tactics

ltac_no_arg is a constant that can be used to simulate optional arguments in tactic definitions. Use *mytactic* ltac_no_arg on the tactic invokation, and use `match` *arg* `with` ltac_no_arg ⇒ .. or `match type of` *arg* `with` **ltac_No_arg** ⇒ .. to test whether an argument was provided.

```
Inductive ltac_No_arg : Set :=
  | ltac_no_arg : ltac_No_arg.
```

19.1.4 Wildcard Arguments for Tactics

ltac_wild is a constant that can be used to simulate wildcard arguments in tactic definitions. Notation is __.

```
Inductive ltac_Wild : Set :=
  | ltac_wild : ltac_Wild.
```

```
Notation "'__'" := ltac_wild : ltac_scope.
```

ltac_wilds is another constant that is typically used to simulate a sequence of N wildcards, with N chosen appropriately depending on the context. Notation is ___.

```
Inductive ltac_Wilds : Set :=
  | ltac_wilds : ltac_Wilds.
```

```
Notation "'___'" := ltac_wilds : ltac_scope.
```

```
Open Scope ltac_scope.
```

19.1.5 Position Markers

ltac_Mark and ltac_mark are dummy definitions used as sentinel by tactics, to mark a certain position in the context or in the goal.

```
Inductive ltac_Mark : Type :=
  | ltac_mark : ltac_Mark.
```

gen_until_mark repeats `generalize` on hypotheses from the context, starting from the bottom and stopping as soon as reaching an hypothesis of type *Mark*. If fails if *Mark* does not appear in the context.

```
Ltac gen_until_mark :=
  match goal with H: ?T ⊢ _ ⇒
  match T with
  | ltac_Mark ⇒ clear H
  | _ ⇒ generalize H; clear H; gen_until_mark
  end end.
```

gen_until_mark_with_processing F is similar to *gen_until_mark* except that it calls F on each hypothesis immediately before generalizing it. This is useful for processing the hypotheses.

```
Ltac gen_until_mark_with_processing cont :=
  match goal with H: ?T ⊢ _ ⇒
  match T with
  | ltac_Mark ⇒ clear H
  | _ ⇒ cont H; generalize H; clear H;
          gen_until_mark_with_processing cont
  end end.
```

intro_until_mark repeats `intro` until reaching an hypothesis of type *Mark*. It throws away the hypothesis *Mark*. It fails if *Mark* does not appear as an hypothesis in the goal.

```
Ltac intro_until_mark :=
  match goal with
  | ⊢ (ltac_Mark → _) ⇒ intros _
  | _ ⇒ intro; intro_until_mark
  end.
```

19.1.6 List of Arguments for Tactics

A datatype of type **list Boxer** is used to manipulate list of Coq values in ltac. Notation is » *v1 v2 ... vN* for building a list containing the values *v1* through *vN*.

```
Notation "'»'" :=
  (@nil Boxer)
  (at level 0)
  : ltac_scope.
Notation "'»' v1" :=
  ((boxer v1)::nil)
  (at level 0, v1 at level 0)
  : ltac_scope.
Notation "'»' v1 v2" :=
  ((boxer v1)::(boxer v2)::nil)
  (at level 0, v1 at level 0, v2 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0, v5 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
```

```
  v4 at level 0, v5 at level 0, v6 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::(boxer v7)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7 v8" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::(boxer v7)::(boxer v8)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0,
   v8 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7 v8 v9" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::(boxer v7)::(boxer v8)::(boxer v9)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0,
   v8 at level 0, v9 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7 v8 v9 v10" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::(boxer v7)::(boxer v8)::(boxer v9)::(boxer v10)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0,
   v8 at level 0, v9 at level 0, v10 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::(boxer v7)::(boxer v8)::(boxer v9)::(boxer v10)
   ::(boxer v11)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
   v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0,
   v8 at level 0, v9 at level 0, v10 at level 0, v11 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
   ::(boxer v6)::(boxer v7)::(boxer v8)::(boxer v9)::(boxer v10)
   ::(boxer v11)::(boxer v12)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
```

```
    v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0,
    v8 at level 0, v9 at level 0, v10 at level 0, v11 at level 0,
    v12 at level 0)
  : ltac_scope.
Notation "'»' v1 v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v12 v13" :=
  ((boxer v1)::(boxer v2)::(boxer v3)::(boxer v4)::(boxer v5)
    ::(boxer v6)::(boxer v7)::(boxer v8)::(boxer v9)::(boxer v10)
    ::(boxer v11)::(boxer v12)::(boxer v13)::nil)
  (at level 0, v1 at level 0, v2 at level 0, v3 at level 0,
    v4 at level 0, v5 at level 0, v6 at level 0, v7 at level 0,
    v8 at level 0, v9 at level 0, v10 at level 0, v11 at level 0,
    v12 at level 0, v13 at level 0)
  : ltac_scope.
```

The tactic *list_boxer_of* inputs a term E and returns a term of type "list boxer", according to the following rules:

- if E is already of type "list Boxer", then it returns E;
- otherwise, it returns the list (boxer E)::nil.

```
Ltac list_boxer_of E :=
  match type of E with
  | List.list Boxer ⇒ constr:(E)
  | _ ⇒ constr:((boxer E)::nil)
  end.
```

19.1.7 Databases of Lemmas

Use the hint facility to implement a database mapping terms to terms. To declare a new database, use a definition: `Definition` *mydatabase* := **True**.

Then, to map *mykey* to *myvalue*, write the hint: `Hint Extern` 1 (*Register mydatabase mykey*) ⇒ *Provide myvalue*.

Finally, to query the value associated with a key, run the tactic *ltac_database_get mydatabase mykey*. This will leave at the head of the goal the term *myvalue*. It can then be named and exploited using `intro`.

```
Inductive Ltac_database_token : Prop := ltac_database_token.

Definition ltac_database (D:Boxer) (T:Boxer) (A:Boxer) := Ltac_database_token.

Notation "'Register' D T" := (ltac_database (boxer D) (boxer T) _)
  (at level 69, D at level 0, T at level 0).

Lemma ltac_database_provide : ∀ (A:Boxer) (D:Boxer) (T:Boxer),
  ltac_database D T A.
Proof using. split. Qed.
```

```
Ltac Provide T := apply (@ltac_database_provide (boxer T)).

Ltac ltac_database_get D T :=
  let A := fresh "TEMP" in evar (A:Boxer);
  let H := fresh "TEMP" in
  assert (H : ltac_database (boxer D) (boxer T) A);
  [ subst A; auto
  | subst A; match type of H with ltac_database _ _ (boxer ?L) ⇒
               generalize L end; clear H ].
```

19.1.8 On-the-Fly Removal of Hypotheses

In a list of arguments » *H1 H2 .. HN* passed to a tactic such as *lets* or *applys* or *forwards* or *specializes*, the term rm, an identity function, can be placed in front of the name of an hypothesis to be deleted.

```
Definition rm (A:Type) (X:A) := X.
```

rm_term E removes one hypothesis that admits the same type as *E*.

```
Ltac rm_term E :=
  let T := type of E in
  match goal with H: T ⊢ _ ⇒ try clear H end.
```

rm_inside E calls *rm_term Ei* for any subterm of the form rm *Ei* found in E

```
Ltac rm_inside E :=
  let go E := rm_inside E in
  match E with
  | rm ?X ⇒ rm_term X
  | ?X1 ?X2 ⇒
     go X1; go X2
  | ?X1 ?X2 ?X3 ⇒
     go X1; go X2; go X3
  | ?X1 ?X2 ?X3 ?X4 ⇒
     go X1; go X2; go X3; go X4
  | ?X1 ?X2 ?X3 ?X4 ?X5 ⇒
     go X1; go X2; go X3; go X4; go X5
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ⇒
     go X1; go X2; go X3; go X4; go X5; go X6
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7 ⇒
     go X1; go X2; go X3; go X4; go X5; go X6; go X7
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7 ?X8 ⇒
     go X1; go X2; go X3; go X4; go X5; go X6; go X7; go X8
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7 ?X8 ?X9 ⇒
     go X1; go X2; go X3; go X4; go X5; go X6; go X7; go X8; go X9
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7 ?X8 ?X9 ?X10 ⇒
```

```
    go X1; go X2; go X3; go X4; go X5; go X6; go X7; go X8; go X9; go X10
  | _ ⇒ idtac
  end.
```

For faster performance, one may deactivate *rm_inside* by replacing the body of this definition with `idtac`.

```
Ltac fast_rm_inside E :=
  rm_inside E.
```

19.1.9 Numbers as Arguments

When tactic takes a natural number as argument, it may be parsed either as a natural number or as a relative number. In order for tactics to convert their arguments into natural numbers, we provide a conversion tactic.

```
Require BinPos Coq.ZArith.BinInt.

Require Coq.Numbers.BinNums Coq.ZArith.BinInt.

Definition ltac_int_to_nat (x:BinInt.Z) : nat :=
  match x with
  | BinInt.Z0 ⇒ 0%nat
  | BinInt.Zpos p ⇒ BinPos.nat_of_P p
  | BinInt.Zneg p ⇒ 0%nat
  end.

Ltac number_to_nat N :=
  match type of N with
  | nat ⇒ constr:(N)
  | BinInt.Z ⇒ let N' := constr:(ltac_int_to_nat N) in eval compute in N'
  end.
```

ltac_pattern E `at` K is the same as `pattern` E `at` K except that K is a Coq number (nat or Z) rather than a Ltac integer. Syntax *ltac_pattern* E `as` K `in` H is also available.

```
Tactic Notation "ltac_pattern" constr(E) "at" constr(K) :=
  match number_to_nat K with
  | 1 ⇒ pattern E at 1
  | 2 ⇒ pattern E at 2
  | 3 ⇒ pattern E at 3
  | 4 ⇒ pattern E at 4
  | 5 ⇒ pattern E at 5
  | 6 ⇒ pattern E at 6
  | 7 ⇒ pattern E at 7
  | 8 ⇒ pattern E at 8
  | _ ⇒ fail "ltac_pattern: arity not supported"
  end.
```

```
Tactic Notation "ltac_pattern" constr(E) "at" constr(K) "in" hyp(H) :=
  match number_to_nat K with
  | 1 => pattern E at 1 in H
  | 2 => pattern E at 2 in H
  | 3 => pattern E at 3 in H
  | 4 => pattern E at 4 in H
  | 5 => pattern E at 5 in H
  | 6 => pattern E at 6 in H
  | 7 => pattern E at 7 in H
  | 8 => pattern E at 8 in H
  | _ => fail "ltac_pattern: arity not supported"
  end.
```

ltac_set (x := E) at K is the same as set (x := E) at K except that K is a Coq number (nat or Z) rather than a Ltac integer.

```
Tactic Notation "ltac_set" "(" ident(X) ":=" constr(E) ")" "at" constr(K) :=
  match number_to_nat K with
  | 1%nat => set (X := E) at 1
  | 2%nat => set (X := E) at 2
  | 3%nat => set (X := E) at 3
  | 4%nat => set (X := E) at 4
  | 5%nat => set (X := E) at 5
  | 6%nat => set (X := E) at 6
  | 7%nat => set (X := E) at 7
  | 8%nat => set (X := E) at 8
  | 9%nat => set (X := E) at 9
  | 10%nat => set (X := E) at 10
  | 11%nat => set (X := E) at 11
  | 12%nat => set (X := E) at 12
  | 13%nat => set (X := E) at 13
  | _ => fail "ltac_set: arity not supported"
  end.
```

19.1.10 Testing Tactics

show tac executes a tactic *tac* that produces a result, and then display its result.

```
Tactic Notation "show" tactic(tac) :=
  let R := tac in pose R.
```

dup N produces N copies of the current goal. It is useful for building examples on which to illustrate behaviour of tactics. *dup* is short for *dup* 2.

```
Lemma dup_lemma : ∀ P, P → P → P.
Proof using. auto. Qed.
```

```
Ltac dup_tactic N :=
  match number_to_nat N with
  | 0 ⇒ idtac
  | S 0 ⇒ idtac
  | S ?N' ⇒ apply dup_lemma; [ | dup_tactic N' ]
  end.
Tactic Notation "dup" constr(N) :=
  dup_tactic N.
Tactic Notation "dup" :=
  dup 2.
```

19.1.11 Testing evars and non-evars

is_not_evar E succeeds only if *E* is not an evar; it fails otherwise. It thus implements the negation of *is_evar*

```
Ltac is_not_evar E :=
  first [ is_evar E; fail 1
        | idtac ].
```

is_evar_as_bool E evaluates to true if *E* is an evar and to false otherwise.

```
Ltac is_evar_as_bool E :=
  constr:(ltac:(first
    [ is_evar E; exact true
    | exact false ])).
```

19.1.12 Check No Evar in Goal

```
Ltac check_noevar M :=
  first [ has_evar M; fail 2 | idtac ].
Ltac check_noevar_hyp H :=
  let T := type of H in check_noevar T.
Ltac check_noevar_goal :=
  match goal with ⊢ ?G ⇒ check_noevar G end.
```

19.1.13 Helper Function for Introducing Evars

with_evar T (fun *M* ⇒ *tac*) creates a new evar that can be used in the tactic *tac* under the name *M*.

```
Ltac with_evar_base T cont :=
  let x := fresh "TEMP" in evar (x:T); cont x; subst x.
Tactic Notation "with_evar" constr(T) tactic(cont) :=
  with_evar_base T cont.
```

19.1.14 Tagging of Hypotheses

get_last_hyp tt is a function that returns the last hypothesis at the bottom of the context. It is useful to obtain the default name associated with the hypothesis, e.g. intro; let *H* := *get_last_hyp* tt in let *H'* := fresh "P" *H* in ...

```
Ltac get_last_hyp tt :=
  match goal with H: _ ⊢ _ ⇒ constr:(H) end.
```

19.1.15 More Tagging of Hypotheses

ltac_tag_subst is a specific marker for hypotheses which is used to tag hypotheses that are equalities to be substituted.

```
Definition ltac_tag_subst (A:Type) (x:A) := x.
```

ltac_to_generalize is a specific marker for hypotheses to be generalized.

```
Definition ltac_to_generalize (A:Type) (x:A) := x.

Ltac gen_to_generalize :=
  repeat match goal with
    H: ltac_to_generalize _ ⊢ _ ⇒ generalize H; clear H end.

Ltac mark_to_generalize H :=
  let T := type of H in
  change T with (ltac_to_generalize T) in H.
```

19.1.16 Deconstructing Terms

get_head *E* is a tactic that returns the head constant of the term *E*, ie, when applied to a term of the form P *x1* ... *xN* it returns P. If *E* is not an application, it returns *E*. Warning: the tactic seems to loop in some cases when the goal is a product and one uses the result of this function.

```
Ltac get_head E :=
  match E with
  | ?P _ _ _ _ _ _ _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ _ ⇒ constr:(P)
  | ?P _ _ _ ⇒ constr:(P)
  | ?P _ _ ⇒ constr:(P)
```

```
| ?P _ ⇒ constr:(P)
| ?P ⇒ constr:(P)
end.
```

get_fun_arg E is a tactic that decomposes an application term E, ie, when applied to a term of the form *X1* ... *XN* it returns a pair made of *X1* .. X(N-1) and *XN*.

```
Ltac get_fun_arg E :=
  match E with
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X7 ?X ⇒ constr:((X1 X2 X3 X4 X5 X6 X7,X))
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X6 ?X ⇒ constr:((X1 X2 X3 X4 X5 X6,X))
  | ?X1 ?X2 ?X3 ?X4 ?X5 ?X ⇒ constr:((X1 X2 X3 X4 X5,X))
  | ?X1 ?X2 ?X3 ?X4 ?X ⇒ constr:((X1 X2 X3 X4,X))
  | ?X1 ?X2 ?X3 ?X ⇒ constr:((X1 X2 X3,X))
  | ?X1 ?X2 ?X ⇒ constr:((X1 X2,X))
  | ?X1 ?X ⇒ constr:((X1,X))
  end.
```

19.1.17 Action at Occurence and Action Not at Occurence

ltac_action_at K `of` E `do` *Tac* isolates the K-th occurence of E in the goal, setting it in the form P E for some named pattern P, then calls tactic *Tac*, and finally unfolds P. Syntax *ltac_action_at* K `of` E `in` H `do` *Tac* is also available.

```
Tactic Notation "ltac_action_at" constr(K) "of" constr(E) "do" tactic(Tac) :=
  let p := fresh "TEMP" in ltac_pattern E at K;
  match goal with ⊢ ?P _ ⇒ set (p:=P) end;
  Tac; unfold p; clear p.
```

```
Tactic Notation "ltac_action_at" constr(K) "of" constr(E) "in" hyp(H) "do" tactic(Tac)
:=
  let p := fresh "TEMP" in ltac_pattern E at K in H;
  match type of H with ?P _ ⇒ set (p:=P) in H end;
  Tac; unfold p in H; clear p.
```

protects E `do` *Tac* temporarily assigns a name to the expression E so that the execution of tactic *Tac* will not modify E. This is useful for instance to restrict the action of `simpl`.

```
Tactic Notation "protects" constr(E) "do" tactic(Tac) :=

  let x := fresh "TEMP" in let H := fresh "TEMP" in
  set (X := E) in *; assert (H : X = E) by reflexivity;
  clearbody X; Tac; subst x.
```

```
Tactic Notation "protects" constr(E) "do" tactic(Tac) "/" :=
  protects E do Tac.
```

19.1.18 An Alias for eq

eq' is an alias for eq to be used for equalities in inductive definitions, so that they don't get mixed with equalities generated by inversion.

```
Definition eq' := @eq.
```

```
Hint Unfold eq'.
```

```
Notation "x '='' y" := (@eq' _ x y)
  (at level 70, y at next level).
```

19.2 Common Tactics for Simplifying Goals Like intuition

```
Ltac jauto_set_hyps :=
  repeat match goal with H: ?T ⊢ _ ⇒
    match T with
    | _ ∧ _ ⇒ destruct H
    | ∃ a , _ ⇒ destruct H
    | _ ⇒ generalize H; clear H
    end
  end.
```

```
Ltac jauto_set_goal :=
  repeat match goal with
  | ⊢ ∃ a , _ ⇒ esplit
  | ⊢ _ ∧ _ ⇒ split
  end.
```

```
Ltac jauto_set :=
  intros; jauto_set_hyps;
  intros; jauto_set_goal;
  unfold not in *.
```

19.3 Backward and Forward Chaining

19.3.1 Application

```
Ltac old_refine f :=
  refine f.
```

rapply is a tactic similar to eapply except that it is based on the refine tactics, and thus is strictly more powerful (at least in theory :). In short, it is able to perform on-the-fly conversions when required for arguments to match, and it is able to instantiate existentials when required.

```
Tactic Notation "rapply" constr(t) :=
  first
  [ eexact (@t)
  | old_refine (@t)
  | old_refine (@t _)
  | old_refine (@t _ _)
  | old_refine (@t _ _ _)
  | old_refine (@t _ _ _ _)
  | old_refine (@t _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _ _ _ _ _ _)
  | old_refine (@t _ _ _ _ _ _ _ _ _ _ _ _ _ _ _)
  ].
```

The tactics *applys_N* T, where N is a natural number, provides a more efficient way of using *applys* T. It avoids trying out all possible arities, by specifying explicitely the arity of function T.

```
Tactic Notation "rapply_0" constr(t) :=
  old_refine (@t).
Tactic Notation "rapply_1" constr(t) :=
  old_refine (@t _).
Tactic Notation "rapply_2" constr(t) :=
  old_refine (@t _ _).
Tactic Notation "rapply_3" constr(t) :=
  old_refine (@t _ _ _).
Tactic Notation "rapply_4" constr(t) :=
  old_refine (@t _ _ _ _).
Tactic Notation "rapply_5" constr(t) :=
  old_refine (@t _ _ _ _ _).
Tactic Notation "rapply_6" constr(t) :=
  old_refine (@t _ _ _ _ _ _).
Tactic Notation "rapply_7" constr(t) :=
  old_refine (@t _ _ _ _ _ _ _).
Tactic Notation "rapply_8" constr(t) :=
  old_refine (@t _ _ _ _ _ _ _ _).
Tactic Notation "rapply_9" constr(t) :=
```

old_refine (@*t* _ _ _ _ _ _ _ _ _).
Tactic Notation "rapply_10" constr(*t*) :=
old_refine (@*t* _ _ _ _ _ _ _ _ _ _).

lets_base H E adds an hypothesis *H* : T to the context, where T is the type of term *E*. If *H* is an introduction pattern, it will destruct *H* according to the pattern.

Ltac *lets_base I E* := generalize *E*; intros *I*.

applys_to H E transform the type of hypothesis *H* by replacing it by the result of the application of the term *E* to *H*. Intuitively, it is equivalent to *lets H*: (*E H*).

```
Tactic Notation "applys_to" hyp(H) constr(E) :=
  let H' := fresh "TEMP" in rename H into H';
  (first [ lets_base H (E H')
         | lets_base H (E _ H')
         | lets_base H (E _ _ H')
         | lets_base H (E _ _ _ H')
         | lets_base H (E _ _ _ _ H')
         | lets_base H (E _ _ _ _ _ H')
         | lets_base H (E _ _ _ _ _ _ H')
         | lets_base H (E _ _ _ _ _ _ _ H')
         | lets_base H (E _ _ _ _ _ _ _ _ H')
         | lets_base H (E _ _ _ _ _ _ _ _ _ H') ]
  ); clear H'.
```

applys_to H1,...,HN E applys *E* to several hypotheses

```
Tactic Notation "applys_to" hyp(H1) "," hyp(H2) constr(E) :=
  applys_to H1 E; applys_to H2 E.
Tactic Notation "applys_to" hyp(H1) "," hyp(H2) "," hyp(H3) constr(E) :=
  applys_to H1 E; applys_to H2 E; applys_to H3 E.
Tactic Notation "applys_to" hyp(H1) "," hyp(H2) "," hyp(H3) "," hyp(H4) constr(E)
:=
  applys_to H1 E; applys_to H2 E; applys_to H3 E; applys_to H4 E.
```

constructors calls constructor or econstructor.

```
Tactic Notation "constructors" :=
  first [ constructor | econstructor ]; unfold eq'.
```

19.3.2 Assertions

asserts H: T is another syntax for assert (*H* : T), which also works with introduction patterns. For instance, one can write: *asserts* \[x *P*\] (∃ n, n = 3), or *asserts* \[*H*|*H*\] (n = 0 ∨ n = 1).

```
Tactic Notation "asserts" simple_intropattern(I) ":" constr(T) :=
  let H := fresh "TEMP" in assert (H : T);
```

[| generalize *H*; clear *H*; intros *I*].

asserts H1 .. HN: T is a shorthand for *asserts* \[*H1* \[*H2* \[.. *HN*\]\]\]\: T].

```
Tactic Notation "asserts" simple_intropattern(I1)
 simple_intropattern(I2) ":" constr(T) :=
  asserts [I1 I2]: T.
Tactic Notation "asserts" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) ":" constr(T) :=
  asserts [I1 [I2 I3]]: T.
Tactic Notation "asserts" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3)
 simple_intropattern(I4) ":" constr(T) :=
  asserts [I1 [I2 [I3 I4]]]: T.
Tactic Notation "asserts" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3)
 simple_intropattern(I4) simple_intropattern(I5) ":" constr(T) :=
  asserts [I1 [I2 [I3 [I4 I5]]]]: T.
Tactic Notation "asserts" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3)
 simple_intropattern(I4) simple_intropattern(I5)
 simple_intropattern(I6) ":" constr(T) :=
  asserts [I1 [I2 [I3 [I4 [I5 I6]]]]]: T.
```

asserts: T is *asserts H*: T with *H* being chosen automatically.

```
Tactic Notation "asserts" ":" constr(T) :=
  let H := fresh "TEMP" in asserts H : T.
```

cuts H: T is the same as *asserts H*: T except that the two subgoals generated are swapped: the subgoal T comes second. Note that contrary to `cut`, it introduces the hypothesis.

```
Tactic Notation "cuts" simple_intropattern(I) ":" constr(T) :=
  cut (T); [ intros I | idtac ].
```

cuts: T is *cuts H*: T with *H* being chosen automatically.

```
Tactic Notation "cuts" ":" constr(T) :=
  let H := fresh "TEMP" in cuts H: T.
```

cuts H1 .. HN: T is a shorthand for *cuts* \[*H1* \[*H2* \[.. *HN*\]\]\]\: T].

```
Tactic Notation "cuts" simple_intropattern(I1)
 simple_intropattern(I2) ":" constr(T) :=
  cuts [I1 I2]: T.
Tactic Notation "cuts" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) ":" constr(T) :=
  cuts [I1 [I2 I3]]: T.
Tactic Notation "cuts" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3)
```

```
  simple_intropattern(I4) ":" constr(T) :=
    cuts [I1 [I2 [I3 I4]]]: T.
Tactic Notation "cuts" simple_intropattern(I1)
  simple_intropattern(I2) simple_intropattern(I3)
  simple_intropattern(I4) simple_intropattern(I5) ":" constr(T) :=
    cuts [I1 [I2 [I3 [I4 I5]]]]: T.
Tactic Notation "cuts" simple_intropattern(I1)
  simple_intropattern(I2) simple_intropattern(I3)
  simple_intropattern(I4) simple_intropattern(I5)
  simple_intropattern(I6) ":" constr(T) :=
    cuts [I1 [I2 [I3 [I4 [I5 I6]]]]]: T.
```

19.3.3 Instantiation and Forward-Chaining

The instantiation tactics are used to instantiate a lemma E (whose type is a product) on some arguments. The type of E is made of implications and universal quantifications, e.g. ∀ x, *P* x → ∀ y z, *Q* x y z → R z.

The first possibility is to provide arguments in order: first x, then a proof of *P* x, then y etc... In this mode, called "Args", all the arguments are to be provided. If a wildcard is provided (written __), then an existential variable will be introduced in place of the argument.

It is very convenient to give some arguments the lemma should be instantiated on, and let the tactic find out automatically where underscores should be insterted. Underscore arguments __ are interpret as follows: an underscore means that we want to skip the argument that has the same type as the next real argument provided (real means not an underscore). If there is no real argument after underscore, then the underscore is used for the first possible argument.

The general syntax is *tactic* (» *E1* .. *EN*) where *tactic* is the name of the tactic (possibly with some arguments) and *Ei* are the arguments. Moreover, some tactics accept the syntax *tactic E1* .. *EN* as short for *tactic* (» *E1* .. *EN*) for values of N up to 5.

Finally, if the argument *EN* given is a triple-underscore ___, then it is equivalent to providing a list of wildcards, with the appropriate number of wildcards. This means that all the remaining arguments of the lemma will be instantiated. Definitions in the conclusion are not unfolded in this case.

```
Ltac app_assert t P cont :=
  let H := fresh "TEMP" in
  assert (H : P); [ | cont(t H); clear H ].

Ltac app_evar t A cont :=
  let x := fresh "TEMP" in
  evar (x:A);
  let t' := constr:(t x) in
  let t'' := (eval unfold x in t') in
```

```
  subst x; cont t''.
Ltac app_arg t P v cont :=
  let H := fresh "TEMP" in
  assert (H : P); [ apply v | cont(t H); try clear H ].
Ltac build_app_alls t final :=
  let rec go t :=
    match type of t with
    | ?P -> ?Q => app_assert t P go
    | ∀ _:?A, _ => app_evar t A go
    | _ => final t
    end in
  go t.
Ltac boxerlist_next_type vs :=
  match vs with
  | nil => constr:(ltac_wild)
  | (boxer ltac_wild)::?vs' => boxerlist_next_type vs'
  | (boxer ltac_wilds)::_ => constr:(ltac_wild)
  | (@boxer ?T _)::_ => constr:(T)
  end.

Ltac build_app_hnts t vs final :=
  let rec go t vs :=
    match vs with
    | nil => first [ final t | fail 1 ]
    | (boxer ltac_wilds)::_ => first [ build_app_alls t final | fail 1 ]
    | (boxer ?v)::?vs' =>
      let cont t' := go t' vs in
      let cont' t' := go t' vs' in
      let T := type of t in
      let T := eval hnf in T in
      match v with
      | ltac_wild =>
         first [ let U := boxerlist_next_type vs' in
           match U with
           | ltac_wild =>
             match T with
             | ?P -> ?Q => first [ app_assert t P cont' | fail 3 ]
             | ∀ _:?A, _ => first [ app_evar t A cont' | fail 3 ]
             end
           | _ =>
             match T with
             | U -> ?Q => first [ app_assert t U cont' | fail 3 ]
```

```
              | ∀ _:U, _ ⇒ first [ app_evar t U cont' | fail 3 ]
              | ?P → ?Q ⇒ first [ app_assert t P cont | fail 3 ]
              | ∀ _:?A, _ ⇒ first [ app_evar t A cont | fail 3 ]
              end
            end
          | fail 2 ]
      | _ ⇒
          match T with
          | ?P → ?Q ⇒ first [ app_arg t P v cont'
                            | app_assert t P cont
                            | fail 3 ]
           | ∀ _:Type, _ ⇒
              match type of v with
              | Type ⇒ first [ cont' (t v)
                              | app_evar t Type cont
                              | fail 3 ]
              | _ ⇒ first [ app_evar t Type cont
                           | fail 3 ]
              end
          | ∀ _:?A, _ ⇒
             let V := type of v in
             match type of V with
             | Prop ⇒ first [ app_evar t A cont
                             | fail 3 ]
             | _ ⇒ first [ cont' (t v)
                          | app_evar t A cont
                          | fail 3 ]
             end
          end
      end
    end in
  go t vs.
```

newer version : support for typeclasses

```
Ltac app_typeclass t cont :=
  let t' := constr:(t _) in
  cont t'.

Ltac build_app_alls t final ::=
  let rec go t :=
    match type of t with
    | ?P → ?Q ⇒ app_assert t P go
    | ∀ _:?A, _ ⇒
        first [ app_evar t A go
```

```
                | app_typeclass t go
                | fail 3 ]
     | _ ⇒ final t
     end in
   go t.

Ltac build_app_hnts t vs final ::=
   let rec go t vs :=
     match vs with
     | nil ⇒ first [ final t | fail 1 ]
     | (boxer ltac_wilds)::_ ⇒ first [ build_app_alls t final | fail 1 ]
     | (boxer ?v)::?vs' ⇒
       let cont t' := go t' vs in
       let cont' t' := go t' vs' in
       let T := type of t in
       let T := eval hnf in T in
       match v with
       | ltac_wild ⇒
          first [ let U := boxerlist_next_type vs' in
            match U with
            | ltac_wild ⇒
              match T with
              | ?P → ?Q ⇒ first [ app_assert t P cont' | fail 3 ]
              | ∀ _:?A, _ ⇒ first [ app_typeclass t cont'
                                  | app_evar t A cont'
                                  | fail 3 ]
              end
            | _ ⇒
              match T with
              | U → ?Q ⇒ first [ app_assert t U cont' | fail 3 ]
              | ∀ _:U, _ ⇒ first
                   [ app_typeclass t cont'
                   | app_evar t U cont'
                   | fail 3 ]
              | ?P → ?Q ⇒ first [ app_assert t P cont | fail 3 ]
              | ∀ _:?A, _ ⇒ first
                   [ app_typeclass t cont
                   | app_evar t A cont
                   | fail 3 ]
              end
            end
          | fail 2 ]
       | _ ⇒
```

```
          match T with
          | ?P → ?Q ⇒ first [ app_arg t P v cont'
                            | app_assert t P cont
                            | fail 3 ]
           | ∀ _:Type, _ ⇒
              match type of v with
              | Type ⇒ first [ cont' (t v)
                            | app_evar t Type cont
                            | fail 3 ]
              | _ ⇒ first [ app_evar t Type cont
                          | fail 3 ]
              end
          | ∀ _:?A, _ ⇒
             let V := type of v in
             match type of V with
             | Prop ⇒ first [ app_typeclass t cont
                            | app_evar t A cont
                            | fail 3 ]
             | _ ⇒ first [ cont' (t v)
                         | app_typeclass t cont
                         | app_evar t A cont
                         | fail 3 ]
             end
          end
      end
    end in
  go t vs.

Ltac build_app args final :=
  first [
    match args with (@boxer ?T ?t)::?vs ⇒
      let t := constr:(t:T) in
      build_app_hnts t vs final;
      fast_rm_inside args
    end
  | fail 1 "Instantiation fails for:" args].

Ltac unfold_head_until_product T :=
  eval hnf in T.

Ltac args_unfold_head_if_not_product args :=
  match args with (@boxer ?T ?t)::?vs ⇒
    let T' := unfold_head_until_product T in
    constr:((@boxer T' t)::vs)
```

```
  end.
Ltac args_unfold_head_if_not_product_but_params args :=
  match args with
  | (boxer ?t)::(boxer ?v)::?vs ⇒
      args_unfold_head_if_not_product args
  | _ ⇒ constr:(args)
  end.
```

lets H: (» *E0 E1* .. *EN*) will instantiate lemma *E0* on the arguments *Ei* (which may be wildcards __), and name *H* the resulting term. *H* may be an introduction pattern, or a sequence of introduction patterns *I1 I2 IN*, or empty. Syntax *lets H*: *E0 E1* .. *EN* is also available. If the last argument *EN* is ___ (triple-underscore), then all arguments of *H* will be instantiated.

```
Ltac lets_build I Ei :=
  let args := list_boxer_of Ei in
  let args := args_unfold_head_if_not_product_but_params args in

  build_app args ltac:(fun R ⇒ lets_base I R).
Tactic Notation "lets" simple_intropattern(I) ":" constr(E) :=
  lets_build I E.
Tactic Notation "lets" ":" constr(E) :=
  let H := fresh in lets H: E.
Tactic Notation "lets" ":" constr(E0)
 constr(A1) :=
  lets: (» E0 A1).
Tactic Notation "lets" ":" constr(E0)
 constr(A1) constr(A2) :=
  lets: (» E0 A1 A2).
Tactic Notation "lets" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  lets: (» E0 A1 A2 A3).
Tactic Notation "lets" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  lets: (» E0 A1 A2 A3 A4).
Tactic Notation "lets" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  lets: (» E0 A1 A2 A3 A4 A5).

Tactic Notation "lets" simple_intropattern(I1) simple_intropattern(I2)
 ":" constr(E) :=
  lets [I1 I2]: E.
Tactic Notation "lets" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) ":" constr(E) :=
```

lets [*I1* [*I2* *I3*]]: *E*.
Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
simple_intropattern(*I3*) *simple_intropattern*(*I4*) ":" constr(*E*) :=
lets [*I1* [*I2* [*I3* *I4*]]]: *E*.
Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
":" constr(*E*) :=
lets [*I1* [*I2* [*I3* [*I4* *I5*]]]]: *E*.

Tactic Notation "lets" *simple_intropattern*(*I*) ":" constr(*E0*)
constr(*A1*) :=
lets *I*: (» *E0* *A1*).
Tactic Notation "lets" *simple_intropattern*(*I*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) :=
lets *I*: (» *E0* *A1* *A2*).
Tactic Notation "lets" *simple_intropattern*(*I*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) constr(*A3*) :=
lets *I*: (» *E0* *A1* *A2* *A3*).
Tactic Notation "lets" *simple_intropattern*(*I*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) constr(*A3*) constr(*A4*) :=
lets *I*: (» *E0* *A1* *A2* *A3* *A4*).
Tactic Notation "lets" *simple_intropattern*(*I*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) constr(*A3*) constr(*A4*) constr(*A5*) :=
lets *I*: (» *E0* *A1* *A2* *A3* *A4* *A5*).

Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) ":" constr(*E0*)
constr(*A1*) :=
lets [*I1* *I2*]: *E0* *A1*.
Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) :=
lets [*I1* *I2*]: *E0* *A1* *A2*.
Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) constr(*A3*) :=
lets [*I1* *I2*]: *E0* *A1* *A2* *A3*.
Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) constr(*A3*) constr(*A4*) :=
lets [*I1* *I2*]: *E0* *A1* *A2* *A3* *A4*.
Tactic Notation "lets" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) ":" constr(*E0*)
constr(*A1*) constr(*A2*) constr(*A3*) constr(*A4*) constr(*A5*) :=
lets [*I1* *I2*]: *E0* *A1* *A2* *A3* *A4* *A5*.

forwards *H*: (» *E0* *E1* .. *EN*) is short for *forwards* *H*: (» *E0* *E1* .. *EN* ___). The arguments *Ei* can be wildcards __ (except *E0*). *H* may be an introduction pattern, or a sequence of introduction pattern, or empty. Syntax *forwards* *H*: *E0* *E1* .. *EN* is also available.

```
Ltac forwards_build_app_arg Ei :=
  let args := list_boxer_of Ei in
  let args := (eval simpl in (args ++ ((boxer ___)::nil))) in
  let args := args_unfold_head_if_not_product args in
  args.

Ltac forwards_then Ei cont :=
  let args := forwards_build_app_arg Ei in
  let args := args_unfold_head_if_not_product_but_params args in
  build_app args cont.

Tactic Notation "forwards" simple_intropattern(I) ":" constr(Ei) :=
  let args := forwards_build_app_arg Ei in
  lets I: args.

Tactic Notation "forwards" ":" constr(E) :=
  let H := fresh in forwards H: E.
Tactic Notation "forwards" ":" constr(E0)
 constr(A1) :=
  forwards: (» E0 A1).
Tactic Notation "forwards" ":" constr(E0)
 constr(A1) constr(A2) :=
  forwards: (» E0 A1 A2).
Tactic Notation "forwards" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  forwards: (» E0 A1 A2 A3).
Tactic Notation "forwards" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  forwards: (» E0 A1 A2 A3 A4).
Tactic Notation "forwards" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  forwards: (» E0 A1 A2 A3 A4 A5).

Tactic Notation "forwards" simple_intropattern(I1) simple_intropattern(I2)
 ":" constr(E) :=
  forwards [I1 I2]: E.
Tactic Notation "forwards" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) ":" constr(E) :=
  forwards [I1 [I2 I3]]: E.
Tactic Notation "forwards" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) ":" constr(E) :=
  forwards [I1 [I2 [I3 I4]]]: E.
Tactic Notation "forwards" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) simple_intropattern(I5)
 ":" constr(E) :=
```

forwards [*I1* [*I2* [*I3* [*I4* *I5*]]]]: *E*.

```
Tactic Notation "forwards" simple_intropattern(I) ":" constr(E0)
 constr(A1) :=
  forwards I: (» E0 A1).
Tactic Notation "forwards" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) :=
  forwards I: (» E0 A1 A2).
Tactic Notation "forwards" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  forwards I: (» E0 A1 A2 A3).
Tactic Notation "forwards" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  forwards I: (» E0 A1 A2 A3 A4).
Tactic Notation "forwards" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  forwards I: (» E0 A1 A2 A3 A4 A5).
```

forwards_nounfold I: *E* is like *forwards* I: *E* but does not unfold the head constant of *E* if there is no visible quantification or hypothesis in *E*. It is meant to be used mainly by tactics.

```
Tactic Notation "forwards_nounfold" simple_intropattern(I) ":" constr(Ei) :=
  let args := list_boxer_of Ei in
  let args := (eval simpl in (args ++ ((boxer ___)::nil))) in
  build_app args ltac:(fun R ⇒ lets_base I R).
```

forwards_nounfold_then *E* ltac:(fun *K* ⇒ ..) is like *forwards*: *E* but it provides the resulting term to a continuation, under the name *K*.

```
Ltac forwards_nounfold_then Ei cont :=
  let args := list_boxer_of Ei in
  let args := (eval simpl in (args ++ ((boxer ___)::nil))) in
  build_app args cont.
```

applys (» *E0* *E1* .. *EN*) instantiates lemma *E0* on the arguments *Ei* (which may be wildcards __), and apply the resulting term to the current goal, using the tactic *applys* defined earlier on. *applys* *E0* *E1* *E2* .. *EN* is also available.

```
Ltac applys_build Ei :=
  let args := list_boxer_of Ei in
  let args := args_unfold_head_if_not_product_but_params args in
  build_app args ltac:(fun R ⇒
   first [ apply R | eapply R | rapply R ]).

Ltac applys_base E :=
  match type of E with
  | list Boxer ⇒ applys_build E
```

```
  | _ ⇒ first [ rapply E | applys_build E ]
  end; fast_rm_inside E.
Tactic Notation "applys" constr(E) :=
  applys_base E.
Tactic Notation "applys" constr(E0) constr(A1) :=
  applys (» E0 A1).
Tactic Notation "applys" constr(E0) constr(A1) constr(A2) :=
  applys (» E0 A1 A2).
Tactic Notation "applys" constr(E0) constr(A1) constr(A2) constr(A3) :=
  applys (» E0 A1 A2 A3).
Tactic Notation "applys" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  applys (» E0 A1 A2 A3 A4).
Tactic Notation "applys" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  applys (» E0 A1 A2 A3 A4 A5).
```

fapplys (» *E0 E1* .. *EN*) instantiates lemma *E0* on the arguments *Ei* and on the argument ___ meaning that all evars should be explicitly instantiated, and apply the resulting term to the current goal. *fapplys E0 E1 E2* .. *EN* is also available.

```
Ltac fapplys_build Ei :=
  let args := list_boxer_of Ei in
  let args := (eval simpl in (args ++ ((boxer ___)::nil))) in
  let args := args_unfold_head_if_not_product_but_params args in
  build_app args ltac:(fun R ⇒ apply R).
Tactic Notation "fapplys" constr(E0) :=
  match type of E0 with
  | list Boxer ⇒ fapplys_build E0
  | _ ⇒ fapplys_build (» E0)
  end.
Tactic Notation "fapplys" constr(E0) constr(A1) :=
  fapplys (» E0 A1).
Tactic Notation "fapplys" constr(E0) constr(A1) constr(A2) :=
  fapplys (» E0 A1 A2).
Tactic Notation "fapplys" constr(E0) constr(A1) constr(A2) constr(A3) :=
  fapplys (» E0 A1 A2 A3).
Tactic Notation "fapplys" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  fapplys (» E0 A1 A2 A3 A4).
Tactic Notation "fapplys" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  fapplys (» E0 A1 A2 A3 A4 A5).
```

specializes H (» *E1 E2* .. *EN*) will instantiate hypothesis *H* on the arguments *Ei* (which may be wildcards __). If the last argument *EN* is ___ (triple-underscore), then all arguments of *H* get instantiated.

```
Ltac specializes_build H Ei :=
  let H' := fresh "TEMP" in rename H into H';
  let args := list_boxer_of Ei in
  let args := constr:((boxer H')::args) in
  let args := args_unfold_head_if_not_product args in
  build_app args ltac:(fun R ⇒ lets H: R);
  clear H'.

Ltac specializes_base H Ei :=
  specializes_build H Ei; fast_rm_inside Ei.

Tactic Notation "specializes" hyp(H) :=
  specializes_base H (___).
Tactic Notation "specializes" hyp(H) constr(A) :=
  specializes_base H A.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) :=
  specializes H (» A1 A2).
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) :=
  specializes H (» A1 A2 A3).
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  specializes H (» A1 A2 A3 A4).
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  specializes H (» A1 A2 A3 A4 A5).
```

specializes_vars H is equivalent to *specializes H* __ .. __ with as many double underscore as the number of dependent arguments visible from the type of *H*. Note that no unfolding is currently being performed (this behavior might change in the future). The current implementation is restricted to the case where *H* is an existing hypothesis – TODO: generalize.

```
Ltac specializes_var_base H :=
  match type of H with
  | ?P → ?Q ⇒ fail 1
  | ∀ _:_, _ ⇒ specializes H __
  end.

Ltac specializes_vars_base H :=
  repeat (specializes_var_base H).

Tactic Notation "specializes_var" hyp(H) :=
  specializes_var_base H.

Tactic Notation "specializes_vars" hyp(H) :=
  specializes_vars_base H.
```

19.3.4 Experimental Tactics for Application

fapply is a version of `apply` based on *forwards*.

```
Tactic Notation "fapply" constr(E) :=
  let H := fresh "TEMP" in forwards H: E;
  first [ apply H | eapply H | rapply H | hnf; apply H
        | hnf; eapply H | applys H ].
```

sapply stands for "super apply". It tries `apply`, `eapply`, *applys* and *fapply*, and also tries to head-normalize the goal first.

```
Tactic Notation "sapply" constr(H) :=
  first [ apply H | eapply H | rapply H | applys H
        | hnf; apply H | hnf; eapply H | hnf; applys H
        | fapply H ].
```

19.3.5 Adding Assumptions

lets_simpl H: E is the same as *lets* H: E excepts that it calls `simpl` on the hypothesis H. *lets_simpl*: E is also provided.

```
Tactic Notation "lets_simpl" ident(H) ":" constr(E) :=
  lets H: E; try simpl in H.
```

```
Tactic Notation "lets_simpl" ":" constr(T) :=
  let H := fresh "TEMP" in lets_simpl H: T.
```

lets_hnf H: E is the same as *lets* H: E excepts that it calls `hnf` to set the definition in head normal form. *lets_hnf*: E is also provided.

```
Tactic Notation "lets_hnf" ident(H) ":" constr(E) :=
  lets H: E; hnf in H.
```

```
Tactic Notation "lets_hnf" ":" constr(T) :=
  let H := fresh "TEMP" in lets_hnf H: T.
```

puts X: E is a synonymous for `pose` (X := E). Alternative syntax is *puts*: E.

```
Tactic Notation "puts" ident(X) ":" constr(E) :=
  pose (X := E).
Tactic Notation "puts" ":" constr(E) :=
  let X := fresh "X" in pose (X := E).
```

19.3.6 Application of Tautologies

logic E, where E is a fact, is equivalent to `assert` H:E; [`tauto` | `eapply` H; `clear` H]. *It* `is` *useful* `for` *instance to prove* a *conjunction* [A ∧ B] `by` *showing* `first` [A] *and* `then` [A → B], *through the command* [*logic* (*foral* A B, A → (A → B) → A ∧ B)]

```
Ltac logic_base E cont :=
```

```
  assert (H:E); [ cont tt | eapply H; clear H ].
Tactic Notation "logic" constr(E) :=
  logic_base E ltac:(fun _ ⇒ tauto).
```

19.3.7 Application Modulo Equalities

The tactic *equates* replaces a goal of the form P x y z with a goal of the form P x ?a z and a subgoal ?a = y. The introduction of the evar ?a makes it possible to apply lemmas that would not apply to the original goal, for example a lemma of the form ∀ n m, P n n m, because x and y might be equal but not convertible.

Usage is *equates* i1 ... *ik*, where the indices are the positions of the arguments to be replaced by evars, counting from the right-hand side. If 0 is given as argument, then the entire goal is replaced by an evar.

```
Section equatesLemma.
Variables (A0 A1 : Type).
Variables (A2 : ∀ (x1 : A1), Type).
Variables (A3 : ∀ (x1 : A1) (x2 : A2 x1), Type).
Variables (A4 : ∀ (x1 : A1) (x2 : A2 x1) (x3 : A3 x2), Type).
Variables (A5 : ∀ (x1 : A1) (x2 : A2 x1) (x3 : A3 x2) (x4 : A4 x3), Type).
Variables (A6 : ∀ (x1 : A1) (x2 : A2 x1) (x3 : A3 x2) (x4 : A4 x3) (x5 : A5 x4), Type).
Lemma equates_0 : ∀ (P Q:Prop),
  P → P = Q → Q.
Proof. intros. subst. auto. Qed.
Lemma equates_1 :
  ∀ (P:A0→Prop) x1 y1,
  P y1 → x1 = y1 → P x1.
Proof. intros. subst. auto. Qed.
Lemma equates_2 :
  ∀ y1 (P:A0→∀(x1:A1),Prop) x1 x2,
  P y1 x2 → x1 = y1 → P x1 x2.
Proof. intros. subst. auto. Qed.
Lemma equates_3 :
  ∀ y1 (P:A0→∀(x1:A1)(x2:A2 x1),Prop) x1 x2 x3,
  P y1 x2 x3 → x1 = y1 → P x1 x2 x3.
Proof. intros. subst. auto. Qed.
Lemma equates_4 :
  ∀ y1 (P:A0→∀(x1:A1)(x2:A2 x1)(x3:A3 x2),Prop) x1 x2 x3 x4,
  P y1 x2 x3 x4 → x1 = y1 → P x1 x2 x3 x4.
Proof. intros. subst. auto. Qed.
Lemma equates_5 :
```

```
  ∀ y1 (P:A0→∀(x1:A1)(x2:A2 x1)(x3:A3 x2)(x4:A4 x3),Prop) x1 x2 x3 x4 x5,
  P y1 x2 x3 x4 x5 → x1 = y1 → P x1 x2 x3 x4 x5.
Proof. intros. subst. auto. Qed.

Lemma equates_6 :
  ∀ y1 (P:A0→∀(x1:A1)(x2:A2 x1)(x3:A3 x2)(x4:A4 x3)(x5:A5 x4),Prop)
  x1 x2 x3 x4 x5 x6,
  P y1 x2 x3 x4 x5 x6 → x1 = y1 → P x1 x2 x3 x4 x5 x6.
Proof. intros. subst. auto. Qed.

End equatesLemma.

Ltac equates_lemma n :=
  match number_to_nat n with
  | 0 ⇒ constr:(equates_0)
  | 1 ⇒ constr:(equates_1)
  | 2 ⇒ constr:(equates_2)
  | 3 ⇒ constr:(equates_3)
  | 4 ⇒ constr:(equates_4)
  | 5 ⇒ constr:(equates_5)
  | 6 ⇒ constr:(equates_6)
  end.

Ltac equates_one n :=
  let L := equates_lemma n in
  eapply L.

Ltac equates_several E cont :=
  let all_pos := match type of E with
    | List.list Boxer ⇒ constr:(E)
    | _ ⇒ constr:((boxer E)::nil)
    end in
  let rec go pos :=
     match pos with
     | nil ⇒ cont tt
     | (boxer ?n)::?pos' ⇒ equates_one n; [ instantiate; go pos' | ]
     end in
  go all_pos.

Tactic Notation "equates" constr(E) :=
  equates_several E ltac:(fun _ ⇒ idtac).
Tactic Notation "equates" constr(n1) constr(n2) :=
  equates (» n1 n2).
Tactic Notation "equates" constr(n1) constr(n2) constr(n3) :=
  equates (» n1 n2 n3).
Tactic Notation "equates" constr(n1) constr(n2) constr(n3) constr(n4) :=
  equates (» n1 n2 n3 n4).
```

applys_eq H i1 .. *iK* is the same as *equates* i1 .. *iK* followed by `apply` H on the first subgoal.

```
Tactic Notation "applys_eq" constr(H) constr(E) :=
  equates_several E ltac:(fun _ ⇒ sapply H).
Tactic Notation "applys_eq" constr(H) constr(n1) constr(n2) :=
  applys_eq H (» n1 n2).
Tactic Notation "applys_eq" constr(H) constr(n1) constr(n2) constr(n3) :=
  applys_eq H (» n1 n2 n3).
Tactic Notation "applys_eq" constr(H) constr(n1) constr(n2) constr(n3) constr(n4)
:=
  applys_eq H (» n1 n2 n3 n4).
```

19.3.8 Absurd Goals

false_goal replaces any goal by the goal **False**. Contrary to the tactic false (below), it does not try to do anything else

```
Tactic Notation "false_goal" :=
  elimtype False.
```

false_post is the underlying tactic used to prove goals of the form **False**. In the default implementation, it proves the goal if the context contains **False** or an hypothesis of the form C *x1* .. *xN* = *D* *y1* .. *yM*, or if the `congruence` tactic finds a proof of x ≠ x for some x.

```
Ltac false_post :=
  solve [ assumption | discriminate | congruence ].
```

false replaces any goal by the goal **False**, and calls *false_post*

```
Tactic Notation "false" :=
  false_goal; try false_post.
```

tryfalse tries to solve a goal by contradiction, and leaves the goal unchanged if it cannot solve it. It is equivalent to `try solve` \[false \].

```
Tactic Notation "tryfalse" :=
  try solve [ false ].
```

false *E* tries to exploit lemma *E* to prove the goal false. false *E1* .. *EN* is equivalent to false (» *E1* .. *EN*), which tries to apply *applys* (» *E1* .. *EN*) and if it does not work then tries *forwards* *H*: (» *E1* .. *EN*) followed with false

```
Ltac false_then E cont :=
  false_goal; first
  [ applys E; instantiate
  | forwards_then E ltac:(fun M ⇒
      pose M; jauto_set_hyps; intros; false) ];
  cont tt.
```

```
Tactic Notation "false" constr(E) :=
  false_then E ltac:(fun _ ⇒ idtac).
Tactic Notation "false" constr(E) constr(E1) :=
  false (» E E1).
Tactic Notation "false" constr(E) constr(E1) constr(E2) :=
  false (» E E1 E2).
Tactic Notation "false" constr(E) constr(E1) constr(E2) constr(E3) :=
  false (» E E1 E2 E3).
Tactic Notation "false" constr(E) constr(E1) constr(E2) constr(E3) constr(E4) :=
  false (» E E1 E2 E3 E4).
```

false_invert H proves a goal if it absurd after calling `inversion` H and false

```
Ltac false_invert_for H :=
  let M := fresh "TEMP" in pose (M := H); inversion H; false.
```

```
Tactic Notation "false_invert" constr(H) :=
  try solve [ false_invert_for H | false ].
```

false_invert proves any goal provided there is at least one hypothesis H in the context (or as a universally quantified hypothesis visible at the head of the goal) that can be proved absurd by calling `inversion` H.

```
Ltac false_invert_iter :=
  match goal with H:_ ⊢ _ ⇒
    solve [ inversion H; false
          | clear H; false_invert_iter
          | fail 2 ] end.
```

```
Tactic Notation "false_invert" :=
  intros; solve [ false_invert_iter | false ].
```

tryfalse_invert H and *tryfalse_invert* are like the above but leave the goal unchanged if they don't solve it.

```
Tactic Notation "tryfalse_invert" constr(H) :=
  try (false_invert H).
```

```
Tactic Notation "tryfalse_invert" :=
  try false_invert.
```

false_neq_self_hyp proves any goal if the context contains an hypothesis of the form $E \neq E$. It is a restricted and optimized version of false. It is intended to be used by other tactics only.

```
Ltac false_neq_self_hyp :=
  match goal with H: ?x ≠ ?x ⊢ _ ⇒
    false_goal; apply H; reflexivity end.
```

19.4 Introduction and Generalization

19.4.1 Introduction

introv is used to name only non-dependent hypothesis.

- If *introv* is called on a goal of the form ∀ x, H, it should introduce all the variables quantified with a ∀ at the head of the goal, but it does not introduce hypotheses that preceed an arrow constructor, like in $P \rightarrow Q$.
- If *introv* is called on a goal that is not of the form ∀ x, H nor $P \rightarrow Q$, the tactic unfolds definitions until the goal takes the form ∀ x, H or $P \rightarrow Q$. If unfolding definitions does not produces a goal of this form, then the tactic *introv* does nothing at all.

```
Ltac introv_rec :=
  match goal with
  | ⊢ ?P → ?Q ⇒ idtac
  | ⊢ ∀ _, _ ⇒ intro; introv_rec
  | ⊢ _ ⇒ idtac
  end.

Ltac introv_noarg :=
  match goal with
  | ⊢ ?P → ?Q ⇒ idtac
  | ⊢ ∀ _, _ ⇒ introv_rec
  | ⊢ ?G ⇒ hnf;
     match goal with
     | ⊢ ?P → ?Q ⇒ idtac
     | ⊢ ∀ _, _ ⇒ introv_rec
     end
  | ⊢ _ ⇒ idtac
  end.

  Ltac introv_noarg_not_optimized :=
    intro; match goal with H:_⊢_ ⇒ revert H end; introv_rec.

Ltac introv_arg H :=
  hnf; match goal with
  | ⊢ ?P → ?Q ⇒ intros H
  | ⊢ ∀ _, _ ⇒ intro; introv_arg H
  end.

Tactic Notation "introv" :=
  introv_noarg.
Tactic Notation "introv" simple_intropattern(I1) :=
```

introv_arg I1.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) :=
  introv I1; *introv I2*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) :=
  introv I1; *introv I2 I3*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) :=
  introv I1; *introv I2 I3 I4*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*) :=
  introv I1; *introv I2 I3 I4 I5*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) :=
  introv I1; *introv I2 I3 I4 I5 I6*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) :=
  introv I1; *introv I2 I3 I4 I5 I6 I7*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*) :=
  introv I1; *introv I2 I3 I4 I5 I6 I7 I8*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*)
 simple_intropattern(*I9*) :=
  introv I1; *introv I2 I3 I4 I5 I6 I7 I8 I9*.

Tactic Notation "introv" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*)
 simple_intropattern(*I9*) *simple_intropattern*(*I10*) :=
  introv I1; *introv I2 I3 I4 I5 I6 I7 I8 I9 I10*.

intros_all repeats `intro` as long as possible. Contrary to `intros`, it unfolds any definition on the way. Remark that it also unfolds the definition of negation, so applying *intros_all* to a goal of the form $\forall$ x, P x $\rightarrow \neg Q$ will introduce x and P x and Q, and will leave **False** in the goal.

Tactic Notation "intros_all" :=
  `repeat intro`.

intros_hnf introduces an hypothesis and sets in head normal form

Tactic Notation "intro_hnf" :=
  intro; match goal with H: _ ⊢ _ ⇒ hnf in H end.

19.4.2 Introduction using ⇒ and =»

Ltac *ltac_intros_post* := idtac.

Tactic Notation "=>" :=
  intros.
Tactic Notation "=>" *simple_intropattern*(*I1*) :=
  intros *I1*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*) :=
  intros *I1* *I2*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) :=
  intros *I1* *I2* *I3*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) :=
  intros *I1* *I2* *I3* *I4*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*) :=
  intros *I1* *I2* *I3* *I4* *I5*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) :=
  intros *I1* *I2* *I3* *I4* *I5* *I6*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) :=
  intros *I1* *I2* *I3* *I4* *I5* *I6* *I7*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*) :=
  intros *I1* *I2* *I3* *I4* *I5* *I6* *I7* *I8*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*)
 simple_intropattern(*I9*) :=
  intros *I1* *I2* *I3* *I4* *I5* *I6* *I7* *I8* *I9*.
Tactic Notation "=>" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
 simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
 simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*)
 simple_intropattern(*I9*) *simple_intropattern*(*I10*) :=

```
  intros I1 I2 I3 I4 I5 I6 I7 I8 I9 I10.

Ltac intro_nondeps_aux_special_intro G :=
  fail.

Ltac intro_nondeps_aux is_already_hnf :=
  match goal with
  | ⊢ (?P → ?Q) ⇒ idtac
  | ⊢ ?G → _ ⇒ intro_nondeps_aux_special_intro G;
                 intro; intro_nondeps_aux true
  | ⊢ (∀ _,_) ⇒ intros ?; intro_nondeps_aux true
  | ⊢ _ ⇒
     match is_already_hnf with
     | true ⇒ idtac
     | false ⇒ hnf; intro_nondeps_aux true
     end
  end.

Ltac intro_nondeps tt := intro_nondeps_aux false.

Tactic Notation "=»" :=
  intro_nondeps tt.
Tactic Notation "=»" simple_intropattern(I1) :=
  =»; intros I1.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2) :=
  =»; intros I1 I2.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) :=
  =»; intros I1 I2 I3.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) :=
  =»; intros I1 I2 I3 I4.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) simple_intropattern(I5) :=
  =»; intros I1 I2 I3 I4 I5.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) simple_intropattern(I5)
 simple_intropattern(I6) :=
  =»; intros I1 I2 I3 I4 I5 I6.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) simple_intropattern(I5)
 simple_intropattern(I6) simple_intropattern(I7) :=
  =»; intros I1 I2 I3 I4 I5 I6 I7.
Tactic Notation "=»" simple_intropattern(I1) simple_intropattern(I2)
 simple_intropattern(I3) simple_intropattern(I4) simple_intropattern(I5)
```

simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*) :=
=»; `intros` *I1 I2 I3 I4 I5 I6 I7 I8*.
`Tactic Notation` "=»" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*)
simple_intropattern(*I9*) :=
=»; `intros` *I1 I2 I3 I4 I5 I6 I7 I8 I9*.
`Tactic Notation` "=»" *simple_intropattern*(*I1*) *simple_intropattern*(*I2*)
simple_intropattern(*I3*) *simple_intropattern*(*I4*) *simple_intropattern*(*I5*)
simple_intropattern(*I6*) *simple_intropattern*(*I7*) *simple_intropattern*(*I8*)
simple_intropattern(*I9*) *simple_intropattern*(*I10*) :=
=»; `intros` *I1 I2 I3 I4 I5 I6 I7 I8 I9 I10*.

19.4.3 Generalization

gen X1 .. *XN* is a shorthand for calling `generalize dependent` successively on variables *XN...X1*. Note that the variables are generalized in reverse order, following the convention of the `generalize` tactic: it means that *X1* will be the first quantified variable in the resulting goal.

`Tactic Notation` "gen" *ident*(*X1*) :=
`generalize dependent` *X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) :=
gen X2; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) :=
gen X3; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) :=
gen X4; *gen X3*; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) *ident*(*X5*) :=
gen X5; *gen X4*; *gen X3*; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) *ident*(*X5*)
ident(*X6*) :=
gen X6; *gen X5*; *gen X4*; *gen X3*; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) *ident*(*X5*)
ident(*X6*) *ident*(*X7*) :=
gen X7; *gen X6*; *gen X5*; *gen X4*; *gen X3*; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) *ident*(*X5*)
ident(*X6*) *ident*(*X7*) *ident*(*X8*) :=
gen X8; *gen X7*; *gen X6*; *gen X5*; *gen X4*; *gen X3*; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) *ident*(*X5*)
ident(*X6*) *ident*(*X7*) *ident*(*X8*) *ident*(*X9*) :=
gen X9; *gen X8*; *gen X7*; *gen X6*; *gen X5*; *gen X4*; *gen X3*; *gen X2*; *gen X1*.
`Tactic Notation` "gen" *ident*(*X1*) *ident*(*X2*) *ident*(*X3*) *ident*(*X4*) *ident*(*X5*)

ident($X6$) *ident*($X7$) *ident*($X8$) *ident*($X9$) *ident*($X10$) :=
gen $X10$; *gen* $X9$; *gen* $X8$; *gen* $X7$; *gen* $X6$; *gen* $X5$; *gen* $X4$; *gen* $X3$; *gen* $X2$; *gen* $X1$.

generalizes X is a shorthand for calling `generalize` X; `clear` X. It is weaker than tactic *gen* X since it does not support dependencies. It is mainly intended for writing tactics.

```
Tactic Notation "generalizes" hyp(X) :=
  generalize X; clear X.
Tactic Notation "generalizes" hyp(X1) hyp(X2) :=
  generalizes X1; generalizes X2.
Tactic Notation "generalizes" hyp(X1) hyp(X2) hyp(X3) :=
  generalizes X1 X2; generalizes X3.
Tactic Notation "generalizes" hyp(X1) hyp(X2) hyp(X3) hyp(X4) :=
  generalizes X1 X2 X3; generalizes X4.
```

19.4.4 Naming

sets X: E is the same as `set` (X := E) `in` *, that is, it replaces all occurences of E by a fresh meta-variable X whose definition is E.

```
Tactic Notation "sets" ident(X) ":" constr(E) :=
  set (X := E) in *.
```

def_to_eq E X H applies when X := E is a local definition. It adds an assumption H: X = E and then clears the definition of X. *def_to_eq_sym* is similar except that it generates the equality H: E = X.

```
Ltac def_to_eq X HX E :=
  assert (HX : X = E) by reflexivity; clearbody X.
Ltac def_to_eq_sym X HX E :=
  assert (HX : E = X) by reflexivity; clearbody X.
```

set_eq X H: E generates the equality H: X = E, for a fresh name X, and replaces E by X in the current goal. Syntaxes *set_eq* X: E and *set_eq*: E are also available. Similarly, *set_eq* ← X H: E generates the equality H: E = X.

sets_eq X HX: E does the same but replaces E by X everywhere in the goal. *sets_eq* X HX: E `in` H replaces in H. *set_eq* X HX: E `in` ⊢ performs no substitution at all.

```
Tactic Notation "set_eq" ident(X) ident(HX) ":" constr(E) :=
  set (X := E); def_to_eq X HX E.
Tactic Notation "set_eq" ident(X) ":" constr(E) :=
  let HX := fresh "EQ" X in set_eq X HX: E.
Tactic Notation "set_eq" ":" constr(E) :=
  let X := fresh "X" in set_eq X: E.

Tactic Notation "set_eq" "<-" ident(X) ident(HX) ":" constr(E) :=
  set (X := E); def_to_eq_sym X HX E.
Tactic Notation "set_eq" "<-" ident(X) ":" constr(E) :=
```

```
  let HX := fresh "EQ" X in set_eq ← X HX: E.
Tactic Notation "set_eq" "<-" ":" constr(E) :=
  let X := fresh "X" in set_eq ← X: E.

Tactic Notation "sets_eq" ident(X) ident(HX) ":" constr(E) :=
  set (X := E) in *; def_to_eq X HX E.
Tactic Notation "sets_eq" ident(X) ":" constr(E) :=
  let HX := fresh "EQ" X in sets_eq X HX: E.
Tactic Notation "sets_eq" ":" constr(E) :=
  let X := fresh "X" in sets_eq X: E.

Tactic Notation "sets_eq" "<-" ident(X) ident(HX) ":" constr(E) :=
  set (X := E) in *; def_to_eq_sym X HX E.
Tactic Notation "sets_eq" "<-" ident(X) ":" constr(E) :=
  let HX := fresh "EQ" X in sets_eq ← X HX: E.
Tactic Notation "sets_eq" "<-" ":" constr(E) :=
  let X := fresh "X" in sets_eq ← X: E.

Tactic Notation "set_eq" ident(X) ident(HX) ":" constr(E) "in" hyp(H) :=
  set (X := E) in H; def_to_eq X HX E.
Tactic Notation "set_eq" ident(X) ":" constr(E) "in" hyp(H) :=
  let HX := fresh "EQ" X in set_eq X HX: E in H.
Tactic Notation "set_eq" ":" constr(E) "in" hyp(H) :=
  let X := fresh "X" in set_eq X: E in H.

Tactic Notation "set_eq" "<-" ident(X) ident(HX) ":" constr(E) "in" hyp(H) :=
  set (X := E) in H; def_to_eq_sym X HX E.
Tactic Notation "set_eq" "<-" ident(X) ":" constr(E) "in" hyp(H) :=
  let HX := fresh "EQ" X in set_eq ← X HX: E in H.
Tactic Notation "set_eq" "<-" ":" constr(E) "in" hyp(H) :=
  let X := fresh "X" in set_eq ← X: E in H.

Tactic Notation "set_eq" ident(X) ident(HX) ":" constr(E) "in" "|-" :=
  set (X := E) in |-; def_to_eq X HX E.
Tactic Notation "set_eq" ident(X) ":" constr(E) "in" "|-" :=
  let HX := fresh "EQ" X in set_eq X HX: E in ⊢.
Tactic Notation "set_eq" ":" constr(E) "in" "|-" :=
  let X := fresh "X" in set_eq X: E in ⊢.

Tactic Notation "set_eq" "<-" ident(X) ident(HX) ":" constr(E) "in" "|-" :=
  set (X := E) in |-; def_to_eq_sym X HX E.
Tactic Notation "set_eq" "<-" ident(X) ":" constr(E) "in" "|-" :=
  let HX := fresh "EQ" X in set_eq ← X HX: E in ⊢.
Tactic Notation "set_eq" "<-" ":" constr(E) "in" "|-" :=
  let X := fresh "X" in set_eq ← X: E in ⊢.
```

gen_eq X: E is a tactic whose purpose is to introduce equalities so as to work around the limitation of the `induction` tactic which typically loses information. *gen_eq* E `as` X replaces all occurences of term E with a fresh variable X and the equality X $= E$ as extra hypothesis to the current conclusion. In other words a conclusion C will be turned into (X $= E$) $\rightarrow$ C. *gen_eq*: E and *gen_eq*: E as X are also accepted.

```
Tactic Notation "gen_eq" ident(X) ":" constr(E) :=
  let EQ := fresh "EQ" X in sets_eq X EQ: E; revert EQ.
Tactic Notation "gen_eq" ":" constr(E) :=
  let X := fresh "X" in gen_eq X: E.
Tactic Notation "gen_eq" ":" constr(E) "as" ident(X) :=
  gen_eq X: E.
Tactic Notation "gen_eq" ident(X1) ":" constr(E1) ","
  ident(X2) ":" constr(E2) :=
  gen_eq X2: E2; gen_eq X1: E1.
Tactic Notation "gen_eq" ident(X1) ":" constr(E1) ","
  ident(X2) ":" constr(E2) "," ident(X3) ":" constr(E3) :=
  gen_eq X3: E3; gen_eq X2: E2; gen_eq X1: E1.
```

sets_let X finds the first let-expression in the goal and names its body X. *sets_eq_let* X is similar, except that it generates an explicit equality. Tactics *sets_let* X in H and *sets_eq_let* X in H allow specifying a particular hypothesis (by default, the first one that contains a let is considered).

Known limitation: it does not seem possible to support naming of multiple let-in constructs inside a term, from ltac.

```
Ltac sets_let_base tac :=
  match goal with
  | ⊢ context[let _ := ?E in _] ⇒ tac E; cbv zeta
  | H: context[let _ := ?E in _] ⊢ _ ⇒ tac E; cbv zeta in H
  end.

Ltac sets_let_in_base H tac :=
  match type of H with context[let _ := ?E in _] ⇒
    tac E; cbv zeta in H end.

Tactic Notation "sets_let" ident(X) :=
  sets_let_base ltac:(fun E ⇒ sets X: E).
Tactic Notation "sets_let" ident(X) "in" hyp(H) :=
  sets_let_in_base H ltac:(fun E ⇒ sets X: E).
Tactic Notation "sets_eq_let" ident(X) :=
  sets_let_base ltac:(fun E ⇒ sets_eq X: E).
Tactic Notation "sets_eq_let" ident(X) "in" hyp(H) :=
  sets_let_in_base H ltac:(fun E ⇒ sets_eq X: E).
```

19.5 Rewriting

rewrites E is similar to `rewrite` except that it supports the rm directives to clear hypotheses on the fly, and that it supports a list of arguments in the form *rewrites* (» *E1 E2 E3*) to indicate that *forwards* should be invoked first before *rewrites* is called.

```
Ltac rewrites_base E cont :=
  match type of E with
  | List.list Boxer ⇒ forwards_then E cont
  | _ ⇒ cont E; fast_rm_inside E
  end.
```

```
Tactic Notation "rewrites" constr(E) :=
  rewrites_base E ltac:(fun M ⇒ rewrite M ).
Tactic Notation "rewrites" constr(E) "in" hyp(H) :=
  rewrites_base E ltac:(fun M ⇒ rewrite M in H).
Tactic Notation "rewrites" constr(E) "in" "*" :=
  rewrites_base E ltac:(fun M ⇒ rewrite M in *).
Tactic Notation "rewrites" "<-" constr(E) :=
  rewrites_base E ltac:(fun M ⇒ rewrite ← M ).
Tactic Notation "rewrites" "<-" constr(E) "in" hyp(H) :=
  rewrites_base E ltac:(fun M ⇒ rewrite ← M in H).
Tactic Notation "rewrites" "<-" constr(E) "in" "*" :=
  rewrites_base E ltac:(fun M ⇒ rewrite ← M in *).
```

rewrite_all E iterates version of `rewrite` E as long as possible. Warning: this tactic can easily get into an infinite loop. Syntax for rewriting from right to left and/or into an hypothese is similar to the one of `rewrite`.

```
Tactic Notation "rewrite_all" constr(E) :=
  repeat rewrite E.
Tactic Notation "rewrite_all" "<-" constr(E) :=
  repeat rewrite ← E.
Tactic Notation "rewrite_all" constr(E) "in" ident(H) :=
  repeat rewrite E in H.
Tactic Notation "rewrite_all" "<-" constr(E) "in" ident(H) :=
  repeat rewrite ← E in H.
Tactic Notation "rewrite_all" constr(E) "in" "*" :=
  repeat rewrite E in *.
Tactic Notation "rewrite_all" "<-" constr(E) "in" "*" :=
  repeat rewrite ← E in *.
```

asserts_rewrite E asserts that an equality E holds (generating a corresponding subgoal) and rewrite it straight away in the current goal. It avoids giving a name to the equality and later clearing it. Syntax for rewriting from right to left and/or into an hypothese is similar to the one of `rewrite`. Note: the tactic *replaces* plays a similar role.

```
Ltac asserts_rewrite_tactic E action :=
  let EQ := fresh "TEMP" in (assert (EQ : E);
  [ idtac | action EQ; clear EQ ]).
```

```
Tactic Notation "asserts_rewrite" constr(E) :=
  asserts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite EQ).
Tactic Notation "asserts_rewrite" "<-" constr(E) :=
  asserts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite ← EQ).
Tactic Notation "asserts_rewrite" constr(E) "in" hyp(H) :=
  asserts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite EQ in H).
Tactic Notation "asserts_rewrite" "<-" constr(E) "in" hyp(H) :=
  asserts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite ← EQ in H).
Tactic Notation "asserts_rewrite" constr(E) "in" "*" :=
  asserts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite EQ in *).
Tactic Notation "asserts_rewrite" "<-" constr(E) "in" "*" :=
  asserts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite ← EQ in *).
```

cuts_rewrite E is the same as *asserts_rewrite* E except that subgoals are permuted.

```
Ltac cuts_rewrite_tactic E action :=
  let EQ := fresh "TEMP" in (cuts EQ: E;
  [ action EQ; clear EQ | idtac ]).
```

```
Tactic Notation "cuts_rewrite" constr(E) :=
  cuts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite EQ).
Tactic Notation "cuts_rewrite" "<-" constr(E) :=
  cuts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite ← EQ).
Tactic Notation "cuts_rewrite" constr(E) "in" hyp(H) :=
  cuts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite EQ in H).
Tactic Notation "cuts_rewrite" "<-" constr(E) "in" hyp(H) :=
  cuts_rewrite_tactic E ltac:(fun EQ ⇒ rewrite ← EQ in H).
```

rewrite_except H EQ rewrites equality EQ everywhere but in hypothesis H. Mainly useful for other tactics.

```
Ltac rewrite_except H EQ :=
  let K := fresh "TEMP" in let T := type of H in
  set (K := T) in H;
  rewrite EQ in *; unfold K in H; clear K.
```

rewrites E `at` K applies when E is of the form $T1 = T2$ rewrites the equality E at the K-th occurence of $T1$ in the current goal. Syntaxes *rewrites* ← E `at` K and *rewrites* E `at` K `in` H are also available.

```
Tactic Notation "rewrites" constr(E) "at" constr(K) :=
  match type of E with ?T1 = ?T2 ⇒
    ltac_action_at K of T1 do (rewrites E) end.
Tactic Notation "rewrites" "<-" constr(E) "at" constr(K) :=
```

```
  match type of E with ?T1 = ?T2 ⇒
    ltac_action_at K of T2 do (rewrites ← E) end.
Tactic Notation "rewrites" constr(E) "at" constr(K) "in" hyp(H) :=
  match type of E with ?T1 = ?T2 ⇒
    ltac_action_at K of T1 in H do (rewrites E in H) end.
Tactic Notation "rewrites" "<-" constr(E) "at" constr(K) "in" hyp(H) :=
  match type of E with ?T1 = ?T2 ⇒
    ltac_action_at K of T2 in H do (rewrites ← E in H) end.
```

19.5.1 Replace

replaces E `with` F is the same as `replace` E `with` F except that the equality $E = F$ is generated as first subgoal. Syntax *replaces* E `with` F `in` H is also available. Note that contrary to `replace`, *replaces* does not try to solve the equality by `assumption`. Note: *replaces* E `with` F is similar to *asserts_rewrite* $(E = F)$.

```
Tactic Notation "replaces" constr(E) "with" constr(F) :=
  let T := fresh "TEMP" in assert (T: E = F); [ | replace E with F; clear T ].
```

```
Tactic Notation "replaces" constr(E) "with" constr(F) "in" hyp(H) :=
  let T := fresh "TEMP" in assert (T: E = F); [ | replace E with F in H; clear T ].
```

replaces E `at` K `with` F replaces the K-th occurence of E with F in the current goal. Syntax *replaces* E `at` K `with` F `in` H is also available.

```
Tactic Notation "replaces" constr(E) "at" constr(K) "with" constr(F) :=
  let T := fresh "TEMP" in assert (T: E = F); [ | rewrites T at K; clear T ].
```

```
Tactic Notation "replaces" constr(E) "at" constr(K) "with" constr(F) "in" hyp(H)
:=
  let T := fresh "TEMP" in assert (T: E = F); [ | rewrites T at K in H; clear T ].
```

19.5.2 Change

changes is like `change` except that it does not silently fail to perform its task. (Note that, *changes* is implemented using `rewrite`, meaning that it might perform additional beta-reductions compared with the original `change` tactic.

```
Tactic Notation "changes" constr(E1) "with" constr(E2) "in" hyp(H) :=
  asserts_rewrite (E1 = E2) in H; [ reflexivity | ].
```

```
Tactic Notation "changes" constr(E1) "with" constr(E2) :=
  asserts_rewrite (E1 = E2); [ reflexivity | ].
```

```
Tactic Notation "changes" constr(E1) "with" constr(E2) "in" "*" :=
  asserts_rewrite (E1 = E2) in *; [ reflexivity | ].
```

19.5.3 Renaming

renames X1 to Y1, ..., *XN to YN* is a shorthand for a sequence of renaming operations `rename` *Xi into Yi*.

```
Tactic Notation "renames" ident(X1) "to" ident(Y1) :=
  rename X1 into Y1.
Tactic Notation "renames" ident(X1) "to" ident(Y1) ","
 ident(X2) "to" ident(Y2) :=
  renames X1 to Y1; renames X2 to Y2.
Tactic Notation "renames" ident(X1) "to" ident(Y1) ","
 ident(X2) "to" ident(Y2) "," ident(X3) "to" ident(Y3) :=
  renames X1 to Y1; renames X2 to Y2, X3 to Y3.
Tactic Notation "renames" ident(X1) "to" ident(Y1) ","
 ident(X2) "to" ident(Y2) "," ident(X3) "to" ident(Y3) ","
 ident(X4) "to" ident(Y4) :=
  renames X1 to Y1; renames X2 to Y2, X3 to Y3, X4 to Y4.
Tactic Notation "renames" ident(X1) "to" ident(Y1) ","
 ident(X2) "to" ident(Y2) "," ident(X3) "to" ident(Y3) ","
 ident(X4) "to" ident(Y4) "," ident(X5) "to" ident(Y5) :=
  renames X1 to Y1; renames X2 to Y2, X3 to Y3, X4 to Y4, X5 to Y5.
Tactic Notation "renames" ident(X1) "to" ident(Y1) ","
 ident(X2) "to" ident(Y2) "," ident(X3) "to" ident(Y3) ","
 ident(X4) "to" ident(Y4) "," ident(X5) "to" ident(Y5) ","
 ident(X6) "to" ident(Y6) :=
  renames X1 to Y1; renames X2 to Y2, X3 to Y3, X4 to Y4, X5 to Y5, X6 to Y6.
```

19.5.4 Unfolding

unfolds unfolds the head definition in the goal, i.e. if the goal has form *P x1 ... xN* then it calls `unfold` *P*. If the goal is an equality, it tries to unfold the head constant on the left-hand side, and otherwise tries on the right-hand side. If the goal is a product, it calls `intros` first. warning: this tactic is overriden in LibReflect.

```
Ltac apply_to_head_of E cont :=
  let go E :=
    let P := get_head E in cont P in
  match E with
  | ∀ _,_ ⇒ intros; apply_to_head_of E cont
  | ?A = ?B ⇒ first [ go A | go B ]
  | ?A ⇒ go A
  end.

Ltac unfolds_base :=
  match goal with ⊢ ?G ⇒
```

```
  apply_to_head_of G ltac:(fun P => unfold P) end.
Tactic Notation "unfolds" :=
  unfolds_base.
```

unfolds in H unfolds the head definition of hypothesis H, i.e. if H has type P *x1* ... *xN* then it calls unfold P in H.

```
Ltac unfolds_in_base H :=
  match type of H with ?G =>
   apply_to_head_of G ltac:(fun P => unfold P in H) end.
Tactic Notation "unfolds" "in" hyp(H) :=
  unfolds_in_base H.
```

unfolds in *H1*,*H2*,..,*HN* allows unfolding the head constant in several hypotheses at once.

```
Tactic Notation "unfolds" "in" hyp(H1) hyp(H2) :=
  unfolds in H1; unfolds in H2.
Tactic Notation "unfolds" "in" hyp(H1) hyp(H2) hyp(H3) :=
  unfolds in H1; unfolds in H2 H3.
Tactic Notation "unfolds" "in" hyp(H1) hyp(H2) hyp(H3) hyp(H4) :=
  unfolds in H1; unfolds in H2 H3 H4.
Tactic Notation "unfolds" "in" hyp(H1) hyp(H2) hyp(H3) hyp(H4) hyp(H5) :=
  unfolds in H1; unfolds in H2 H3 H4 H5.
```

unfolds *P1*,..,*PN* is a shortcut for unfold *P1*,..,*PN* in *.

```
Tactic Notation "unfolds" constr(F1) :=
  unfold F1 in *.
Tactic Notation "unfolds" constr(F1) "," constr(F2) :=
  unfold F1,F2 in *.
Tactic Notation "unfolds" constr(F1) "," constr(F2)
 "," constr(F3) :=
  unfold F1,F2,F3 in *.
Tactic Notation "unfolds" constr(F1) "," constr(F2)
 "," constr(F3) "," constr(F4) :=
  unfold F1,F2,F3,F4 in *.
Tactic Notation "unfolds" constr(F1) "," constr(F2)
 "," constr(F3) "," constr(F4) "," constr(F5) :=
  unfold F1,F2,F3,F4,F5 in *.
Tactic Notation "unfolds" constr(F1) "," constr(F2)
 "," constr(F3) "," constr(F4) "," constr(F5) "," constr(F6) :=
  unfold F1,F2,F3,F4,F5,F6 in *.
Tactic Notation "unfolds" constr(F1) "," constr(F2)
 "," constr(F3) "," constr(F4) "," constr(F5)
 "," constr(F6) "," constr(F7) :=
  unfold F1,F2,F3,F4,F5,F6,F7 in *.
```

```
Tactic Notation "unfolds" constr(F1) "," constr(F2)
 "," constr(F3) "," constr(F4) "," constr(F5)
 "," constr(F6) "," constr(F7) "," constr(F8) :=
  unfold F1,F2,F3,F4,F5,F6,F7,F8 in *.
```

folds P1,..,PN is a shortcut for `fold` *P1* `in` *; ..; `fold` *PN* `in` *.

```
Tactic Notation "folds" constr(H) :=
  fold H in *.
Tactic Notation "folds" constr(H1) "," constr(H2) :=
  folds H1; folds H2.
Tactic Notation "folds" constr(H1) "," constr(H2) "," constr(H3) :=
  folds H1; folds H2; folds H3.
Tactic Notation "folds" constr(H1) "," constr(H2) "," constr(H3)
 "," constr(H4) :=
  folds H1; folds H2; folds H3; folds H4.
Tactic Notation "folds" constr(H1) "," constr(H2) "," constr(H3)
 "," constr(H4) "," constr(H5) :=
  folds H1; folds H2; folds H3; folds H4; folds H5.
```

19.5.5 Simplification

simpls is a shortcut for `simpl in` *.

```
Tactic Notation "simpls" :=
  simpl in *.
```

simpls P1,..,PN is a shortcut for `simpl` *P1* `in` *; ..; `simpl` *PN* `in` *.

```
Tactic Notation "simpls" constr(F1) :=
  simpl F1 in *.
Tactic Notation "simpls" constr(F1) "," constr(F2) :=
  simpls F1; simpls F2.
Tactic Notation "simpls" constr(F1) "," constr(F2)
 "," constr(F3) :=
  simpls F1; simpls F2; simpls F3.
Tactic Notation "simpls" constr(F1) "," constr(F2)
 "," constr(F3) "," constr(F4) :=
  simpls F1; simpls F2; simpls F3; simpls F4.
```

unsimpl E replaces all occurence of X by *E*, where X is the result which the tactic `simpl` would give when applied to *E*. It is useful to undo what `simpl` has simplified too far.

```
Tactic Notation "unsimpl" constr(E) :=
  let F := (eval simpl in E) in change F with E.
```

unsimpl E `in` *H* is similar to *unsimpl E* but it applies inside a particular hypothesis *H*.

```
Tactic Notation "unsimpl" constr(E) "in" hyp(H) :=
```

```
  let F := (eval simpl in E) in change F with E in H.
```

unsimpl E in * applies *unsimpl* E everywhere possible. *unsimpls* E is a synonymous.

```
Tactic Notation "unsimpl" constr(E) "in" "*" :=
  let F := (eval simpl in E) in change F with E in *.
Tactic Notation "unsimpls" constr(E) :=
  unsimpl E in *.
```

nosimpl t protects the Coq term$t$ against some forms of simplification. See Gonthier's work for details on this trick.

```
Notation "'nosimpl' t" := (match tt with tt ⇒ t end)
  (at level 10).
```

19.5.6 Reduction

```
Tactic Notation "hnfs" := hnf in *.
```

19.5.7 Substitution

substs does the same as `subst`, except that it does not fail when there are circular equalities in the context.

```
Tactic Notation "substs" :=
  repeat (match goal with H: ?x = ?y ⊢ _ ⇒
            first [ subst x | subst y ] end).
```

Implementation of *substs below*, which allows to call `subst` on all the hypotheses that lie beyond a given position in the proof context.

```
Ltac substs_below limit :=
  match goal with H: ?T ⊢ _ ⇒
  match T with
  | limit ⇒ idtac
  | ?x = ?y ⇒
    first [ subst x; substs_below limit
          | subst y; substs_below limit
          | generalizes H; substs_below limit; intro ]
  end end.
```

substs below body E applies `subst` on all equalities that appear in the context below the first hypothesis whose body is E. If there is no such hypothesis in the context, it is equivalent to `subst`. For instance, if H is an hypothesis, then *substs below* H will substitute equalities below hypothesis H.

```
Tactic Notation "substs" "below" "body" constr(M) :=
  substs_below M.
```

substs below H applies `subst` on all equalities that appear in the context below the hypothesis named *H*. Note that the current implementation is technically incorrect since it will confuse different hypotheses with the same body.

```
Tactic Notation "substs" "below" hyp(H) :=
  match type of H with ?M ⇒ substs below body M end.
```

subst_hyp H substitutes the equality contained in the first hypothesis from the context.

```
Ltac intro_subst_hyp := fail.
```

subst_hyp H substitutes the equality contained in *H*.

```
Ltac subst_hyp_base H :=
  match type of H with
  | (_,_,_,_,_) = (_,_,_,_,_) ⇒ injection H; clear H; do 4 intro_subst_hyp
  | (_,_,_,_) = (_,_,_,_) ⇒ injection H; clear H; do 4 intro_subst_hyp
  | (_,_,_) = (_,_,_) ⇒ injection H; clear H; do 3 intro_subst_hyp
  | (_,_) = (_,_) ⇒ injection H; clear H; do 2 intro_subst_hyp
  | ?x = ?x ⇒ clear H
  | ?x = ?y ⇒ first [ subst x | subst y ]
  end.
Tactic Notation "subst_hyp" hyp(H) := subst_hyp_base H.
Ltac intro_subst_hyp ::=
  let H := fresh "TEMP" in intros H; subst_hyp H.
```

intro_subst is a shorthand for `intro` *H*; *subst_hyp H*: it introduces and substitutes the equality at the head of the current goal.

```
Tactic Notation "intro_subst" :=
  let H := fresh "TEMP" in intros H; subst_hyp H.
```

subst_local substitutes all local definition from the context

```
Ltac subst_local :=
  repeat match goal with H:=_ ⊢ _ ⇒ subst H end.
```

subst_eq E takes an equality $\mathsf{x} = t$ and replace x with t everywhere in the goal

```
Ltac subst_eq_base E :=
  let H := fresh "TEMP" in lets H: E; subst_hyp H.
Tactic Notation "subst_eq" constr(E) :=
  subst_eq_base E.
```

19.5.8 Tactics to Work with Proof Irrelevance

```
Require Import ProofIrrelevance.
```

pi_rewrite E replaces *E* of type `Prop` with a fresh unification variable, and is thus a practical way to exploit proof irrelevance, without writing explicitly `rewrite` (*proof_irrelevance* *E E'*). Particularly useful when *E'* is a big expression.

```
Ltac pi_rewrite_base E rewrite_tac :=
  let E' := fresh "TEMP" in let T := type of E in evar (E':T);
  rewrite_tac (@proof_irrelevance _ E E'); subst E'.

Tactic Notation "pi_rewrite" constr(E) :=
  pi_rewrite_base E ltac:(fun X => rewrite X).
Tactic Notation "pi_rewrite" constr(E) "in" hyp(H) :=
  pi_rewrite_base E ltac:(fun X => rewrite X in H).
```

19.5.9 Proving Equalities

Note: current implementation only supports up to arity 5

fequal is a variation on `f_equal` which has a better behaviour on equalities between n-ary tuples.

```
Ltac fequal_base :=
  let go := f_equal; [ fequal_base | ] in
  match goal with
  | ⊢ (_,_,_) = (_,_,_) => go
  | ⊢ (_,_,_,_) = (_,_,_,_) => go
  | ⊢ (_,_,_,_,_) = (_,_,_,_,_) => go
  | ⊢ (_,_,_,_,_,_) = (_,_,_,_,_,_) => go
  | ⊢ _ => f_equal
  end.

Tactic Notation "fequal" :=
  fequal_base.
```

fequals is the same as *fequal* except that it tries and solve all trivial subgoals, using `reflexivity` and `congruence` (as well as the proof-irrelevance principle). *fequals* applies to goals of the form f *x1* .. *xN* = f *y1* .. *yN* and produces some subgoals of the form *xi* = *yi*).

```
Ltac fequal_post :=
  first [ reflexivity | congruence | apply proof_irrelevance | idtac ].

Tactic Notation "fequals" :=
  fequal; fequal_post.
```

fequals_rec calls *fequals* recursively. It is equivalent to `repeat (progress fequals)`.

```
Tactic Notation "fequals_rec" :=
  repeat (progress fequals).
```

19.6 Inversion

19.6.1 Basic Inversion

invert keep H is same to `inversion` *H* except that it puts all the facts obtained in the goal. The keyword *keep* means that the hypothesis *H* should not be removed.

```
Tactic Notation "invert" "keep" hyp(H) :=
  pose ltac_mark; inversion H; gen_until_mark.
```

invert keep H `as` *X1 .. XN* is the same as `inversion` *H* `as` ... except that only hypotheses which are not variable need to be named explicitely, in a similar fashion as *introv* is used to name only hypotheses.

```
Tactic Notation "invert" "keep" hyp(H) "as" simple_intropattern(I1) :=
  invert keep H; introv I1.
Tactic Notation "invert" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) :=
  invert keep H; introv I1 I2.
Tactic Notation "invert" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) :=
  invert keep H; introv I1 I2 I3.
```

invert H is same to `inversion` *H* except that it puts all the facts obtained in the goal and clears hypothesis *H*. In other words, it is equivalent to *invert keep H*; `clear` *H*.

```
Tactic Notation "invert" hyp(H) :=
  invert keep H; clear H.
```

invert H `as` *X1 .. XN* is the same as *invert keep H* `as` *X1 .. XN* but it also clears hypothesis *H*.

```
Tactic Notation "invert_tactic" hyp(H) tactic(tac) :=
  let H' := fresh "TEMP" in rename H into H'; tac H'; clear H'.
Tactic Notation "invert" hyp(H) "as" simple_intropattern(I1) :=
  invert_tactic H (fun H => invert keep H as I1).
Tactic Notation "invert" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) :=
  invert_tactic H (fun H => invert keep H as I1 I2).
Tactic Notation "invert" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) :=
  invert_tactic H (fun H => invert keep H as I1 I2 I3).
```

19.6.2 Inversion with Substitution

Our inversion tactics is able to get rid of dependent equalities generated by `inversion`, using proof irrelevance.

```
Axiom inj_pair2 :
  ∀ (U : Type) (P : U → Type) (p : U) (x y : P p),
  existT P p x = existT P p y → x = y.

Ltac inverts_tactic H i1 i2 i3 i4 i5 i6 :=
  let rec go i1 i2 i3 i4 i5 i6 :=
    match goal with
    | ⊢ (ltac_Mark → _) ⇒ intros _
    | ⊢ (?x = ?y → _) ⇒ let H := fresh "TEMP" in intro H;
                             first [ subst x | subst y ];
                             go i1 i2 i3 i4 i5 i6
    | ⊢ (existT ?P ?p ?x = existT ?P ?p ?y → _) ⇒
          let H := fresh "TEMP" in intro H;
          generalize (@inj_pair2 _ P p x y H);
          clear H; go i1 i2 i3 i4 i5 i6
    | ⊢ (?P → ?Q) ⇒ i1; go i2 i3 i4 i5 i6 ltac:(intro)
    | ⊢ (∀ _, _) ⇒ intro; go i1 i2 i3 i4 i5 i6
    end in
  generalize ltac_mark; invert keep H; go i1 i2 i3 i4 i5 i6;
  unfold eq' in *.
```

inverts keep H is same to *invert keep H* except that it applies `subst` to all the equalities generated by the inversion.

```
Tactic Notation "inverts" "keep" hyp(H) :=
  inverts_tactic H ltac:(intro) ltac:(intro) ltac:(intro)
                   ltac:(intro) ltac:(intro) ltac:(intro).
```

inverts keep H `as` *X1* .. *XN* is the same as *invert keep H* `as` *X1* .. *XN* except that it applies `subst` to all the equalities generated by the inversion

```
Tactic Notation "inverts" "keep" hyp(H) "as" simple_intropattern(I1) :=
  inverts_tactic H ltac:(intros I1)
   ltac:(intro) ltac:(intro) ltac:(intro) ltac:(intro) ltac:(intro).
Tactic Notation "inverts" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) :=
  inverts_tactic H ltac:(intros I1) ltac:(intros I2)
   ltac:(intro) ltac:(intro) ltac:(intro) ltac:(intro).
Tactic Notation "inverts" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) :=
  inverts_tactic H ltac:(intros I1) ltac:(intros I2) ltac:(intros I3)
   ltac:(intro) ltac:(intro) ltac:(intro).
Tactic Notation "inverts" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4) :=
  inverts_tactic H ltac:(intros I1) ltac:(intros I2) ltac:(intros I3)
   ltac:(intros I4) ltac:(intro) ltac:(intro).
```

```
Tactic Notation "inverts" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4)
 simple_intropattern(I5) :=
  inverts_tactic H ltac:(intros I1) ltac:(intros I2) ltac:(intros I3)
   ltac:(intros I4) ltac:(intros I5) ltac:(intro).
Tactic Notation "inverts" "keep" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4)
 simple_intropattern(I5) simple_intropattern(I6) :=
  inverts_tactic H ltac:(intros I1) ltac:(intros I2) ltac:(intros I3)
   ltac:(intros I4) ltac:(intros I5) ltac:(intros I6).
```

inverts H is same to *inverts keep H* except that it clears hypothesis *H*.

```
Tactic Notation "inverts" hyp(H) :=
  inverts keep H; try clear H.
```

inverts H `as` *X1 .. XN* is the same as *inverts keep H* `as` *X1 .. XN* but it also clears the hypothesis *H*.

```
Tactic Notation "inverts_tactic" hyp(H) tactic(tac) :=
  let H' := fresh "TEMP" in rename H into H'; tac H'; clear H'.
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1) :=
  invert_tactic H (fun H ⇒ inverts keep H as I1).
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) :=
  invert_tactic H (fun H ⇒ inverts keep H as I1 I2).
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) :=
  invert_tactic H (fun H ⇒ inverts keep H as I1 I2 I3).
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4) :=
  invert_tactic H (fun H ⇒ inverts keep H as I1 I2 I3 I4).
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4)
 simple_intropattern(I5) :=
  invert_tactic H (fun H ⇒ inverts keep H as I1 I2 I3 I4 I5).
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4)
 simple_intropattern(I5) simple_intropattern(I6) :=
  invert_tactic H (fun H ⇒ inverts keep H as I1 I2 I3 I4 I5 I6).
```

inverts H `as` performs an inversion on hypothesis *H*, substitutes generated equalities, and put in the goal the other freshly-created hypotheses, for the user to name explicitly. *inverts keep H* `as` is the same except that it does not clear *H*. TODO: reimplement *inverts* above using this one

```
Ltac inverts_as_tactic H :=
```

```
  let rec go tt :=
    match goal with
    | ⊢ (ltac_Mark → _) ⇒ intros _
    | ⊢ (?x = ?y → _) ⇒ let H := fresh "TEMP" in intro H;
                            first [ subst x | subst y ];
                            go tt
    | ⊢ (existT ?P ?p ?x = existT ?P ?p ?y → _) ⇒
          let H := fresh "TEMP" in intro H;
          generalize (@inj_pair2 _ P p x y H);
          clear H; go tt
    | ⊢ (∀ _, _) ⇒
        intro; let H := get_last_hyp tt in mark_to_generalize H; go tt
    end in
  pose ltac_mark; inversion H;
  generalize ltac_mark; gen_until_mark;
  go tt; gen_to_generalize; unfolds ltac_to_generalize;
  unfold eq' in *.

Tactic Notation "inverts" "keep" hyp(H) "as" :=
  inverts_as_tactic H.

Tactic Notation "inverts" hyp(H) "as" :=
  inverts_as_tactic H; clear H.

Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4)
 simple_intropattern(I5) simple_intropattern(I6) simple_intropattern(I7) :=
  inverts H as; introv I1 I2 I3 I4 I5 I6 I7.
Tactic Notation "inverts" hyp(H) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4)
 simple_intropattern(I5) simple_intropattern(I6) simple_intropattern(I7)
 simple_intropattern(I8) :=
  inverts H as; introv I1 I2 I3 I4 I5 I6 I7 I8.
```

lets_inverts E as *I1* .. *IN* is intuitively equivalent to *inverts* E, with the difference that it applies to any expression and not just to the name of an hypothesis.

```
Ltac lets_inverts_base E cont :=
  let H := fresh "TEMP" in lets H: E; try cont H.

Tactic Notation "lets_inverts" constr(E) :=
  lets_inverts_base E ltac:(fun H ⇒ inverts H).
Tactic Notation "lets_inverts" constr(E) "as" simple_intropattern(I1) :=
  lets_inverts_base E ltac:(fun H ⇒ inverts H as I1).
Tactic Notation "lets_inverts" constr(E) "as" simple_intropattern(I1)
 simple_intropattern(I2) :=
  lets_inverts_base E ltac:(fun H ⇒ inverts H as I1 I2).
```

```
Tactic Notation "lets_inverts" constr(E) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) :=
  lets_inverts_base E ltac:(fun H => inverts H as I1 I2 I3).
Tactic Notation "lets_inverts" constr(E) "as" simple_intropattern(I1)
 simple_intropattern(I2) simple_intropattern(I3) simple_intropattern(I4) :=
  lets_inverts_base E ltac:(fun H => inverts H as I1 I2 I3 I4).
```

19.6.3 Injection with Substitution

Underlying implementation of *injects*

```
Ltac injects_tactic H :=
  let rec go _ :=
    match goal with
    | ⊢ (ltac_Mark → _) => intros _
    | ⊢ (?x = ?y → _) => let H := fresh "TEMP" in intro H;
                          first [ subst x | subst y | idtac ];
                          go tt
    end in
  generalize ltac_mark; injection H; go tt.
```

injects keep H takes an hypothesis *H* of the form C *a1* .. *aN* = C *b1* .. *bN* and substitute all equalities *ai* = *bi* that have been generated.

```
Tactic Notation "injects" "keep" hyp(H) :=
  injects_tactic H.
```

injects H is similar to *injects keep H* but clears the hypothesis *H*.

```
Tactic Notation "injects" hyp(H) :=
  injects_tactic H; clear H.
```

inject H as *X1* .. *XN* is the same as `injection` followed by `intros` *X1* .. *XN*

```
Tactic Notation "inject" hyp(H) :=
  injection H.
Tactic Notation "inject" hyp(H) "as" ident(X1) :=
  injection H; intros X1.
Tactic Notation "inject" hyp(H) "as" ident(X1) ident(X2) :=
  injection H; intros X1 X2.
Tactic Notation "inject" hyp(H) "as" ident(X1) ident(X2) ident(X3) :=
  injection H; intros X1 X2 X3.
Tactic Notation "inject" hyp(H) "as" ident(X1) ident(X2) ident(X3)
 ident(X4) :=
  injection H; intros X1 X2 X3 X4.
Tactic Notation "inject" hyp(H) "as" ident(X1) ident(X2) ident(X3)
 ident(X4) ident(X5) :=
  injection H; intros X1 X2 X3 X4 X5.
```

19.6.4 Inversion and Injection with Substitution –rough implementation

The tactics *inversions* and *injections* provided in this section are similar to *inverts* and *injects* except that they perform substitution on all equalities from the context and not only the ones freshly generated. The counterpart is that they have simpler implementations.

DEPRECATED: these tactics should no longer be used.

inversions keep H is the same as *inversions H* but it does not clear hypothesis *H*.

```
Tactic Notation "inversions" "keep" hyp(H) :=
  inversion H; subst.
```

inversions H is a shortcut for `inversion` *H* followed by `subst` and `clear` *H*. It is a rough implementation of *inverts keep H* which behave badly when the proof context already contains equalities. It is provided in case the better implementation turns out to be too slow.

```
Tactic Notation "inversions" hyp(H) :=
  inversion H; subst; try clear H.
```

injections keep H is the same as `injection` *H* followed by `intros` and `subst`. It is a rough implementation of *injects keep H* which behave badly when the proof context already contains equalities, or when the goal starts with a forall or an implication.

```
Tactic Notation "injections" "keep" hyp(H) :=
  injection H; intros; subst.
```

injections H is the same as `injection` *H* followed by `clear` *H* and `intros` and `subst`. It is a rough implementation of *injects keep H* which behave badly when the proof context already contains equalities, or when the goal starts with a forall or an implication.

```
Tactic Notation "injections" "keep" hyp(H) :=
  injection H; clear H; intros; subst.
```

19.6.5 Case Analysis

cases is similar to *case_eq E* except that it generates the equality in the context and not in the goal, and generates the equality the other way round. The syntax *cases E* `as` *H* allows specifying the name *H* of that hypothesis.

```
Tactic Notation "cases" constr(E) "as" ident(H) :=
  let X := fresh "TEMP" in
  set (X := E) in *; def_to_eq_sym X H E;
  destruct X.
```

```
Tactic Notation "cases" constr(E) :=
  let H := fresh "Eq" in cases E as H.
```

case_if_post H is to be defined later as a tactic to clean up hypothesis *H* and the goal. By defaults, it looks for obvious contradictions. Currently, this tactic is extended in LibReflect to clean up boolean propositions.

```
Ltac case_if_post H :=
  tryfalse.
```

case_if looks for a pattern of the form `if ?B then ?E1 else ?E2` in the goal, and perform a case analysis on B by calling `destruct B`. Subgoals containing a contradiction are discarded. *case_if* looks in the goal first, and otherwise in the first hypothesis that contains an `if` statement. *case_if* `in` *H* can be used to specify which hypothesis to consider. Syntaxes *case_if* `as` *Eq* and *case_if* `in` *H* `as` *Eq* allows to name the hypothesis coming from the case analysis.

```
Ltac case_if_on_tactic_core E Eq :=
  match type of E with
  | {_}+{_} ⇒ destruct E as [Eq | Eq]
  | _ ⇒ let X := fresh "TEMP" in
         sets_eq ← X Eq: E;
         destruct X
  end.

Ltac case_if_on_tactic E Eq :=
  case_if_on_tactic_core E Eq; case_if_post Eq.

Tactic Notation "case_if_on" constr(E) "as" simple_intropattern(Eq) :=
  case_if_on_tactic E Eq.

Tactic Notation "case_if" "as" simple_intropattern(Eq) :=
  match goal with
  | ⊢ context [if ?B then _ else _] ⇒ case_if_on B as Eq
  | K: context [if ?B then _ else _] ⊢ _ ⇒ case_if_on B as Eq
  end.

Tactic Notation "case_if" "in" hyp(H) "as" simple_intropattern(Eq) :=
  match type of H with context [if ?B then _ else _] ⇒
    case_if_on B as Eq end.

Tactic Notation "case_if" :=
  let Eq := fresh "C" in case_if as Eq.

Tactic Notation "case_if" "in" hyp(H) :=
  let Eq := fresh "C" in case_if in H as Eq.
```

cases_if is similar to *case_if* with two main differences: if it creates an equality of the form $x = y$ and then substitutes it in the goal

```
Ltac cases_if_on_tactic_core E Eq :=
  match type of E with
  | {_}+{_} ⇒ destruct E as [Eq|Eq]; try subst_hyp Eq
```

```
  | _ ⇒ let X := fresh "TEMP" in
          sets_eq ← X Eq: E;
          destruct X
  end.
```

```
Ltac cases_if_on_tactic E Eq :=
  cases_if_on_tactic_core E Eq; tryfalse; case_if_post Eq.
```

```
Tactic Notation "cases_if_on" constr(E) "as" simple_intropattern(Eq) :=
  cases_if_on_tactic E Eq.
```

```
Tactic Notation "cases_if" "as" simple_intropattern(Eq) :=
  match goal with
  | ⊢ context [if ?B then _ else _] ⇒ cases_if_on B as Eq
  | K: context [if ?B then _ else _] ⊢ _ ⇒ cases_if_on B as Eq
  end.
```

```
Tactic Notation "cases_if" "in" hyp(H) "as" simple_intropattern(Eq) :=
  match type of H with context [if ?B then _ else _] ⇒
    cases_if_on B as Eq end.
```

```
Tactic Notation "cases_if" :=
  let Eq := fresh "C" in cases_if as Eq.
```

```
Tactic Notation "cases_if" "in" hyp(H) :=
  let Eq := fresh "C" in cases_if in H as Eq.
```

case_ifs is like `repeat` *case_if*

```
Ltac case_ifs_core :=
  repeat case_if.
```

```
Tactic Notation "case_ifs" :=
  case_ifs_core.
```

destruct_if looks for a pattern of the form `if ?B then ?E1 else ?E2` in the goal, and perform a case analysis on B by calling `destruct B`. It looks in the goal first, and otherwise in the first hypothesis that contains an `if` statement.

```
Ltac destruct_if_post := tryfalse.
```

```
Tactic Notation "destruct_if"
 "as" simple_intropattern(Eq1) simple_intropattern(Eq2) :=
  match goal with
  | ⊢ context [if ?B then _ else _] ⇒ destruct B as [Eq1|Eq2]
  | K: context [if ?B then _ else _] ⊢ _ ⇒ destruct B as [Eq1|Eq2]
  end;
  destruct_if_post.
```

```
Tactic Notation "destruct_if" "in" hyp(H)
 "as" simple_intropattern(Eq1) simple_intropattern(Eq2) :=
  match type of H with context [if ?B then _ else _] ⇒
```

```
    destruct B as [Eq1|Eq2] end;
  destruct_if_post.

Tactic Notation "destruct_if" "as" simple_intropattern(Eq) :=
  destruct_if as Eq Eq.
Tactic Notation "destruct_if" "in" hyp(H) "as" simple_intropattern(Eq) :=
  destruct_if in H as Eq Eq.

Tactic Notation "destruct_if" :=
  let Eq := fresh "C" in destruct_if as Eq Eq.
Tactic Notation "destruct_if" "in" hyp(H) :=
  let Eq := fresh "C" in destruct_if in H as Eq Eq.
```

BROKEN since v8.5beta2. TODO: cleanup.

destruct_head_match performs a case analysis on the argument of the head pattern matching when the goal has the form `match ?E with ...` or `match ?E with ... = _` or `_ = match ?E with ...`. Due to the limits of Ltac, this tactic will not fail if a match does not occur. Instead, it might perform a case analysis on an unspecified subterm from the goal. Warning: experimental.

```
Ltac find_head_match T :=
  match T with context [?E] ⇒
    match T with
    | E ⇒ fail 1
    | _ ⇒ constr:(E)
    end
  end.

Ltac destruct_head_match_core cont :=
  match goal with
  | ⊢ ?T1 = ?T2 ⇒ first [ let E := find_head_match T1 in cont E
                         | let E := find_head_match T2 in cont E ]
  | ⊢ ?T1 ⇒ let E := find_head_match T1 in cont E
  end;
  destruct_if_post.

Tactic Notation "destruct_head_match" "as" simple_intropattern(I) :=
  destruct_head_match_core ltac:(fun E ⇒ destruct E as I).

Tactic Notation "destruct_head_match" :=
  destruct_head_match_core ltac:(fun E ⇒ destruct E).
```

cases' E is similar to *case_eq E* except that it generates the equality in the context and not in the goal. The syntax *cases' E* `as` *H* allows specifying the name *H* of that hypothesis.

```
Tactic Notation "cases'" constr(E) "as" ident(H) :=
  let X := fresh "TEMP" in
  set (X := E) in *; def_to_eq X H E;
  destruct X.
```

```
Tactic Notation "cases'" constr(E) :=
  let x := fresh "Eq" in cases' E as H.
```

cases_if' is similar to *cases_if* except that it generates the symmetric equality.

```
Ltac cases_if_on' E Eq :=
  match type of E with
  | {_}+{_} ⇒ destruct E as [Eq|Eq]; try subst_hyp Eq
  | _ ⇒ let X := fresh "TEMP" in
         sets_eq X Eq: E;
         destruct X
  end; case_if_post Eq.

Tactic Notation "cases_if'" "as" simple_intropattern(Eq) :=
  match goal with
  | ⊢ context [if ?B then _ else _] ⇒ cases_if_on' B Eq
  | K: context [if ?B then _ else _] ⊢ _ ⇒ cases_if_on' B Eq
  end.

Tactic Notation "cases_if'" :=
  let Eq := fresh "C" in cases_if' as Eq.
```

19.7 Induction

inductions E is a shorthand for `dependent induction` *E*. *inductions E gen X1 .. XN* is a shorthand for `dependent induction` *E generalizing X1 .. XN*.

```
From Coq Require Import Program.Equality.

Ltac inductions_post :=
  unfold eq' in *.

Tactic Notation "inductions" ident(E) :=
  dependent induction E; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) :=
  dependent induction E generalizing X1; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2) :=
  dependent induction E generalizing X1 X2; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2)
 ident(X3) :=
  dependent induction E generalizing X1 X2 X3; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2)
 ident(X3) ident(X4) :=
  dependent induction E generalizing X1 X2 X3 X4; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2)
 ident(X3) ident(X4) ident(X5) :=
  dependent induction E generalizing X1 X2 X3 X4 X5; inductions_post.
```

```
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2)
 ident(X3) ident(X4) ident(X5) ident(X6) :=
  dependent induction E generalizing X1 X2 X3 X4 X5 X6; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2)
 ident(X3) ident(X4) ident(X5) ident(X6) ident(X7) :=
  dependent induction E generalizing X1 X2 X3 X4 X5 X6 X7; inductions_post.
Tactic Notation "inductions" ident(E) "gen" ident(X1) ident(X2)
 ident(X3) ident(X4) ident(X5) ident(X6) ident(X7) ident(X8) :=
  dependent induction E generalizing X1 X2 X3 X4 X5 X6 X7 X8; inductions_post.
```

induction_wf IH: *E* X is used to apply the well-founded induction principle, for a given well-founded relation. It applies to a goal *PX* where *PX* is a proposition on X. First, it sets up the goal in the form (fun a ⇒ *P* a) X, using pattern X, and then it applies the well-founded induction principle instantiated on *E*.

Here *E* may be either:

- a proof of wf R for R of type A→A→Prop
- a binary relation of type A→A→Prop
- a measure of type A → **nat** // only when LibWf is used

  Syntaxes *induction_wf*: *E* X and *induction_wf E* X.

```
Ltac induction_wf_core_then IH E X cont :=
  let T := type of E in
  let T := eval hnf in T in
  let clearX tt :=
    first [ clear X | fail 3 "the variable on which the induction is done appears in the
hypotheses" ] in
  match T with

  | ?A → ?A → Prop ⇒
     pattern X;
     first [
       applys well_founded_ind E;
       clearX tt;
       [
       | intros X IH; cont tt ]
     | fail 2 ]
  | _ ⇒
    pattern X;
    applys well_founded_ind E;
    clearX tt;
```

```
    intros X IH;
    cont tt
  end.
Ltac induction_wf_core IH E X :=
  induction_wf_core_then IH E X ltac:(fun _ ⇒ idtac).
Tactic Notation "induction_wf" ident(IH) ":" constr(E) ident(X) :=
  induction_wf_core IH E X.
Tactic Notation "induction_wf" ":" constr(E) ident(X) :=
  let IH := fresh "IH" in induction_wf IH: E X.
Tactic Notation "induction_wf" ":" constr(E) ident(X) :=
  induction_wf: E X.
```

Induction on the height of a derivation: the helper tactic *induct_height* helps proving the equivalence of the auxiliary judgment that includes a counter for the maximal height (see LibTacticsDemos for an example)

```
From Coq Require Import Arith.Compare_dec.
From Coq Require Import omega.Omega.
Lemma induct_height_max2 : ∀ n1 n2 : nat,
  ∃ n, n1 < n ∧ n2 < n.
Proof using.
  intros. destruct (lt_dec n1 n2).
  ∃ (S n2). omega.
  ∃ (S n1). omega.
Qed.
Ltac induct_height_step x :=
  match goal with
  | H: ∃ _, _ ⊢ _ ⇒
     let n := fresh "n" in let y := fresh "x" in
     destruct H as [n ?];
     forwards (y&?&?): induct_height_max2 n x;
     induct_height_step y
  | _ ⇒ ∃ (S x); eauto
 end.
Ltac induct_height := induct_height_step O.
```

19.8 Coinduction

Tactic *cofixs IH* is like `cofix` *IH* except that the coinduction hypothesis is tagged in the form *IH*: COIND *P* instead of being just *IH*: *P*. This helps other tactics clearing the coinduction hypothesis using *clear_coind*

```
Definition COIND (P:Prop) := P.
```

```
Tactic Notation "cofixs" ident(IH) :=
  cofix IH;
  match type of IH with ?P ⇒ change P with (COIND P) in IH end.
```

Tactic *clear_coind* clears all the coinduction hypotheses, assuming that they have been tagged

```
Ltac clear_coind :=
  repeat match goal with H: COIND _ ⊢ _ ⇒ clear H end.
```

Tactic *abstracts tac* is like `abstract` *tac* except that it clears the coinduction hypotheses so that the productivity check will be happy. For example, one can use *abstracts* `omega` to obtain the same behavior as `omega` but with an auxiliary lemma being generated.

```
Tactic Notation "abstracts" tactic(tac) :=
  clear_coind; tac.
```

19.9 Decidable Equality

decides_equality is the same as *decide equality* excepts that it is able to unfold definitions at head of the current goal.

```
Ltac decides_equality_tactic :=
  first [ decide equality | progress(unfolds); decides_equality_tactic ].
Tactic Notation "decides_equality" :=
  decides_equality_tactic.
```

19.10 Equivalence

iff H can be used to prove an equivalence $P \leftrightarrow Q$ and name H the hypothesis obtained in each case. The syntaxes *iff* and *iff H1 H2* are also available to specify zero or two names. The tactic *iff* $\leftarrow$ *H* swaps the two subgoals, i.e. produces (Q -> P) as first subgoal.

```
Lemma iff_intro_swap : ∀ (P Q : Prop),
  (Q → P) → (P → Q) → (P ↔ Q).
Proof using. intuition. Qed.
Tactic Notation "iff" simple_intropattern(H1) simple_intropattern(H2) :=
  split; [ intros H1 | intros H2 ].
Tactic Notation "iff" simple_intropattern(H) :=
  iff H H.
Tactic Notation "iff" :=
  let H := fresh "H" in iff H.
Tactic Notation "iff" "<-" simple_intropattern(H1) simple_intropattern(H2) :=
  apply iff_intro_swap; [ intros H1 | intros H2 ].
Tactic Notation "iff" "<-" simple_intropattern(H) :=
```

```
  iff ← H H.
Tactic Notation "iff" "<-" :=
  let H := fresh "H" in iff ← H.
```

19.11 N-ary Conjunctions and Disjunctions

N-ary Conjunctions Splitting in Goals

Underlying implementation of *splits*.

```
Ltac splits_tactic N :=
  match N with
  | O ⇒ fail
  | S O ⇒ idtac
  | S ?N' ⇒ split; [| splits_tactic N']
  end.

Ltac unfold_goal_until_conjunction :=
  match goal with
  | ⊢ _ ∧ _ ⇒ idtac
  | _ ⇒ progress(unfolds); unfold_goal_until_conjunction
  end.

Ltac get_term_conjunction_arity T :=
  match T with
  | _ ∧ _ ∧ _ ∧ _ ∧ _ ∧ _ ∧ _ ∧ _ ⇒ constr:(8)
  | _ ∧ _ ∧ _ ∧ _ ∧ _ ∧ _ ∧ _ ⇒ constr:(7)
  | _ ∧ _ ∧ _ ∧ _ ∧ _ ∧ _ ⇒ constr:(6)
  | _ ∧ _ ∧ _ ∧ _ ∧ _ ⇒ constr:(5)
  | _ ∧ _ ∧ _ ∧ _ ⇒ constr:(4)
  | _ ∧ _ ∧ _ ⇒ constr:(3)
  | _ ∧ _ ⇒ constr:(2)
  | _ → ?T' ⇒ get_term_conjunction_arity T'
  | _ ⇒ let P := get_head T in
        let T' := eval unfold P in T in
        match T' with
        | T ⇒ fail 1
        | _ ⇒ get_term_conjunction_arity T'
        end

  end.

Ltac get_goal_conjunction_arity :=
  match goal with ⊢ ?T ⇒ get_term_conjunction_arity T end.
```

splits applies to a goal of the form (*T1* ∧ .. ∧ *TN*) and destruct it into *N* subgoals *T1* .. *TN*. If the goal is not a conjunction, then it unfolds the head definition.

```
Tactic Notation "splits" :=
  unfold_goal_until_conjunction;
  let N := get_goal_conjunction_arity in
  splits_tactic N.
```

splits N is similar to *splits*, except that it will unfold as many definitions as necessary to obtain an N-ary conjunction.

```
Tactic Notation "splits" constr(N) :=
  let N := number_to_nat N in
  splits_tactic N.
```

N-ary Conjunctions Deconstruction

Underlying implementation of *destructs*.

```
Ltac destructs_conjunction_tactic N T :=
  match N with
  | 2 ⇒ destruct T as [? ?]
  | 3 ⇒ destruct T as [? [? ?]]
   | 4 ⇒ destruct T as [? [? [? ?]]]
   | 5 ⇒ destruct T as [? [? [? [? ?]]]]
   | 6 ⇒ destruct T as [? [? [? [? [? ?]]]]]
   | 7 ⇒ destruct T as [? [? [? [? [? [? ?]]]]]]
   end.
```

destructs `T` allows destructing a term `T` which is a N-ary conjunction. It is equivalent to `destruct T as` (*H1* .. *HN*), except that it does not require to manually specify N different names.

```
Tactic Notation "destructs" constr(T) :=
  let TT := type of T in
  let N := get_term_conjunction_arity TT in
  destructs_conjunction_tactic N T.
```

destructs N `T` is equivalent to `destruct T as` (*H1* .. *HN*), except that it does not require to manually specify N different names. Remark that it is not restricted to N-ary conjunctions.

```
Tactic Notation "destructs" constr(N) constr(T) :=
  let N := number_to_nat N in
  destructs_conjunction_tactic N T.
```

Proving goals which are N-ary disjunctions

Underlying implementation of *branch*.

```
Ltac branch_tactic K N :=
  match constr:((K,N)) with
  | (_,0) ⇒ fail 1
  | (0,_) ⇒ fail 1
  | (1,1) ⇒ idtac
```

```
  | (1,_) ⇒ left
  | (S ?K', S ?N') ⇒ right; branch_tactic K' N'
  end.

Ltac unfold_goal_until_disjunction :=
  match goal with
  | ⊢ _ ∨ _ ⇒ idtac
  | _ ⇒ progress(unfolds); unfold_goal_until_disjunction
  end.

Ltac get_term_disjunction_arity T :=
  match T with
  | _ ∨ _ ∨ _ ∨ _ ∨ _ ∨ _ ∨ _ ∨ _ ⇒ constr:(8)
  | _ ∨ _ ∨ _ ∨ _ ∨ _ ∨ _ ∨ _ ⇒ constr:(7)
  | _ ∨ _ ∨ _ ∨ _ ∨ _ ∨ _ ⇒ constr:(6)
  | _ ∨ _ ∨ _ ∨ _ ∨ _ ⇒ constr:(5)
  | _ ∨ _ ∨ _ ∨ _ ⇒ constr:(4)
  | _ ∨ _ ∨ _ ⇒ constr:(3)
  | _ ∨ _ ⇒ constr:(2)
  | _ → ?T' ⇒ get_term_disjunction_arity T'
  | _ ⇒ let P := get_head T in
         let T' := eval unfold P in T in
         match T' with
         | T ⇒ fail 1
         | _ ⇒ get_term_disjunction_arity T'
         end
  end.

Ltac get_goal_disjunction_arity :=
  match goal with ⊢ ?T ⇒ get_term_disjunction_arity T end.
```

branch N applies to a goal of the form *P1* ∨ ... ∨ *PK* ∨ ... ∨ *PN* and leaves the goal *PK*. It only able to unfold the head definition (if there is one), but for more complex unfolding one should use the tactic *branch K* `of` *N*.

```
Tactic Notation "branch" constr(K) :=
  let K := number_to_nat K in
  unfold_goal_until_disjunction;
  let N := get_goal_disjunction_arity in
  branch_tactic K N.
```

branch K `of` *N* is similar to *branch K* except that the arity of the disjunction *N* is given manually, and so this version of the tactic is able to unfold definitions. In other words, applies to a goal of the form *P1* ∨ ... ∨ *PK* ∨ ... ∨ *PN* and leaves the goal *PK*.

```
Tactic Notation "branch" constr(K) "of" constr(N) :=
  let N := number_to_nat N in
  let K := number_to_nat K in
```

branch_tactic K N.

N-ary Disjunction Deconstruction
Underlying implementation of *branches*.

```
Ltac destructs_disjunction_tactic N T :=
  match N with
  | 2 ⇒ destruct T as [? | ?]
  | 3 ⇒ destruct T as [? | [? | ?]]
   | 4 ⇒ destruct T as [? | [? | [? | ?]]]
   | 5 ⇒ destruct T as [? | [? | [? | [? | ?]]]]
   end.
```

branches T allows destructing a term T which is a N-ary disjunction. It is equivalent to `destruct T as [ H1 | .. | HN ]` , and produces *N* subgoals corresponding to the *N* possible cases.

```
Tactic Notation "branches" constr(T) :=
  let TT := type of T in
  let N := get_term_disjunction_arity TT in
  destructs_disjunction_tactic N T.
```

branches N T is the same as *branches* T except that the arity is forced to *N*. This version is useful to unfold definitions on the fly.

```
Tactic Notation "branches" constr(N) constr(T) :=
  let N := number_to_nat N in
  destructs_disjunction_tactic N T.
```

branches automatically finds a hypothesis *h* that is a disjunction and destructs it.

```
Tactic Notation "branches" :=
  match goal with h: _ ∨ _ ⊢ _ ⇒ branches h end.

Ltac get_term_existential_arity T :=
  match T with
  | ∃ x1 x2 x3 x4 x5 x6 x7 x8 , _ ⇒ constr:(8)
  | ∃ x1 x2 x3 x4 x5 x6 x7 , _ ⇒ constr:(7)
  | ∃ x1 x2 x3 x4 x5 x6 , _ ⇒ constr:(6)
  | ∃ x1 x2 x3 x4 x5 , _ ⇒ constr:(5)
  | ∃ x1 x2 x3 x4 , _ ⇒ constr:(4)
  | ∃ x1 x2 x3 , _ ⇒ constr:(3)
  | ∃ x1 x2 , _ ⇒ constr:(2)
  | ∃ x1 , _ ⇒ constr:(1)
  | _ → ?T' ⇒ get_term_existential_arity T'
  | _ ⇒ let P := get_head T in
        let T' := eval unfold P in T in
        match T' with
```

```
      | T ⇒ fail 1
      | _ ⇒ get_term_existential_arity T'
      end
  end.
Ltac get_goal_existential_arity :=
  match goal with ⊢ ?T ⇒ get_term_existential_arity T end.
```

∃ *T1* ... *TN* is a shorthand for ∃ *T1*; ...; ∃ *TN*. It is intended to prove goals of the form *exist X1* .. *XN*, *P*. If an argument provided is __ (double underscore), then an evar is introduced. ∃ *T1* .. *TN* ___ is equivalent to ∃ *T1* .. *TN* __ __ __ with as many __ as possible.

```
Tactic Notation "exists_original" constr(T1) :=
  ∃ T1.
Tactic Notation "exists" constr(T1) :=
  match T1 with
  | ltac_wild ⇒ esplit
  | ltac_wilds ⇒ repeat esplit
  | _ ⇒ ∃ T1
  end.
Tactic Notation "exists" constr(T1) constr(T2) :=
  ∃ T1; ∃ T2.
Tactic Notation "exists" constr(T1) constr(T2) constr(T3) :=
  ∃ T1; ∃ T2; ∃ T3.
Tactic Notation "exists" constr(T1) constr(T2) constr(T3) constr(T4) :=
  ∃ T1; ∃ T2; ∃ T3; ∃ T4.
Tactic Notation "exists" constr(T1) constr(T2) constr(T3) constr(T4)
 constr(T5) :=
  ∃ T1; ∃ T2; ∃ T3; ∃ T4; ∃ T5.
Tactic Notation "exists" constr(T1) constr(T2) constr(T3) constr(T4)
 constr(T5) constr(T6) :=
  ∃ T1; ∃ T2; ∃ T3; ∃ T4; ∃ T5; ∃ T6.
```

For compatibility with Coq syntax, ∃ *T1*, .., *TN* is also provided.

```
Tactic Notation "exists" constr(T1) "," constr(T2) :=
  ∃ T1 T2.
Tactic Notation "exists" constr(T1) "," constr(T2) "," constr(T3) :=
  ∃ T1 T2 T3.
Tactic Notation "exists" constr(T1) "," constr(T2) "," constr(T3) "," constr(T4) :=
  ∃ T1 T2 T3 T4.
Tactic Notation "exists" constr(T1) "," constr(T2) "," constr(T3) "," constr(T4) ","
 constr(T5) :=
  ∃ T1 T2 T3 T4 T5.
Tactic Notation "exists" constr(T1) "," constr(T2) "," constr(T3) "," constr(T4) ","
```

```
 constr(T5) "," constr(T6) :=
  ∃ T1 T2 T3 T4 T5 T6.

Tactic Notation "exists___" constr(N) :=
  let rec aux N :=
    match N with
    | 0 ⇒ idtac
    | S ?N' ⇒ esplit; aux N'
    end in
  let N := number_to_nat N in aux N.

Tactic Notation "exists___" :=
  let N := get_goal_existential_arity in
  exists___ N.

Tactic Notation "exists" :=
  exists___.

Tactic Notation "exists_all" := exists___.
```

Existentials and conjunctions in hypotheses

unpack or *unpack H* destructs conjunctions and existentials in all or one hypothesis.

```
Ltac unpack_core :=
  repeat match goal with
  | H: _ ∧ _ ⊢ _ ⇒ destruct H
  | H: ∃ (varname: _), _ ⊢ _ ⇒

      let name := fresh varname in
      destruct H as [name ?]
  end.

Ltac unpack_hypothesis H :=
  try match type of H with
  | _ ∧ _ ⇒
      let h1 := fresh "TEMP" in
      let h2 := fresh "TEMP" in
      destruct H as [ h1 h2 ];
      unpack_hypothesis h1;
      unpack_hypothesis h2
  | ∃ (varname: _), _ ⇒

      let name := fresh varname in
      let body := fresh "TEMP" in
      destruct H as [name body];
      unpack_hypothesis body
  end.
```

```
Tactic Notation "unpack" :=
  unpack_core.
Tactic Notation "unpack" constr(H) :=
  unpack_hypothesis H.
```

19.12 Tactics to Prove Typeclass Instances

typeclass is an automation tactic specialized for finding typeclass instances.

```
Tactic Notation "typeclass" :=
  let go _ := eauto with typeclass_instances in
  solve [ go tt | constructor; go tt ].
```

solve_typeclass is a simpler version of *typeclass*, to use in hint tactics for resolving instances

```
Tactic Notation "solve_typeclass" :=
  solve [ eauto with typeclass_instances ].
```

19.13 Tactics to Invoke Automation

19.13.1 Definitions for Parsing Compatibility

```
Tactic Notation "f_equal" :=
  f_equal.
Tactic Notation "constructor" :=
  constructor.
Tactic Notation "simple" :=
  simpl.

Tactic Notation "split" :=
  split.

Tactic Notation "right" :=
  right.
Tactic Notation "left" :=
  left.
```

19.13.2 *hint* to Add Hints Local to a Lemma

hint E adds *E* as an hypothesis so that automation can use it. Syntax *hint E1,..,EN* is available

```
Tactic Notation "hint" constr(E) :=
  let H := fresh "Hint" in lets H: E.
```

```
Tactic Notation "hint" constr(E1) "," constr(E2) :=
  hint E1; hint E2.
Tactic Notation "hint" constr(E1) "," constr(E2) "," constr(E3) :=
  hint E1; hint E2; hint(E3).
Tactic Notation "hint" constr(E1) "," constr(E2) "," constr(E3) "," constr(E4) :=
  hint E1; hint E2; hint(E3); hint(E4 ).
```

19.13.3 *jauto*, a New Automation Tactic

jauto is better at `intuition eauto` because it can open existentials from the context. In the same time, *jauto* can be faster than `intuition eauto` because it does not destruct disjunctions from the context. The strategy of *jauto* can be summarized as follows:

- open all the existentials and conjunctions from the context
- call esplit and split on the existentials and conjunctions in the goal
- call eauto.

```
Tactic Notation "jauto" :=
  try solve [ jauto_set; eauto ].
```

```
Tactic Notation "jauto_fast" :=
  try solve [ auto | eauto | jauto ].
```

iauto is a shorthand for `intuition eauto`

```
Tactic Notation "iauto" := try solve [intuition eauto].
```

19.13.4 Definitions of Automation Tactics

The two following tactics defined the default behaviour of “light automation” and “strong automation”. These tactics may be redefined at any time using the syntax `Ltac` .. ::= ...

auto_tilde is the tactic which will be called each time a symbol ¬ is used after a tactic.

```
Ltac auto_tilde_default := auto.
Ltac auto_tilde := auto_tilde_default.
```

auto_star is the tactic which will be called each time a symbol × is used after a tactic.

```
Ltac auto_star_default := try solve [ jauto ].
Ltac auto_star := auto_star_default.
```

autos¬ is a notation for tactic *auto_tilde*. It may be followed by lemmas (or proofs terms) which auto will be able to use for solving the goal.

autos is an alias for *autos¬*

```
Tactic Notation "autos" :=
  auto_tilde.
```

```
Tactic Notation "autos" "~" :=
  auto_tilde.
Tactic Notation "autos" "~" constr(E1) :=
  lets: E1; auto_tilde.
Tactic Notation "autos" "~" constr(E1) constr(E2) :=
  lets: E1; lets: E2; auto_tilde.
Tactic Notation "autos" "~" constr(E1) constr(E2) constr(E3) :=
  lets: E1; lets: E2; lets: E3; auto_tilde.
```

autos× is a notation for tactic *auto_star*. It may be followed by lemmas (or proofs terms) which auto will be able to use for solving the goal.

```
Tactic Notation "autos" "*" :=
  auto_star.
Tactic Notation "autos" "*" constr(E1) :=
  lets: E1; auto_star.
Tactic Notation "autos" "*" constr(E1) constr(E2) :=
  lets: E1; lets: E2; auto_star.
Tactic Notation "autos" "*" constr(E1) constr(E2) constr(E3) :=
  lets: E1; lets: E2; lets: E3; auto_star.
```

auto_false is a version of `auto` able to spot some contradictions. There is an ad-hoc support for goals in ↔: split is called first. *auto_false*¬ and *auto_false*× are also available.

```
Ltac auto_false_base cont :=
  try solve [
    intros_all; try match goal with ⊢ _ ↔ _ ⇒ split end;
    solve [ cont tt | intros_all; false; cont tt ] ].

Tactic Notation "auto_false" :=
   auto_false_base ltac:(fun tt ⇒ auto).
Tactic Notation "auto_false" "~" :=
   auto_false_base ltac:(fun tt ⇒ auto_tilde).
Tactic Notation "auto_false" "*" :=
   auto_false_base ltac:(fun tt ⇒ auto_star).

Tactic Notation "dauto" :=
  dintuition eauto.
```

19.13.5 Parsing for Light Automation

Any tactic followed by the symbol ¬ will have *auto_tilde* called on all of its subgoals. Three exceptions:

- *cuts* and *asserts* only call `auto` on their first subgoal,
- `apply`¬ relies on *sapply* rather than `apply`,

- *tryfalse*¬ is defined as *tryfalse* `by` *auto_tilde*.

Some builtin tactics are not defined using tactic notations and thus cannot be extended, e.g., `simpl` and `unfold`. For these, notation such as `simpl`¬ will not be available.

```
Tactic Notation "equates" "~" constr(E) :=
  equates E; auto_tilde.
Tactic Notation "equates" "~" constr(n1) constr(n2) :=
  equates n1 n2; auto_tilde.
Tactic Notation "equates" "~" constr(n1) constr(n2) constr(n3) :=
  equates n1 n2 n3; auto_tilde.
Tactic Notation "equates" "~" constr(n1) constr(n2) constr(n3) constr(n4) :=
  equates n1 n2 n3 n4; auto_tilde.

Tactic Notation "applys_eq" "~" constr(H) constr(E) :=
  applys_eq H E; auto_tilde.
Tactic Notation "applys_eq" "~" constr(H) constr(n1) constr(n2) :=
  applys_eq H n1 n2; auto_tilde.
Tactic Notation "applys_eq" "~" constr(H) constr(n1) constr(n2) constr(n3) :=
  applys_eq H n1 n2 n3; auto_tilde.
Tactic Notation "applys_eq" "~" constr(H) constr(n1) constr(n2) constr(n3) constr(n4)
:=
  applys_eq H n1 n2 n3 n4; auto_tilde.

Tactic Notation "apply" "~" constr(H) :=
  sapply H; auto_tilde.

Tactic Notation "destruct" "~" constr(H) :=
  destruct H; auto_tilde.
Tactic Notation "destruct" "~" constr(H) "as" simple_intropattern(I) :=
  destruct H as I; auto_tilde.
Tactic Notation "f_equal" "~" :=
  f_equal; auto_tilde.
Tactic Notation "induction" "~" constr(H) :=
  induction H; auto_tilde.
Tactic Notation "inversion" "~" constr(H) :=
  inversion H; auto_tilde.
Tactic Notation "split" "~" :=
  split; auto_tilde.
Tactic Notation "subst" "~" :=
  subst; auto_tilde.
Tactic Notation "right" "~" :=
  right; auto_tilde.
Tactic Notation "left" "~" :=
  left; auto_tilde.
Tactic Notation "constructor" "~" :=
```

```
  constructor; auto_tilde.
Tactic Notation "constructors" "~" :=
  constructors; auto_tilde.
Tactic Notation "false" "~" :=
  false; auto_tilde.
Tactic Notation "false" "~" constr(E) :=
  false_then E ltac:(fun _ ⇒ auto_tilde).
Tactic Notation "false" "~" constr(E0) constr(E1) :=
  false¬ (» E0 E1).
Tactic Notation "false" "~" constr(E0) constr(E1) constr(E2) :=
  false¬ (» E0 E1 E2).
Tactic Notation "false" "~" constr(E0) constr(E1) constr(E2) constr(E3) :=
  false¬ (» E0 E1 E2 E3).
Tactic Notation "false" "~" constr(E0) constr(E1) constr(E2) constr(E3) constr(E4)
:=
  false¬ (» E0 E1 E2 E3 E4).
Tactic Notation "tryfalse" "~" :=
  try solve [ false¬ ].
Tactic Notation "asserts" "~" simple_intropattern(H) ":" constr(E) :=
  asserts H: E; [ auto_tilde | idtac ].
Tactic Notation "asserts" "~" ":" constr(E) :=
  let H := fresh "H" in asserts¬ H: E.
Tactic Notation "cuts" "~" simple_intropattern(H) ":" constr(E) :=
  cuts H: E; [ auto_tilde | idtac ].
Tactic Notation "cuts" "~" ":" constr(E) :=
  cuts: E; [ auto_tilde | idtac ].
Tactic Notation "lets" "~" simple_intropattern(I) ":" constr(E) :=
  lets I: E; auto_tilde.
Tactic Notation "lets" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) :=
  lets I: E0 A1; auto_tilde.
Tactic Notation "lets" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) :=
  lets I: E0 A1 A2; auto_tilde.
Tactic Notation "lets" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  lets I: E0 A1 A2 A3; auto_tilde.
Tactic Notation "lets" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  lets I: E0 A1 A2 A3 A4; auto_tilde.
Tactic Notation "lets" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
```

lets I: *E0 A1 A2 A3 A4 A5*; *auto_tilde*.

```
Tactic Notation "lets" "~" ":" constr(E) :=
  lets: E; auto_tilde.
Tactic Notation "lets" "~" ":" constr(E0)
 constr(A1) :=
  lets: E0 A1; auto_tilde.
Tactic Notation "lets" "~" ":" constr(E0)
 constr(A1) constr(A2) :=
  lets: E0 A1 A2; auto_tilde.
Tactic Notation "lets" "~" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  lets: E0 A1 A2 A3; auto_tilde.
Tactic Notation "lets" "~" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  lets: E0 A1 A2 A3 A4; auto_tilde.
Tactic Notation "lets" "~" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  lets: E0 A1 A2 A3 A4 A5; auto_tilde.

Tactic Notation "forwards" "~" simple_intropattern(I) ":" constr(E) :=
  forwards I: E; auto_tilde.
Tactic Notation "forwards" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) :=
  forwards I: E0 A1; auto_tilde.
Tactic Notation "forwards" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) :=
  forwards I: E0 A1 A2; auto_tilde.
Tactic Notation "forwards" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  forwards I: E0 A1 A2 A3; auto_tilde.
Tactic Notation "forwards" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  forwards I: E0 A1 A2 A3 A4; auto_tilde.
Tactic Notation "forwards" "~" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  forwards I: E0 A1 A2 A3 A4 A5; auto_tilde.

Tactic Notation "forwards" "~" ":" constr(E) :=
  forwards: E; auto_tilde.
Tactic Notation "forwards" "~" ":" constr(E0)
 constr(A1) :=
  forwards: E0 A1; auto_tilde.
Tactic Notation "forwards" "~" ":" constr(E0)
 constr(A1) constr(A2) :=
```

```
  forwards: E0 A1 A2; auto_tilde.
Tactic Notation "forwards" "~" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  forwards: E0 A1 A2 A3; auto_tilde.
Tactic Notation "forwards" "~" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  forwards: E0 A1 A2 A3 A4; auto_tilde.
Tactic Notation "forwards" "~" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  forwards: E0 A1 A2 A3 A4 A5; auto_tilde.

Tactic Notation "applys" "~" constr(H) :=
  sapply H; auto_tilde. Tactic Notation "applys" "~" constr(E0) constr(A1) :=
  applys E0 A1; auto_tilde.
Tactic Notation "applys" "~" constr(E0) constr(A1) :=
  applys E0 A1; auto_tilde.
Tactic Notation "applys" "~" constr(E0) constr(A1) constr(A2) :=
  applys E0 A1 A2; auto_tilde.
Tactic Notation "applys" "~" constr(E0) constr(A1) constr(A2) constr(A3) :=
  applys E0 A1 A2 A3; auto_tilde.
Tactic Notation "applys" "~" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  applys E0 A1 A2 A3 A4; auto_tilde.
Tactic Notation "applys" "~" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  applys E0 A1 A2 A3 A4 A5; auto_tilde.

Tactic Notation "specializes" "~" hyp(H) :=
  specializes H; auto_tilde.
Tactic Notation "specializes" "~" hyp(H) constr(A1) :=
  specializes H A1; auto_tilde.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) :=
  specializes H A1 A2; auto_tilde.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) :=
  specializes H A1 A2 A3; auto_tilde.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  specializes H A1 A2 A3 A4; auto_tilde.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  specializes H A1 A2 A3 A4 A5; auto_tilde.

Tactic Notation "fapply" "~" constr(E) :=
  fapply E; auto_tilde.
Tactic Notation "sapply" "~" constr(E) :=
```

```
  sapply E; auto_tilde.
Tactic Notation "logic" "~" constr(E) :=
  logic_base E ltac:(fun _ => auto_tilde).
Tactic Notation "intros_all" "~" :=
  intros_all; auto_tilde.
Tactic Notation "unfolds" "~" :=
  unfolds; auto_tilde.
Tactic Notation "unfolds" "~" constr(F1) :=
  unfolds F1; auto_tilde.
Tactic Notation "unfolds" "~" constr(F1) "," constr(F2) :=
  unfolds F1, F2; auto_tilde.
Tactic Notation "unfolds" "~" constr(F1) "," constr(F2) "," constr(F3) :=
  unfolds F1, F2, F3; auto_tilde.
Tactic Notation "unfolds" "~" constr(F1) "," constr(F2) "," constr(F3) ","
 constr(F4) :=
  unfolds F1, F2, F3, F4; auto_tilde.
Tactic Notation "simple" "~" :=
  simpl; auto_tilde.
Tactic Notation "simple" "~" "in" hyp(H) :=
  simpl in H; auto_tilde.
Tactic Notation "simpls" "~" :=
  simpls; auto_tilde.
Tactic Notation "hnfs" "~" :=
  hnfs; auto_tilde.
Tactic Notation "hnfs" "~" "in" hyp(H) :=
  hnf in H; auto_tilde.
Tactic Notation "substs" "~" :=
  substs; auto_tilde.
Tactic Notation "intro_hyp" "~" hyp(H) :=
  subst_hyp H; auto_tilde.
Tactic Notation "intro_subst" "~" :=
  intro_subst; auto_tilde.
Tactic Notation "subst_eq" "~" constr(E) :=
  subst_eq E; auto_tilde.
Tactic Notation "rewrite" "~" constr(E) :=
  rewrite E; auto_tilde.
Tactic Notation "rewrite" "~" "<-" constr(E) :=
  rewrite <- E; auto_tilde.
Tactic Notation "rewrite" "~" constr(E) "in" hyp(H) :=
  rewrite E in H; auto_tilde.
Tactic Notation "rewrite" "~" "<-" constr(E) "in" hyp(H) :=
```

rewrite ← E in H; *auto_tilde*.

Tactic Notation "rewrites" "~" constr(E) :=
 rewrites E; *auto_tilde*.
Tactic Notation "rewrites" "~" constr(E) "in" *hyp*(H) :=
 rewrites E in H; *auto_tilde*.
Tactic Notation "rewrites" "~" constr(E) "in" "*" :=
 rewrites E in *; *auto_tilde*.
Tactic Notation "rewrites" "~" "<-" constr(E) :=
 rewrites ← E; *auto_tilde*.
Tactic Notation "rewrites" "~" "<-" constr(E) "in" *hyp*(H) :=
 rewrites ← E in H; *auto_tilde*.
Tactic Notation "rewrites" "~" "<-" constr(E) "in" "*" :=
 rewrites ← E in *; *auto_tilde*.

Tactic Notation "rewrite_all" "~" constr(E) :=
 rewrite_all E; *auto_tilde*.
Tactic Notation "rewrite_all" "~" "<-" constr(E) :=
 rewrite_all ← E; *auto_tilde*.
Tactic Notation "rewrite_all" "~" constr(E) "in" *ident*(H) :=
 rewrite_all E in H; *auto_tilde*.
Tactic Notation "rewrite_all" "~" "<-" constr(E) "in" *ident*(H) :=
 rewrite_all ← E in H; *auto_tilde*.
Tactic Notation "rewrite_all" "~" constr(E) "in" "*" :=
 rewrite_all E in *; *auto_tilde*.
Tactic Notation "rewrite_all" "~" "<-" constr(E) "in" "*" :=
 rewrite_all ← E in *; *auto_tilde*.

Tactic Notation "asserts_rewrite" "~" constr(E) :=
 asserts_rewrite E; *auto_tilde*.
Tactic Notation "asserts_rewrite" "~" "<-" constr(E) :=
 asserts_rewrite ← E; *auto_tilde*.
Tactic Notation "asserts_rewrite" "~" constr(E) "in" *hyp*(H) :=
 asserts_rewrite E in H; *auto_tilde*.
Tactic Notation "asserts_rewrite" "~" "<-" constr(E) "in" *hyp*(H) :=
 asserts_rewrite ← E in H; *auto_tilde*.
Tactic Notation "asserts_rewrite" "~" constr(E) "in" "*" :=
 asserts_rewrite E in *; *auto_tilde*.
Tactic Notation "asserts_rewrite" "~" "<-" constr(E) "in" "*" :=
 asserts_rewrite ← E in *; *auto_tilde*.

Tactic Notation "cuts_rewrite" "~" constr(E) :=
 cuts_rewrite E; *auto_tilde*.
Tactic Notation "cuts_rewrite" "~" "<-" constr(E) :=
 cuts_rewrite ← E; *auto_tilde*.

```
Tactic Notation "cuts_rewrite" "~" constr(E) "in" hyp(H) :=
  cuts_rewrite E in H; auto_tilde.
Tactic Notation "cuts_rewrite" "~" "<-" constr(E) "in" hyp(H) :=
  cuts_rewrite ← E in H; auto_tilde.

Tactic Notation "erewrite" "~" constr(E) :=
  erewrite E; auto_tilde.

Tactic Notation "fequal" "~" :=
  fequal; auto_tilde.
Tactic Notation "fequals" "~" :=
  fequals; auto_tilde.
Tactic Notation "pi_rewrite" "~" constr(E) :=
  pi_rewrite E; auto_tilde.
Tactic Notation "pi_rewrite" "~" constr(E) "in" hyp(H) :=
  pi_rewrite E in H; auto_tilde.

Tactic Notation "invert" "~" hyp(H) :=
  invert H; auto_tilde.
Tactic Notation "inverts" "~" hyp(H) :=
  inverts H; auto_tilde.
Tactic Notation "inverts" "~" hyp(E) "as" :=
  inverts E as; auto_tilde.
Tactic Notation "injects" "~" hyp(H) :=
  injects H; auto_tilde.
Tactic Notation "inversions" "~" hyp(H) :=
  inversions H; auto_tilde.

Tactic Notation "cases" "~" constr(E) "as" ident(H) :=
  cases E as H; auto_tilde.
Tactic Notation "cases" "~" constr(E) :=
  cases E; auto_tilde.
Tactic Notation "case_if" "~" :=
  case_if; auto_tilde.
Tactic Notation "case_ifs" "~" :=
  case_ifs; auto_tilde.
Tactic Notation "case_if" "~" "in" hyp(H) :=
  case_if in H; auto_tilde.
Tactic Notation "cases_if" "~" :=
  cases_if; auto_tilde.
Tactic Notation "cases_if" "~" "in" hyp(H) :=
  cases_if in H; auto_tilde.
Tactic Notation "destruct_if" "~" :=
  destruct_if; auto_tilde.
Tactic Notation "destruct_if" "~" "in" hyp(H) :=
```

destruct_if in H; *auto_tilde*.
Tactic Notation "destruct_head_match" "~" :=
destruct_head_match; *auto_tilde*.

Tactic Notation "cases'" "~" constr(E) "as" *ident*(H) :=
cases' E as H; *auto_tilde*.
Tactic Notation "cases'" "~" constr(E) :=
cases' E; *auto_tilde*.
Tactic Notation "cases_if'" "~" "as" *ident*(H) :=
cases_if' as H; *auto_tilde*.
Tactic Notation "cases_if'" "~" :=
cases_if'; *auto_tilde*.

Tactic Notation "decides_equality" "~" :=
decides_equality; *auto_tilde*.

Tactic Notation "iff" "~" :=
iff; *auto_tilde*.
Tactic Notation "iff" "~" *simple_intropattern*(I) :=
iff I; *auto_tilde*.
Tactic Notation "splits" "~" :=
splits; *auto_tilde*.
Tactic Notation "splits" "~" constr(N) :=
splits N; *auto_tilde*.

Tactic Notation "destructs" "~" constr(T) :=
destructs T; *auto_tilde*.
Tactic Notation "destructs" "~" constr(N) constr(T) :=
destructs N T; *auto_tilde*.

Tactic Notation "branch" "~" constr(N) :=
branch N; *auto_tilde*.
Tactic Notation "branch" "~" constr(K) "of" constr(N) :=
branch K of N; *auto_tilde*.

Tactic Notation "branches" "~" :=
branches; *auto_tilde*.
Tactic Notation "branches" "~" constr(T) :=
branches T; *auto_tilde*.
Tactic Notation "branches" "~" constr(N) constr(T) :=
branches N T; *auto_tilde*.

Tactic Notation "exists" "~" :=
∃; *auto_tilde*.
Tactic Notation "exists___" "~" :=
exists___; *auto_tilde*.
Tactic Notation "exists" "~" constr($T1$) :=
∃ $T1$; *auto_tilde*.

```
Tactic Notation "exists" "~" constr(T1) constr(T2) :=
  ∃ T1 T2; auto_tilde.
Tactic Notation "exists" "~" constr(T1) constr(T2) constr(T3) :=
  ∃ T1 T2 T3; auto_tilde.
Tactic Notation "exists" "~" constr(T1) constr(T2) constr(T3) constr(T4) :=
  ∃ T1 T2 T3 T4; auto_tilde.
Tactic Notation "exists" "~" constr(T1) constr(T2) constr(T3) constr(T4)
 constr(T5) :=
  ∃ T1 T2 T3 T4 T5; auto_tilde.
Tactic Notation "exists" "~" constr(T1) constr(T2) constr(T3) constr(T4)
 constr(T5) constr(T6) :=
  ∃ T1 T2 T3 T4 T5 T6; auto_tilde.

Tactic Notation "exists" "~" constr(T1) "," constr(T2) :=
  ∃ T1 T2; auto_tilde.
Tactic Notation "exists" "~" constr(T1) "," constr(T2) "," constr(T3) :=
  ∃ T1 T2 T3; auto_tilde.
Tactic Notation "exists" "~" constr(T1) "," constr(T2) "," constr(T3) ","
 constr(T4) :=
  ∃ T1 T2 T3 T4; auto_tilde.
Tactic Notation "exists" "~" constr(T1) "," constr(T2) "," constr(T3) ","
 constr(T4) "," constr(T5) :=
  ∃ T1 T2 T3 T4 T5; auto_tilde.
Tactic Notation "exists" "~" constr(T1) "," constr(T2) "," constr(T3) ","
 constr(T4) "," constr(T5) "," constr(T6) :=
  ∃ T1 T2 T3 T4 T5 T6; auto_tilde.
```

19.13.6 Parsing for Strong Automation

Any tactic followed by the symbol × will have `auto`× called on all of its subgoals. The exceptions to these rules are the same as for light automation.

Exception: use *subs*× instead of `subst`× if you import the library *Coq.Classes.Equivalence*.

```
Tactic Notation "equates" "*" constr(E) :=
    equates E; auto_star.
Tactic Notation "equates" "*" constr(n1) constr(n2) :=
  equates n1 n2; auto_star.
Tactic Notation "equates" "*" constr(n1) constr(n2) constr(n3) :=
  equates n1 n2 n3; auto_star.
Tactic Notation "equates" "*" constr(n1) constr(n2) constr(n3) constr(n4) :=
  equates n1 n2 n3 n4; auto_star.

Tactic Notation "applys_eq" "*" constr(H) constr(E) :=
  applys_eq H E; auto_star.
```

```
Tactic Notation "applys_eq" "*" constr(H) constr(n1) constr(n2) :=
  applys_eq H n1 n2; auto_star.
Tactic Notation "applys_eq" "*" constr(H) constr(n1) constr(n2) constr(n3) :=
  applys_eq H n1 n2 n3; auto_star.
Tactic Notation "applys_eq" "*" constr(H) constr(n1) constr(n2) constr(n3) constr(n4)
:=
  applys_eq H n1 n2 n3 n4; auto_star.
Tactic Notation "apply" "*" constr(H) :=
  sapply H; auto_star.
Tactic Notation "destruct" "*" constr(H) :=
  destruct H; auto_star.
Tactic Notation "destruct" "*" constr(H) "as" simple_intropattern(I) :=
  destruct H as I; auto_star.
Tactic Notation "f_equal" "*" :=
  f_equal; auto_star.
Tactic Notation "induction" "*" constr(H) :=
  induction H; auto_star.
Tactic Notation "inversion" "*" constr(H) :=
  inversion H; auto_star.
Tactic Notation "split" "*" :=
  split; auto_star.
Tactic Notation "subs" "*" :=
  subst; auto_star.
Tactic Notation "subst" "*" :=
  subst; auto_star.
Tactic Notation "right" "*" :=
  right; auto_star.
Tactic Notation "left" "*" :=
  left; auto_star.
Tactic Notation "constructor" "*" :=
  constructor; auto_star.
Tactic Notation "constructors" "*" :=
  constructors; auto_star.
Tactic Notation "false" "*" :=
  false; auto_star.
Tactic Notation "false" "*" constr(E) :=
  false_then E ltac:(fun _ ⇒ auto_star).
Tactic Notation "false" "*" constr(E0) constr(E1) :=
  false× (» E0 E1).
Tactic Notation "false" "*" constr(E0) constr(E1) constr(E2) :=
  false× (» E0 E1 E2).
Tactic Notation "false" "*" constr(E0) constr(E1) constr(E2) constr(E3) :=
```

```
  false× (» E0 E1 E2 E3).
Tactic Notation "false" "*" constr(E0) constr(E1) constr(E2) constr(E3) constr(E4)
:=
  false× (» E0 E1 E2 E3 E4).
Tactic Notation "tryfalse" "*" :=
  try solve [ false× ].

Tactic Notation "asserts" "*" simple_intropattern(H) ":" constr(E) :=
  asserts H: E; [ auto_star | idtac ].
Tactic Notation "asserts" "*" ":" constr(E) :=
  let H := fresh "H" in asserts× H: E.
Tactic Notation "cuts" "*" simple_intropattern(H) ":" constr(E) :=
  cuts H: E; [ auto_star | idtac ].
Tactic Notation "cuts" "*" ":" constr(E) :=
  cuts: E; [ auto_star | idtac ].

Tactic Notation "lets" "*" simple_intropattern(I) ":" constr(E) :=
  lets I: E; auto_star.
Tactic Notation "lets" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) :=
  lets I: E0 A1; auto_star.
Tactic Notation "lets" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) :=
  lets I: E0 A1 A2; auto_star.
Tactic Notation "lets" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  lets I: E0 A1 A2 A3; auto_star.
Tactic Notation "lets" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  lets I: E0 A1 A2 A3 A4; auto_star.
Tactic Notation "lets" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  lets I: E0 A1 A2 A3 A4 A5; auto_star.

Tactic Notation "lets" "*" ":" constr(E) :=
  lets: E; auto_star.
Tactic Notation "lets" "*" ":" constr(E0)
 constr(A1) :=
  lets: E0 A1; auto_star.
Tactic Notation "lets" "*" ":" constr(E0)
 constr(A1) constr(A2) :=
  lets: E0 A1 A2; auto_star.
Tactic Notation "lets" "*" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  lets: E0 A1 A2 A3; auto_star.
```

```
Tactic Notation "lets" "*" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  lets: E0 A1 A2 A3 A4; auto_star.
Tactic Notation "lets" "*" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  lets: E0 A1 A2 A3 A4 A5; auto_star.

Tactic Notation "forwards" "*" simple_intropattern(I) ":" constr(E) :=
  forwards I: E; auto_star.
Tactic Notation "forwards" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) :=
  forwards I: E0 A1; auto_star.
Tactic Notation "forwards" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) :=
  forwards I: E0 A1 A2; auto_star.
Tactic Notation "forwards" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  forwards I: E0 A1 A2 A3; auto_star.
Tactic Notation "forwards" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  forwards I: E0 A1 A2 A3 A4; auto_star.
Tactic Notation "forwards" "*" simple_intropattern(I) ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  forwards I: E0 A1 A2 A3 A4 A5; auto_star.

Tactic Notation "forwards" "*" ":" constr(E) :=
  forwards: E; auto_star.
Tactic Notation "forwards" "*" ":" constr(E0)
 constr(A1) :=
  forwards: E0 A1; auto_star.
Tactic Notation "forwards" "*" ":" constr(E0)
 constr(A1) constr(A2) :=
  forwards: E0 A1 A2; auto_star.
Tactic Notation "forwards" "*" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) :=
  forwards: E0 A1 A2 A3; auto_star.
Tactic Notation "forwards" "*" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) :=
  forwards: E0 A1 A2 A3 A4; auto_star.
Tactic Notation "forwards" "*" ":" constr(E0)
 constr(A1) constr(A2) constr(A3) constr(A4) constr(A5) :=
  forwards: E0 A1 A2 A3 A4 A5; auto_star.

Tactic Notation "applys" "*" constr(H) :=
  sapply H; auto_star. Tactic Notation "applys" "*" constr(E0) constr(A1) :=
```

```
  applys E0 A1; auto_star.
Tactic Notation "applys" "*" constr(E0) constr(A1) :=
  applys E0 A1; auto_star.
Tactic Notation "applys" "*" constr(E0) constr(A1) constr(A2) :=
  applys E0 A1 A2; auto_star.
Tactic Notation "applys" "*" constr(E0) constr(A1) constr(A2) constr(A3) :=
  applys E0 A1 A2 A3; auto_star.
Tactic Notation "applys" "*" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  applys E0 A1 A2 A3 A4; auto_star.
Tactic Notation "applys" "*" constr(E0) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  applys E0 A1 A2 A3 A4 A5; auto_star.

Tactic Notation "specializes" "*" hyp(H) :=
  specializes H; auto_star.
Tactic Notation "specializes" "~" hyp(H) constr(A1) :=
  specializes H A1; auto_star.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) :=
  specializes H A1 A2; auto_star.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) :=
  specializes H A1 A2 A3; auto_star.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) constr(A4)
:=
  specializes H A1 A2 A3 A4; auto_star.
Tactic Notation "specializes" hyp(H) constr(A1) constr(A2) constr(A3) constr(A4)
constr(A5) :=
  specializes H A1 A2 A3 A4 A5; auto_star.

Tactic Notation "fapply" "*" constr(E) :=
  fapply E; auto_star.
Tactic Notation "sapply" "*" constr(E) :=
  sapply E; auto_star.

Tactic Notation "logic" constr(E) :=
  logic_base E ltac:(fun _ ⇒ auto_star).

Tactic Notation "intros_all" "*" :=
  intros_all; auto_star.

Tactic Notation "unfolds" "*" :=
  unfolds; auto_star.
Tactic Notation "unfolds" "*" constr(F1) :=
  unfolds F1; auto_star.
Tactic Notation "unfolds" "*" constr(F1) "," constr(F2) :=
  unfolds F1, F2; auto_star.
```

```
Tactic Notation "unfolds" "*" constr(F1) "," constr(F2) "," constr(F3) :=
  unfolds F1, F2, F3; auto_star.
Tactic Notation "unfolds" "*" constr(F1) "," constr(F2) "," constr(F3) ","
 constr(F4) :=
  unfolds F1, F2, F3, F4; auto_star.

Tactic Notation "simple" "*" :=
  simpl; auto_star.
Tactic Notation "simple" "*" "in" hyp(H) :=
  simpl in H; auto_star.
Tactic Notation "simpls" "*" :=
  simpls; auto_star.
Tactic Notation "hnfs" "*" :=
  hnfs; auto_star.
Tactic Notation "hnfs" "*" "in" hyp(H) :=
  hnf in H; auto_star.
Tactic Notation "substs" "*" :=
  substs; auto_star.
Tactic Notation "intro_hyp" "*" hyp(H) :=
  subst_hyp H; auto_star.
Tactic Notation "intro_subst" "*" :=
  intro_subst; auto_star.
Tactic Notation "subst_eq" "*" constr(E) :=
  subst_eq E; auto_star.

Tactic Notation "rewrite" "*" constr(E) :=
  rewrite E; auto_star.
Tactic Notation "rewrite" "*" "<-" constr(E) :=
  rewrite ← E; auto_star.
Tactic Notation "rewrite" "*" constr(E) "in" hyp(H) :=
  rewrite E in H; auto_star.
Tactic Notation "rewrite" "*" "<-" constr(E) "in" hyp(H) :=
  rewrite ← E in H; auto_star.

Tactic Notation "rewrites" "*" constr(E) :=
  rewrites E; auto_star.
Tactic Notation "rewrites" "*" constr(E) "in" hyp(H):=
  rewrites E in H; auto_star.
Tactic Notation "rewrites" "*" constr(E) "in" "*":=
  rewrites E in *; auto_star.
Tactic Notation "rewrites" "*" "<-" constr(E) :=
  rewrites ← E; auto_star.
Tactic Notation "rewrites" "*" "<-" constr(E) "in" hyp(H):=
  rewrites ← E in H; auto_star.
Tactic Notation "rewrites" "*" "<-" constr(E) "in" "*":=
```

rewrites ← E `in` *; *auto_star*.

`Tactic Notation` "rewrite_all" "*" `constr`(E) :=
rewrite_all E; *auto_star*.
`Tactic Notation` "rewrite_all" "*" "<-" `constr`(E) :=
rewrite_all ← E; *auto_star*.
`Tactic Notation` "rewrite_all" "*" `constr`(E) "in" *ident*(H) :=
rewrite_all E `in` H; *auto_star*.
`Tactic Notation` "rewrite_all" "*" "<-" `constr`(E) "in" *ident*(H) :=
rewrite_all ← E `in` H; *auto_star*.
`Tactic Notation` "rewrite_all" "*" `constr`(E) "in" "*" :=
rewrite_all E `in` *; *auto_star*.
`Tactic Notation` "rewrite_all" "*" "<-" `constr`(E) "in" "*" :=
rewrite_all ← E `in` *; *auto_star*.

`Tactic Notation` "asserts_rewrite" "*" `constr`(E) :=
asserts_rewrite E; *auto_star*.
`Tactic Notation` "asserts_rewrite" "*" "<-" `constr`(E) :=
asserts_rewrite ← E; *auto_star*.
`Tactic Notation` "asserts_rewrite" "*" `constr`(E) "in" *hyp*(H) :=
asserts_rewrite E; *auto_star*.
`Tactic Notation` "asserts_rewrite" "*" "<-" `constr`(E) "in" *hyp*(H) :=
asserts_rewrite ← E; *auto_star*.
`Tactic Notation` "asserts_rewrite" "*" `constr`(E) "in" "*" :=
asserts_rewrite E `in` *; *auto_tilde*.
`Tactic Notation` "asserts_rewrite" "*" "<-" `constr`(E) "in" "*" :=
asserts_rewrite ← E `in` *; *auto_tilde*.

`Tactic Notation` "cuts_rewrite" "*" `constr`(E) :=
cuts_rewrite E; *auto_star*.
`Tactic Notation` "cuts_rewrite" "*" "<-" `constr`(E) :=
cuts_rewrite ← E; *auto_star*.
`Tactic Notation` "cuts_rewrite" "*" `constr`(E) "in" *hyp*(H) :=
cuts_rewrite E `in` H; *auto_star*.
`Tactic Notation` "cuts_rewrite" "*" "<-" `constr`(E) "in" *hyp*(H) :=
cuts_rewrite ← E `in` H; *auto_star*.

`Tactic Notation` "erewrite" "*" `constr`(E) :=
erewrite E; *auto_star*.

`Tactic Notation` "fequal" "*" :=
fequal; *auto_star*.
`Tactic Notation` "fequals" "*" :=
fequals; *auto_star*.
`Tactic Notation` "pi_rewrite" "*" `constr`(E) :=
pi_rewrite E; *auto_star*.

```
Tactic Notation "pi_rewrite" "*" constr(E) "in" hyp(H) :=
  pi_rewrite E in H; auto_star.

Tactic Notation "invert" "*" hyp(H) :=
  invert H; auto_star.
Tactic Notation "inverts" "*" hyp(H) :=
  inverts H; auto_star.
Tactic Notation "inverts" "*" hyp(E) "as" :=
  inverts E as; auto_star.
Tactic Notation "injects" "*" hyp(H) :=
  injects H; auto_star.
Tactic Notation "inversions" "*" hyp(H) :=
  inversions H; auto_star.

Tactic Notation "cases" "*" constr(E) "as" ident(H) :=
  cases E as H; auto_star.
Tactic Notation "cases" "*" constr(E) :=
  cases E; auto_star.
Tactic Notation "case_if" "*" :=
  case_if; auto_star.
Tactic Notation "case_ifs" "*" :=
  case_ifs; auto_star.
Tactic Notation "case_if" "*" "in" hyp(H) :=
  case_if in H; auto_star.
Tactic Notation "cases_if" "*" :=
  cases_if; auto_star.
Tactic Notation "cases_if" "*" "in" hyp(H) :=
  cases_if in H; auto_star.
 Tactic Notation "destruct_if" "*" :=
  destruct_if; auto_star.
Tactic Notation "destruct_if" "*" "in" hyp(H) :=
  destruct_if in H; auto_star.
Tactic Notation "destruct_head_match" "*" :=
  destruct_head_match; auto_star.

Tactic Notation "cases'" "*" constr(E) "as" ident(H) :=
  cases' E as H; auto_star.
Tactic Notation "cases'" "*" constr(E) :=
  cases' E; auto_star.
Tactic Notation "cases_if'" "*" "as" ident(H) :=
  cases_if' as H; auto_star.
Tactic Notation "cases_if'" "*" :=
  cases_if'; auto_star.

Tactic Notation "decides_equality" "*" :=
```

decides_equality; *auto_star*.

Tactic Notation "iff" "*" :=
  iff; *auto_star*.
Tactic Notation "iff" "*" *simple_intropattern*(I) :=
  iff I; *auto_star*.
Tactic Notation "splits" "*" :=
  splits; *auto_star*.
Tactic Notation "splits" "*" constr(N) :=
  splits N; *auto_star*.

Tactic Notation "destructs" "*" constr(T) :=
  destructs T; *auto_star*.
Tactic Notation "destructs" "*" constr(N) constr(T) :=
  destructs N T; *auto_star*.

Tactic Notation "branch" "*" constr(N) :=
  branch N; *auto_star*.
Tactic Notation "branch" "*" constr(K) "of" constr(N) :=
  branch K of N; *auto_star*.

Tactic Notation "branches" "*" constr(T) :=
  branches T; *auto_star*.
Tactic Notation "branches" "*" constr(N) constr(T) :=
  branches N T; *auto_star*.

Tactic Notation "exists" "*" :=
  ∃; *auto_star*.
Tactic Notation "exists___" "*" :=
  exists___; *auto_star*.
Tactic Notation "exists" "*" constr($T1$) :=
  ∃ $T1$; *auto_star*.
Tactic Notation "exists" "*" constr($T1$) constr($T2$) :=
  ∃ $T1$ $T2$; *auto_star*.
Tactic Notation "exists" "*" constr($T1$) constr($T2$) constr($T3$) :=
  ∃ $T1$ $T2$ $T3$; *auto_star*.
Tactic Notation "exists" "*" constr($T1$) constr($T2$) constr($T3$) constr($T4$) :=
  ∃ $T1$ $T2$ $T3$ $T4$; *auto_star*.
Tactic Notation "exists" "*" constr($T1$) constr($T2$) constr($T3$) constr($T4$)
 constr($T5$) :=
  ∃ $T1$ $T2$ $T3$ $T4$ $T5$; *auto_star*.
Tactic Notation "exists" "*" constr($T1$) constr($T2$) constr($T3$) constr($T4$)
 constr($T5$) constr($T6$) :=
  ∃ $T1$ $T2$ $T3$ $T4$ $T5$ $T6$; *auto_star*.

Tactic Notation "exists" "*" constr($T1$) "," constr($T2$) :=
  ∃ $T1$ $T2$; *auto_star*.

```
Tactic Notation "exists" "*" constr(T1) "," constr(T2) "," constr(T3) :=
  ∃ T1 T2 T3; auto_star.
Tactic Notation "exists" "*" constr(T1) "," constr(T2) "," constr(T3) ","
  constr(T4) :=
  ∃ T1 T2 T3 T4; auto_star.
Tactic Notation "exists" "*" constr(T1) "," constr(T2) "," constr(T3) ","
 constr(T4) "," constr(T5) :=
  ∃ T1 T2 T3 T4 T5; auto_star.
Tactic Notation "exists" "*" constr(T1) "," constr(T2) "," constr(T3) ","
 constr(T4) "," constr(T5) "," constr(T6) :=
  ∃ T1 T2 T3 T4 T5 T6; auto_star.
```

19.14 Tactics to Sort Out the Proof Context

19.14.1 Hiding Hypotheses

```
Definition ltac_something (P:Type) (e:P) := e.
Notation "'Something'" :=
  (@ltac_something _ _).
Lemma ltac_something_eq : ∀ (e:Type),
  e = (@ltac_something _ e).
Proof using. auto. Qed.
Lemma ltac_something_hide : ∀ (e:Type),
  e → (@ltac_something _ e).
Proof using. auto. Qed.
Lemma ltac_something_show : ∀ (e:Type),
  (@ltac_something _ e) → e.
Proof using. auto. Qed.
```

hide_def x and *show_def* x can be used to hide/show the body of the definition x.

```
Tactic Notation "hide_def" hyp(x) :=
  let x' := constr:(x) in
  let T := eval unfold x in x' in
  change T with (@ltac_something _ T) in x.
Tactic Notation "show_def" hyp(x) :=
  let x' := constr:(x) in
  let U := eval unfold x in x' in
  match U with @ltac_something _ ?T ⇒
    change U with T in x end.
```

show_def unfolds *Something* in the goal

```
Tactic Notation "show_def" :=
  unfold ltac_something.
Tactic Notation "show_def" "in" hyp(H) :=
  unfold ltac_something in H.
Tactic Notation "show_def" "in" "*" :=
  unfold ltac_something in *.
```

hide_defs and *show_defs* applies to all definitions

```
Tactic Notation "hide_defs" :=
  repeat match goal with H := ?T ⊢ _ ⇒
    match T with
    | @ltac_something _ _ ⇒ fail 1
    | _ ⇒ change T with (@ltac_something _ T) in H
    end
  end.
```

```
Tactic Notation "show_defs" :=
  repeat match goal with H := (@ltac_something _ ?T) ⊢ _ ⇒
    change (@ltac_something _ T) with T in H end.
```

hide_hyp H replaces the type of H with the notation *Something* and *show_hyp* H reveals the type of the hypothesis. Note that the hidden type of H remains convertible the real type of H.

```
Tactic Notation "show_hyp" hyp(H) :=
  apply ltac_something_show in H.
```

```
Tactic Notation "hide_hyp" hyp(H) :=
  apply ltac_something_hide in H.
```

hide_hyps and *show_hyps* can be used to hide/show all hypotheses of type `Prop`.

```
Tactic Notation "show_hyps" :=
  repeat match goal with
    H: @ltac_something _ _ ⊢ _ ⇒ show_hyp H end.
```

```
Tactic Notation "hide_hyps" :=
  repeat match goal with H: ?T ⊢ _ ⇒
    match type of T with
    | Prop ⇒
      match T with
      | @ltac_something _ _ ⇒ fail 2
      | _ ⇒ hide_hyp H
      end
    | _ ⇒ fail 1
    end
  end.
```

hide H and *show H* automatically select between *hide_hyp* or *hide_def*, and *show_hyp* or *show_def*. Similarly *hide_all* and *show_all* apply to all.

```
Tactic Notation "hide" hyp(H) :=
  first [hide_def H | hide_hyp H].

Tactic Notation "show" hyp(H) :=
  first [show_def H | show_hyp H].

Tactic Notation "hide_all" :=
  hide_hyps; hide_defs.

Tactic Notation "show_all" :=
  unfold ltac_something in *.
```

hide_term E can be used to hide a term from the goal. *show_term* or *show_term E* can be used to reveal it. *hide_term E* in *H* can be used to specify an hypothesis.

```
Tactic Notation "hide_term" constr(E) :=
  change E with (@ltac_something _ E).
Tactic Notation "show_term" constr(E) :=
  change (@ltac_something _ E) with E.
Tactic Notation "show_term" :=
  unfold ltac_something.

Tactic Notation "hide_term" constr(E) "in" hyp(H) :=
  change E with (@ltac_something _ E) in H.
Tactic Notation "show_term" constr(E) "in" hyp(H) :=
  change (@ltac_something _ E) with E in H.
Tactic Notation "show_term" "in" hyp(H) :=
  unfold ltac_something in H.
```

show_unfold R unfolds the definition of R and reveals the hidden definition of R. – todo:test, and implement using unfold simply

```
Tactic Notation "show_unfold" constr(R1) :=
  unfold R1; show_def.
Tactic Notation "show_unfold" constr(R1) "," constr(R2) :=
  unfold R1, R2; show_def.
```

19.14.2 Sorting Hypotheses

sort sorts out hypotheses from the context by moving all the propositions (hypotheses of type Prop) to the bottom of the context.

```
Ltac sort_tactic :=
  try match goal with H: ?T ⊢ _ ⇒
  match type of T with Prop ⇒
    generalizes H; (try sort_tactic); intro
  end end.
```

```
Tactic Notation "sort" :=
  sort_tactic.
```

19.14.3 Clearing Hypotheses

clears X1 ... XN is a variation on `clear` which clears the variables *X1..XN* as well as all the hypotheses which depend on them. Contrary to `clear`, it never fails.

```
Tactic Notation "clears" ident(X1) :=
  let rec doit _ :=
  match goal with
  | H:context[X1] ⊢ _ ⇒ clear H; try (doit tt)
  | _ ⇒ clear X1
  end in doit tt.
Tactic Notation "clears" ident(X1) ident(X2) :=
  clears X1; clears X2.
Tactic Notation "clears" ident(X1) ident(X2) ident(X3) :=
  clears X1; clears X2; clears X3.
Tactic Notation "clears" ident(X1) ident(X2) ident(X3) ident(X4) :=
  clears X1; clears X2; clears X3; clears X4.
Tactic Notation "clears" ident(X1) ident(X2) ident(X3) ident(X4)
 ident(X5) :=
  clears X1; clears X2; clears X3; clears X4; clears X5.
Tactic Notation "clears" ident(X1) ident(X2) ident(X3) ident(X4)
 ident(X5) ident(X6) :=
  clears X1; clears X2; clears X3; clears X4; clears X5; clears X6.
```

clears (without any argument) clears all the unused variables from the context. In other words, it removes any variable which is not a proposition (i.e. not of type Prop) and which does not appear in another hypothesis nor in the goal.

```
Ltac clears_tactic :=
  match goal with H: ?T ⊢ _ ⇒
  match type of T with
  | Prop ⇒ generalizes H; (try clears_tactic); intro
  | ?TT ⇒ clear H; (try clears_tactic)
  | ?TT ⇒ generalizes H; (try clears_tactic); intro
  end end.
```

```
Tactic Notation "clears" :=
  clears_tactic.
```

clears_all clears all the hypotheses from the context that can be cleared. It leaves only the hypotheses that are mentioned in the goal.

```
Ltac clears_or_generalizes_all_core :=
  repeat match goal with H: _ ⊢ _ ⇒
```

```
      first [ clear H | generalizes H] end.
Tactic Notation "clears_all" :=
  generalize ltac_mark;
  clears_or_generalizes_all_core;
  intro_until_mark.
```

clears_but H1 H2 .. HN clears all hypotheses except the one that are mentioned and those that cannot be cleared.

```
Ltac clears_but_core cont :=
  generalize ltac_mark;
  cont tt;
  clears_or_generalizes_all_core;
  intro_until_mark.

Tactic Notation "clears_but" :=
  clears_but_core ltac:(fun _ ⇒ idtac).
Tactic Notation "clears_but" ident(H1) :=
  clears_but_core ltac:(fun _ ⇒ gen H1).
Tactic Notation "clears_but" ident(H1) ident(H2) :=
  clears_but_core ltac:(fun _ ⇒ gen H1 H2).
Tactic Notation "clears_but" ident(H1) ident(H2) ident(H3) :=
  clears_but_core ltac:(fun _ ⇒ gen H1 H2 H3).
Tactic Notation "clears_but" ident(H1) ident(H2) ident(H3) ident(H4) :=
  clears_but_core ltac:(fun _ ⇒ gen H1 H2 H3 H4).
Tactic Notation "clears_but" ident(H1) ident(H2) ident(H3) ident(H4) ident(H5) :=
  clears_but_core ltac:(fun _ ⇒ gen H1 H2 H3 H4 H5).

Lemma demo_clears_all_and_clears_but :
  ∀ x y:nat, y < 2 → x = x → x ≥ 2 → x < 3 → True.
Proof using.
  introv M1 M2 M3. dup 6.
  clears_all. auto.
  clears_but M3. auto.
  clears_but y. auto.
  clears_but x. auto.
  clears_but M2 M3. auto.
  clears_but x y. auto.
Qed.
```

clears_last clears the last hypothesis in the context. *clears_last N* clears the last *N* hypotheses in the context.

```
Tactic Notation "clears_last" :=
  match goal with H: ?T ⊢ _ ⇒ clear H end.

Ltac clears_last_base N :=
```

```
  match number_to_nat N with
  | 0 ⇒ idtac
  | S ?p ⇒ clears_last; clears_last_base p
  end.
Tactic Notation "clears_last" constr(N) :=
  clears_last_base N.
```

19.15 Tactics for Development Purposes

19.15.1 Skipping Subgoals

```
Tactic Notation "skip" :=
  admit.
```

demo is like *admit* but it documents the fact that admit is intended

```
Tactic Notation "demo" :=
  skip.
```

admits *H*: T adds an assumption named *H* of type T to the current context, blindly assuming that it is true. *admit*: T is another possible syntax. Note that H may be an intro pattern.

```
Tactic Notation "admits" simple_intropattern(I) ":" constr(T) :=
  asserts I: T; [ skip | ].
Tactic Notation "admits" ":" constr(T) :=
  let H := fresh "TEMP" in admits H: T.
Tactic Notation "admits" "~" ":" constr(T) :=
  admits: T; auto_tilde.
Tactic Notation "admits" "*" ":" constr(T) :=
  admits: T; auto_star.
```

admit_cuts T simply replaces the current goal with T.

```
Tactic Notation "admit_cuts" constr(T) :=
  cuts: T; [ skip | ].
```

admit_goal *H* applies to any goal. It simply assumes the current goal to be true. The assumption is named "H". It is useful to set up proof by induction or coinduction. Syntax *admit_goal* is also accepted.

```
Tactic Notation "admit_goal" ident(H) :=
  match goal with ⊢ ?G ⇒ admits H: G end.
Tactic Notation "admit_goal" :=
  let IH := fresh "IH" in admit_goal IH.
```

admit_rewrite T can be applied when T is an equality. It blindly assumes this equality to be true, and rewrite it in the goal.

```
Tactic Notation "admit_rewrite" constr(T) :=
  let M := fresh "TEMP" in admits M: T; rewrite M; clear M.
```

admit_rewrite T in *H* is similar as *admit_rewrite*, except that it rewrites in hypothesis *H*.

```
Tactic Notation "admit_rewrite" constr(T) "in" hyp(H) :=
  let M := fresh "TEMP" in admits M: T; rewrite M in H; clear M.
```

admit_rewrites_all T is similar as *admit_rewrite*, except that it rewrites everywhere (goal and all hypotheses).

```
Tactic Notation "admit_rewrite_all" constr(T) :=
  let M := fresh "TEMP" in admits M: T; rewrite_all M; clear M.
```

forwards_nounfold_admit_sides_then *E* ltac:(fun *K* ⇒ ..) is like *forwards*: *E* but it provides the resulting term to a continuation, under the name *K*, and it admits any side-condition produced by the instantiation of *E*, using the *skip* tactic.

```
Inductive ltac_goal_to_discard := ltac_goal_to_discard_intro.

Ltac forwards_nounfold_admit_sides_then S cont :=
  let MARK := fresh "TEMP" in
  generalize ltac_goal_to_discard_intro;
  intro MARK;
  forwards_nounfold_then S ltac:(fun K ⇒
    clear MARK;
    cont K);
  match goal with
  | MARK: ltac_goal_to_discard ⊢ _ ⇒ skip
  | _ ⇒ idtac
  end.
```

19.16 Compatibility with standard library

The module Program contains definitions that conflict with the current module. If you import Program, either directly or indirectly (e.g., through *Setoid* or ZArith), you will need to import the compability definitions through the top-level command: Import LibTacticsCompatibility.

```
Module LibTacticsCompatibility.
  Tactic Notation "apply" "*" constr(H) :=
    sapply H; auto_star.
  Tactic Notation "subst" "*" :=
    subst; auto_star.
```

```
End LibTacticsCompatibility.
Open Scope nat_scope.
```

Chapter 20

UseTactics: Tactic Library for Coq: A Gentle Introduction

Coq comes with a set of builtin tactics, such as `reflexivity`, `intros`, `inversion` and so on. While it is possible to conduct proofs using only those tactics, you can significantly increase your productivity by working with a set of more powerful tactics. This chapter describes a number of such useful tactics, which, for various reasons, are not yet available by default in Coq. These tactics are defined in the *LibTactics.v* file.

```
Set Warnings "-notation-overridden,-parsing".
From Coq Require Import Arith.Arith.
From PLF Require Import Maps.
From PLF Require Import Imp.
From PLF Require Import Types.
From PLF Require Import Smallstep.
From PLF Require Import LibTactics.
From PLF Require Stlc.
From PLF Require Equiv.
From PLF Require Imp.
From PLF Require References.
From PLF Require Smallstep.
From PLF Require Hoare.
From PLF Require Sub.
```

Remark: SSReflect is another package providing powerful tactics. The library "LibTactics" differs from "SSReflect" in two respects:

- "SSReflect" was primarily developed for proving mathematical theorems, whereas "LibTactics" was primarily developed for proving theorems on programming languages. In particular, "LibTactics" provides a number of useful tactics that have no counterpart in the "SSReflect" package.

- "SSReflect" entirely rethinks the presentation of tactics, whereas "LibTactics" mostly stick to the traditional presentation of Coq tactics, simply providing a number of additional tactics. For this reason, "LibTactics" is probably easier to get started with than "SSReflect".

This chapter is a tutorial focusing on the most useful features from the "LibTactics" library. It does not aim at presenting all the features of "LibTactics". The detailed specification of tactics can be found in the source file *LibTactics.v*. Further documentation as well as demos can be found at http://www.chargueraud.org/softs/tlc/.

In this tutorial, tactics are presented using examples taken from the core chapters of the "Software Foundations" course. To illustrate the various ways in which a given tactic can be used, we use a tactic that duplicates a given goal. More precisely, *dup* produces two copies of the current goal, and *dup* `n` produces `n` copies of it.

20.1 Tactics for Naming and Performing Inversion

This section presents the following tactics:

- *introv*, for naming hypotheses more efficiently,
- *inverts*, for improving the `inversion` tactic.

20.1.1 The Tactic *introv*

```
Module IntrovExamples.
  Import Stlc.
  Import Imp.
  Import STLC.
```

The tactic *introv* allows to automatically introduce the variables of a theorem and explicitly name the hypotheses involved. In the example shown next, the variables *c*, *st*, *st1* and *st2* involved in the statement of determinism need not be named explicitly, because their name where already given in the statement of the lemma. On the contrary, it is useful to provide names for the two hypotheses, which we name *E1* and *E2*, respectively.

```
Theorem ceval_deterministic: ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  introv E1 E2. Abort.
```

When there is no hypothesis to be named, one can call *introv* without any argument.

```
Theorem dist_exists_or : ∀ (X:Type) (P Q : X → Prop),
```

```
  (∃ x, P x ∨ Q x) ↔ (∃ x, P x) ∨ (∃ x, Q x).
Proof.
  introv. Abort.
```

The tactic *introv* also applies to statements in which ∀ and → are interleaved.

```
Theorem ceval_deterministic': ∀ c st st1,
  (st =[ c ]=> st1) →
  ∀ st2,
  (st =[ c ]=> st2) →
  st1 = st2.
Proof.
  introv E1 E2. Abort.
```

Like the arguments of `intros`, the arguments of *introv* can be structured patterns.

```
Theorem exists_impl: ∀ X (P : X → Prop) (Q : Prop) (R : Prop),
  (∀ x, P x → Q) →
  ((∃ x, P x) → Q).
Proof.
  introv [x H2]. eauto.
Qed.
```

Remark: the tactic *introv* works even when definitions need to be unfolded in order to reveal hypotheses.

```
End IntrovExamples.
```

20.1.2 The Tactic *inverts*

```
Module InvertsExamples.
  Import Stlc.
  Import Equiv.
  Import Imp.
  Import STLC.
```

The `inversion` tactic of Coq is not very satisfying for three reasons. First, it produces a bunch of equalities which one typically wants to substitute away, using `subst`. Second, it introduces meaningless names for hypotheses. Third, a call to `inversion` *H* does not remove *H* from the context, even though in most cases an hypothesis is no longer needed after being inverted. The tactic *inverts* address all of these three issues. It is intented to be used in place of the tactic `inversion`.

The following example illustrates how the tactic *inverts H* behaves mostly like `inversion` *H* except that it performs some substitutions in order to eliminate the trivial equalities that are being produced by `inversion`.

```
Theorem skip_left: ∀ c,
  cequiv (SKIP;; c) c.
```

```
Proof.
  introv. split; intros H.
  dup.   - inversion H. subst. inversion H2. subst. assumption.
  - inverts H. inverts H2. assumption.
Abort.
```

A slightly more interesting example appears next.

```
Theorem ceval_deterministic: ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  introv E1 E2. generalize dependent st2.
  induction E1; intros st2 E2.
  admit. admit.   dup.   - inversion E2. subst. admit.
  - inverts E2. admit.
Abort.
```

The tactic *inverts H* `as`. is like *inverts H* except that the variables and hypotheses being produced are placed in the goal rather than in the context. This strategy allows naming those new variables and hypotheses explicitly, using either `intros` or *introv*.

```
Theorem ceval_deterministic': ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  introv E1 E2. generalize dependent st2.
  (induction E1); intros st2 E2;
    inverts E2 as.
  - reflexivity.
  -

    subst n.
    reflexivity.
  -

    intros st3 Red1 Red2.
    assert (st' = st3) as EQ1.
    { apply IHE1_1; assumption. }
    subst st3.
    apply IHE1_2. assumption.
  -
```

```
    intros.
    apply IHE1. assumption.
  -
    intros.
    rewrite H in H5. inversion H5.
Abort.
```

In the particular case where a call to `inversion` produces a single subgoal, one can use the syntax *inverts H* `as` *H1 H2 H3* for calling *inverts* and naming the new hypotheses *H1*, *H2* and *H3*. In other words, the tactic *inverts H* `as` *H1 H2 H3* is equivalent to *inverts H* `as`; *introv H1 H2 H3*. An example follows.

```
Theorem skip_left': ∀ c,
  cequiv (SKIP;; c) c.
Proof.
  introv. split; intros H.
  inverts H as U V.   inverts U. assumption.
Abort.
```

A more involved example appears next. In particular, this example shows that the name of the hypothesis being inverted can be reused.

```
Example typing_nonexample_1 :
  ¬ ∃ T,
      has_type empty
        (abs x Bool
            (abs y Bool
               (app (var x) (var y))))
        T.
Proof.
  dup 3.

  - intros C. destruct C.
  inversion H. subst. clear H.
  inversion H5. subst. clear H5.
  inversion H4. subst. clear H4.
  inversion H2. subst. clear H2.
  inversion H1.

  - intros C. destruct C.
  inverts H as H1.
  inverts H1 as H2.
  inverts H2 as H3 H4.
  inverts H3 as H5.
  inverts H5.

  - intros C. destruct C.
  inverts H as H.
```

```
  inverts H as H.
  inverts H as H1 H2.
  inverts H1 as H1.
  inverts H1.
Qed.
End InvertsExamples.
```

Note: in the rare cases where one needs to perform an inversion on an hypothesis H without clearing H from the context, one can use the tactic *inverts keep H*, where the keyword *keep* indicates that the hypothesis should be kept in the context.

20.2 Tactics for N-ary Connectives

Because Coq encodes conjunctions and disjunctions using binary constructors ∧ and ∨, working with a conjunction or a disjunction of N facts can sometimes be quite cumbursome. For this reason, "LibTactics" provides tactics offering direct support for n-ary conjunctions and disjunctions. It also provides direct support for n-ary existententials.

This section presents the following tactics:

- *splits* for decomposing n-ary conjunctions,
- *branch* for decomposing n-ary disjunctions

```
Module NaryExamples.
  Import References.
  Import Smallstep.
  Import STLCRef.
```

20.2.1 The Tactic *splits*

The tactic *splits* applies to a goal made of a conjunction of `n` propositions and it produces `n` subgoals. For example, it decomposes the goal *G1* ∧ *G2* ∧ *G3* into the three subgoals *G1*, *G2* and *G3*.

```
Lemma demo_splits : ∀ n m,
  n > 0 ∧ n < m ∧ m < n+10 ∧ m ≠ 3.
Proof.
  intros. splits.
Abort.
```

20.2.2 The Tactic *branch*

The tactic *branch* `k` can be used to prove a n-ary disjunction. For example, if the goal takes the form *G1* ∨ *G2* ∨ *G3*, the tactic *branch* 2 leaves only *G2* as subgoal. The following example illustrates the behavior of the *branch* tactic.

```
Lemma demo_branch : ∀ n m,
  n < m ∨ n = m ∨ m < n.
Proof.
  intros.
  destruct (lt_eq_lt_dec n m) as [[H1|H2]|H3].
  - branch 1. apply H1.
  - branch 2. apply H2.
  - branch 3. apply H3.
Qed.
```

End NARYEXAMPLES.

20.3 Tactics for Working with Equality

One of the major weakness of Coq compared with other interactive proof assistants is its relatively poor support for reasoning with equalities. The tactics described next aims at simplifying pieces of proof scripts manipulating equalities.

This section presents the following tactics:

- *asserts_rewrite* for introducing an equality to rewrite with,
- *cuts_rewrite*, which is similar except that its subgoals are swapped,
- *substs* for improving the `subst` tactic,
- *fequals* for improving the `f_equal` tactic,
- *applys_eq* for proving P x y using an hypothesis P x z, automatically producing an equality y = z as subgoal.

Module EQUALITYEXAMPLES.

20.3.1 The Tactics *asserts_rewrite* and *cuts_rewrite*

The tactic *asserts_rewrite* (*E1* = *E2*) replaces *E1* with *E2* in the goal, and produces the goal *E1* = *E2*.

```
Theorem mult_0_plus : ∀ n m : nat,
  (0 + n) × m = n × m.
Proof.
  dup.
  intros n m.
  assert (H: 0 + n = n). reflexivity. rewrite → H.
  reflexivity.

  intros n m.
```

```
  asserts_rewrite (0 + n = n).
    reflexivity.    reflexivity. Qed.
```

The tactic *cuts_rewrite* (*E1* = *E2*) is like *asserts_rewrite* (*E1* = *E2*), except that the equality *E1* = *E2* appears as first subgoal.

```
Theorem mult_0_plus' : ∀ n m : nat,
  (0 + n) × m = n × m.
Proof.
  intros n m.
  cuts_rewrite (0 + n = n).
    reflexivity.    reflexivity. Qed.
```

More generally, the tactics *asserts_rewrite* and *cuts_rewrite* can be provided a lemma as argument. For example, one can write *asserts_rewrite* (∀ a *b*, a*(S *b*) = a×*b*+a). This formulation is useful when a and *b* are big terms, since there is no need to repeat their statements.

```
Theorem mult_0_plus'' : ∀ u v w x y z: nat,
  (u + v) × (S (w × x + y)) = z.
Proof.
  intros. asserts_rewrite (∀ a b, a*(S b) = a×b+a).
Abort.
```

20.3.2 The Tactic *substs*

The tactic *substs* is similar to `subst` except that it does not fail when the goal contains "circular equalities", such as x = f x.

```
Lemma demo_substs : ∀ x y (f:nat→nat),
  x = f x →
  y = x →
  y = f x.
Proof.
  intros. substs.   assumption.
Qed.
```

20.3.3 The Tactic *fequals*

The tactic *fequals* is similar to `f_equal` except that it directly discharges all the trivial subgoals produced. Moreover, the tactic *fequals* features an enhanced treatment of equalities between tuples.

```
Lemma demo_fequals : ∀ (a b c d e : nat) (f : nat→nat→nat→nat→nat),
  a = 1 →
  b = e →
  e = 2 →
```

f a b c d = *f* 1 2 *c* 4.
Proof.
  intros. *fequals*.
Abort.

20.3.4 The Tactic *applys_eq*

The tactic *applys_eq* is a variant of eapply that introduces equalities for subterms that do not unify. For example, assume the goal is the proposition P x y and assume we have the assumption H asserting that P x z holds. We know that we can prove y to be equal to z. So, we could call the tactic *assert_rewrite* (y = z) and change the goal to P x z, but this would require copy-pasting the values of y and z. With the tactic *applys_eq*, we can call *applys_eq* H 1, which proves the goal and leaves only the subgoal y = z. The value 1 given as argument to *applys_eq* indicates that we want an equality to be introduced for the first argument of P x y counting from the right. The three following examples illustrate the behavior of a call to *applys_eq* H 1, a call to *applys_eq* H 2, and a call to *applys_eq* H 1 2.

Axiom *big_expression_using* : **nat**→**nat**.

Lemma demo_applys_eq_1 : ∀ (*P*:**nat**→**nat**→Prop) *x y z*,
  P x (*big_expression_using z*) →
  P x (*big_expression_using y*).
Proof.
  introv H. dup.

  assert (*Eq*: *big_expression_using y* = *big_expression_using z*).
    admit.   rewrite *Eq*. apply *H*.

  applys_eq H 1.
    admit. Abort.

If the mismatch was on the first argument of P instead of the second, we would have written *applys_eq* H 2. Recall that the occurences are counted from the right.

Lemma demo_applys_eq_2 : ∀ (*P*:**nat**→**nat**→Prop) *x y z*,
  P (*big_expression_using z*) *x* →
  P (*big_expression_using y*) *x*.
Proof.
  introv H. applys_eq H 2.
Abort.

When we have a mismatch on two arguments, we want to produce two equalities. To achieve this, we may call *applys_eq* H 1 2. More generally, the tactic *applys_eq* expects a lemma and a sequence of natural numbers as arguments.

Lemma demo_applys_eq_3 : ∀ (*P*:**nat**→**nat**→Prop) *x1 x2 y1 y2*,
  P (*big_expression_using x2*) (*big_expression_using y2*) →
  P (*big_expression_using x1*) (*big_expression_using y1*).

```
Proof.
  introv H. applys_eq H 1 2.
Abort.
End EqualityExamples.
```

20.4 Some Convenient Shorthands

This section of the tutorial introduces a few tactics that help make proof scripts shorter and more readable:

- *unfolds* (without argument) for unfolding the head definition,
- false for replacing the goal with **False**,
- *gen* as a shorthand for `dependent generalize`,
- *admits* for naming an addmited fact,
- *admit_rewrite* for rewriting using an admitted equality,
- *admit_goal* to set up a proof by induction by skipping the justification that some order decreases,
- *sort* for re-ordering the proof context by moving moving all propositions at the bottom.

20.4.1 The Tactic *unfolds*

```
Module UnfoldsExample.
  Import Hoare.
```

The tactic *unfolds* (without any argument) unfolds the head constant of the goal. This tactic saves the need to name the constant explicitly.

```
Lemma bexp_eval_true : ∀ b st,
  beval st b = true →
  (bassn b) st.
Proof.
  intros b st Hbe. dup.
  unfold bassn. assumption.
  unfolds. assumption.
Qed.
```

Remark: contrary to the tactic `hnf`, which may unfold several constants, *unfolds* performs only a single step of unfolding.

Remark: the tactic *unfolds* `in` H can be used to unfold the head definition of the hypothesis H.

End UNFOLDSEXAMPLE.

20.4.2 The Tactics false and *tryfalse*

The tactic false can be used to replace any goal with **False**. In short, it is a shorthand for *exfalso*. Moreover, false proves the goal if it contains an absurd assumption, such as **False** or $0 =$ S n, or if it contains contradictory assumptions, such as x $=$ true and x $=$ false.

Lemma demo_false : ∀ n,
  S n = 1 →
  n = 0.
Proof.
  intros. destruct n. reflexivity. *false*.
Qed.

The tactic false can be given an argument: false H replace the goals with **False** and then applies H.

Lemma demo_false_arg :
  (∀ n, n < 0 → **False**) →
  3 < 0 →
  4 < 0.
Proof.
  intros H L. *false* H. apply L.
Qed.

The tactic *tryfalse* is a shorthand for `try solve` [false]: it tries to find a contradiction in the goal. The tactic *tryfalse* is generally called after a case analysis.

Lemma demo_tryfalse : ∀ n,
  S n = 1 →
  n = 0.
Proof.
  intros. destruct n; *tryfalse*. reflexivity.
Qed.

20.4.3 The Tactic *gen*

The tactic *gen* is a shortand for `generalize dependent` that accepts several arguments at once. An invokation of this tactic takes the form *gen* x y z.

Module GENEXAMPLE.
  Import *Stlc*.
  Import *STLC*.

Lemma substitution_preserves_typing : ∀ *Gamma x U v t S*,
  has_type (update *Gamma x U*) *t S* →
  has_type empty *v U* →
  has_type *Gamma* ([*x*:=*v*] *t*) *S*.
Proof.
  dup.

  intros *Gamma x U v t S Htypt Htypv*.
  generalize dependent *S*. generalize dependent *Gamma*.
  induction *t*; intros; simpl.
  admit. admit. admit. admit. admit. admit.

  introv Htypt Htypv. gen S Gamma.
  induction *t*; intros; simpl.
  admit. admit. admit. admit. admit. admit.
Abort.

End GenExample.

20.4.4 The Tactics *admits*, *admit_rewrite* and *admit_goal*

Temporarily admitting a given subgoal is very useful when constructing proofs. Several tactics are provided as useful wrappers around the builtin *admit* tactic.

Module SkipExample.
  Import *Stlc*.
  Import *STLC*.

The tactic *admits H*: P adds the hypothesis *H*: P to the context, without checking whether the proposition P is true. It is useful for exploiting a fact and postponing its proof. Note: *admits H*: P is simply a shorthand for assert (*H*:P). *admit*.

Theorem demo_admits : **True**.
Proof.
  admits H: (∀ *n m* : **nat**, (0 + *n*) × *m* = *n* × *m*).
Abort.

The tactic *admit_rewrite* (*E1* = *E2*) replaces *E1* with *E2* in the goal, without checking that *E1* is actually equal to *E2*.

Theorem mult_plus_0 : ∀ *n m* : **nat**,
  (*n* + 0) × *m* = *n* × *m*.
Proof.
  dup 3.

  intros *n m*.
  assert (*H*: *n* + 0 = *n*). *admit*. rewrite → *H*. clear *H*.
  reflexivity.

  intros *n m*.

```
  admit_rewrite (n + 0 = n).
  reflexivity.

  intros n m.
  admit_rewrite (∀ a, a + 0 = a).
  reflexivity.
Admitted.
```

The tactic *admit_goal* adds the current goal as hypothesis. This cheat is useful to set up the structure of a proof by induction without having to worry about the induction hypothesis being applied only to smaller arguments. Using *skip_goal*, one can construct a proof in two steps: first, check that the main arguments go through without waisting time on fixing the details of the induction hypotheses; then, focus on fixing the invokations of the induction hypothesis.

```
Theorem ceval_deterministic: ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  admit_goal.
  introv E1 E2. gen st2.
  (induction E1); introv E2; inverts E2 as.
  - reflexivity.
  -
    subst n.
    reflexivity.
  -
    intros st3 Red1 Red2.
    assert (st' = st3) as EQ1.
    {

      eapply IH. eapply E1_1. eapply Red1. }
    subst st3.
   eapply IH. eapply E1_2. eapply Red2.
Abort.

End SkipExample.
```

20.4.5 The Tactic *sort*

```
Module SortExamples.
  Import Imp.
```

The tactic *sort* reorganizes the proof context by placing all the variables at the top and all the hypotheses at the bottom, thereby making the proof context more readable.

```
Theorem ceval_deterministic: ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  intros c st st1 st2 E1 E2.
  generalize dependent st2.
  (induction E1); intros st2 E2; inverts E2.
  admit. admit.   sort. Abort.
End SortExamples.
```

20.5 Tactics for Advanced Lemma Instantiation

This last section describes a mechanism for instantiating a lemma by providing some of its arguments and leaving other implicit. Variables whose instantiation is not provided are turned into existentential variables, and facts whose instantiation is not provided are turned into subgoals.

Remark: this instantion mechanism goes far beyond the abilities of the "Implicit Arguments" mechanism. The point of the instantiation mechanism described in this section is that you will no longer need to spend time figuring out how many underscore symbols you need to write.

In this section, we'll use a useful feature of Coq for decomposing conjunctions and existentials. In short, a tactic like `intros` or `destruct` can be provided with a pattern (*H1* & *H2* & *H3* & *H4* & *H5*), which is a shorthand for [*H1* [*H2* [*H3* [*H4* *H5*]]]]. For example, `destruct` (*H* _ _ _ *Htypt*) `as` [T [*Hctx Hsub*]]. can be rewritten in the form `destruct` (*H* _ _ _ *Htypt*) `as` (T & *Hctx* & *Hsub*).

20.5.1 Working of *lets*

When we have a lemma (or an assumption) that we want to exploit, we often need to explicitly provide arguments to this lemma, writing something like: `destruct` (*typing_inversion_var* _ _ _ *Htypt*) `as` (T & *Hctx* & *Hsub*). The need to write several times the "underscore" symbol is tedious. Not only we need to figure out how many of them to write down, but it also makes the proof scripts look prettly ugly. With the tactic *lets*, one can simply write: *lets* (T & *Hctx* & *Hsub*): *typing_inversion_var* *Htypt*.

In short, this tactic *lets* allows to specialize a lemma on a bunch of variables and hypotheses. The syntax is *lets* I: *E0* *E1* .. *EN*, for building an hypothesis named I by applying the fact *E0* to the arguments *E1* to *EN*. Not all the arguments need to be provided, however the arguments that are provided need to be provided in the correct order. The tactic relies on a first-match algorithm based on types in order to figure out how the to instantiate the lemma with the arguments provided.

```
Module ExamplesLets.
```

```
Import Sub.
```

```
Import Sub.

Axiom typing_inversion_var : ∀ (G:context) (x:string) (T:ty),
  has_type G (var x) T →
  ∃ S, G x = Some S ∧ subtype S T.
```

First, assume we have an assumption H with the type of the form **has_type** G (var x) T. We can obtain the conclusion of the lemma *typing_inversion_var* by invoking the tactics *lets* K: *typing_inversion_var* H, as shown next.

```
Lemma demo_lets_1 : ∀ (G:context) (x:string) (T:ty),
  has_type G (var x) T →
  True.
Proof.
  intros G x T H. dup.

  lets K: typing_inversion_var H.
  destruct K as (S & Eq & Sub).
  admit.

  lets (S & Eq & Sub): typing_inversion_var H.
  admit.
Abort.
```

Assume now that we know the values of G, x and T and we want to obtain S, and have **has_type** G (var x) T be produced as a subgoal. To indicate that we want all the remaining arguments of *typing_inversion_var* to be produced as subgoals, we use a triple-underscore symbol ___. (We'll later introduce a shorthand tactic called *forwards* to avoid writing triple underscores.)

```
Lemma demo_lets_2 : ∀ (G:context) (x:string) (T:ty), True.
Proof.
  intros G x T.
  lets (S & Eq & Sub): typing_inversion_var G x T ___.
Abort.
```

Usually, there is only one context G and one type T that are going to be suitable for proving **has_type** G (var x) T, so we don't really need to bother giving G and T explicitly. It suffices to call *lets* (S & *Eq* & Sub): *typing_inversion_var* x. The variables G and T are then instantiated using existential variables.

```
Lemma demo_lets_3 : ∀ (x:string), True.
Proof.
  intros x.
  lets (S & Eq & Sub): typing_inversion_var x ___.
Abort.
```

We may go even further by not giving any argument to instantiate *typing_inversion_var*. In this case, three unification variables are introduced.

```
Lemma demo_lets_4 : True.
Proof.
  lets (S & Eq & Sub): typing_inversion_var ___.
Abort.
```

Note: if we provide *lets* with only the name of the lemma as argument, it simply adds this lemma in the proof context, without trying to instantiate any of its arguments.

```
Lemma demo_lets_5 : True.
Proof.
  lets H: typing_inversion_var.
Abort.
```

A last useful feature of *lets* is the double-underscore symbol, which allows skipping an argument when several arguments have the same type. In the following example, our assumption quantifies over two variables n and m, both of type **nat**. We would like m to be instantiated as the value 3, but without specifying a value for n. This can be achieved by writting *lets K*: *H* __ 3.

```
Lemma demo_lets_underscore :
  (∀ n m, n ≤ m → n < m+1) →
  True.
Proof.
  intros H.
  lets K: H 3.     clear K.
  lets K: H __ 3.     clear K.
Abort.
```

Note: one can write *lets*: *E0 E1 E2* in place of *lets H*: *E0 E1 E2*. In this case, the name *H* is chosen arbitrarily.

Note: the tactics *lets* accepts up to five arguments. Another syntax is available for providing more than five arguments. It consists in using a list introduced with the special symbol », for example *lets H*: (» *E0 E1 E2 E3 E4 E5 E6 E7 E8 E9* 10).

End ExamplesLets.

20.5.2 Working of *applys*, *forwards* and *specializes*

The tactics *applys*, *forwards* and *specializes* are shorthand that may be used in place of *lets* to perform specific tasks.

- *forwards* is a shorthand for instantiating all the arguments

of a lemma. More precisely, *forwards H*: *E0 E1 E2 E3* is the same as *lets H*: *E0 E1 E2 E3* ___, where the triple-underscore has the same meaning as explained earlier on.

- *applys* allows building a lemma using the advanced instantion mode of *lets*, and then apply that lemma right away. So, *applys E0 E1 E2 E3* is the same as *lets H*: *E0 E1 E2 E3* followed with `eapply` *H* and then `clear` *H*.

- *specializes* is a shorthand for instantiating in-place

an assumption from the context with particular arguments. More precisely, *specializes H E0 E1* is the same as *lets H'*: *H E0 E1* followed with `clear` *H* and `rename` *H' into H*.

Examples of use of *applys* appear further on. Several examples of use of *forwards* can be found in the tutorial chapter UseAuto.

20.5.3 Example of Instantiations

```
Module ExamplesInstantiations.
  Import Sub.
```

The following proof shows several examples where *lets* is used instead of `destruct`, as well as examples where *applys* is used instead of `apply`. The proof also contains some holes that you need to fill in as an exercise.

```
Lemma substitution_preserves_typing : ∀ Gamma x U v t S,
  has_type (update Gamma x U) t S →
  has_type empty v U →
  has_type Gamma ([x:=v]t) S.
Proof with eauto.
  intros Gamma x U v t S Htypt Htypv.
  generalize dependent S. generalize dependent Gamma.
  (induction t); intros; simpl.
  -
    rename s into y.
   lets (T&Hctx&Hsub): typing_inversion_var Htypt.
    unfold update, t_update in Hctx.
    destruct (eqb_stringP x y)...
    +
      subst.
      inversion Hctx; subst. clear Hctx.
      apply context_invariance with empty...
      intros x Hcontra.
        lets [T' HT']: free_in_context S (@empty ty) Hcontra...
        inversion HT'.
  -

    admit.
```

```
-
  rename s into y. rename t into T1.
 lets (T2&Hsub&Htypt2): typing_inversion_abs Htypt.
 applys T_Sub (Arrow T1 T2)...
   apply T_Abs...
  destruct (eqb_stringP x y).
  +
    eapply context_invariance...
    subst.
    intros x Hafi. unfold update, t_update.
    destruct (eqb_stringP y x)...
  +
    apply IHt. eapply context_invariance...
    intros z Hafi. unfold update, t_update.
    destruct (eqb_stringP y z)...
    subst. rewrite false_eqb_string...
-
  lets: typing_inversion_true Htypt...
-
  lets: typing_inversion_false Htypt...
-
  lets (Htyp1&Htyp2&Htyp3): typing_inversion_if Htypt...
-


  lets: typing_inversion_unit Htypt...
Admitted.
End ExamplesInstantiations.
```

20.6 Summary

In this chapter we have presented a number of tactics that help make proof script more concise and more robust on change.

- *introv* and *inverts* improve naming and inversions.
- false and *tryfalse* help discarding absurd goals.
- *unfolds* automatically calls `unfold` on the head definition.
- *gen* helps setting up goals for induction.
- *cases* and *cases_if* help with case analysis.

- *splits* and *branch*, to deal with n-ary constructs.
- *asserts_rewrite*, *cuts_rewrite*, *substs* and *fequals* help working with equalities.
- *lets*, *forwards*, *specializes* and *applys* provide means of very conveniently instantiating lemmas.
- *applys_eq* can save the need to perform manual rewriting steps before being able to apply lemma.
- *admits*, *admit_rewrite* and *admit_goal* give the flexibility to choose which subgoals to try and discharge first.

Making use of these tactics can boost one's productivity in Coq proofs.

If you are interested in using *LibTactics.v* in your own developments, make sure you get the lastest version from: http://www.chargueraud.org/softs/tlc/.

Chapter 21

UseAuto: Theory and Practice of Automation in Coq Proofs

In a machine-checked proof, every single detail has to be justified. This can result in huge proof scripts. Fortunately, Coq comes with a proof-search mechanism and with several decision procedures that enable the system to automatically synthesize simple pieces of proof. Automation is very powerful when set up appropriately. The purpose of this chapter is to explain the basics of how automation works in Coq.

The chapter is organized in two parts. The first part focuses on a general mechanism called "proof search." In short, proof search consists in naively trying to apply lemmas and assumptions in all possible ways. The second part describes "decision procedures", which are tactics that are very good at solving proof obligations that fall in some particular fragments of the logic of Coq.

Many of the examples used in this chapter consist of small lemmas that have been made up to illustrate particular aspects of automation. These examples are completely independent from the rest of the Software Foundations course. This chapter also contains some bigger examples which are used to explain how to use automation in realistic proofs. These examples are taken from other chapters of the course (mostly from STLC), and the proofs that we present make use of the tactics from the library *LibTactics.v*, which is presented in the chapter UseTactics.

```
From Coq Require Import Arith.Arith.

From PLF Require Import Maps.
From PLF Require Import Smallstep.
From PLF Require Import Stlc.
From PLF Require Import LibTactics.

From PLF Require Imp.

From Coq Require Import Lists.List.
Import ListNotations.
```

21.1 Basic Features of Proof Search

The idea of proof search is to replace a sequence of tactics applying lemmas and assumptions with a call to a single tactic, for example `auto`. This form of proof automation saves a lot of effort. It typically leads to much shorter proof scripts, and to scripts that are typically more robust to change. If one makes a little change to a definition, a proof that exploits automation probably won't need to be modified at all. Of course, using too much automation is a bad idea. When a proof script no longer records the main arguments of a proof, it becomes difficult to fix it when it gets broken after a change in a definition. Overall, a reasonable use of automation is generally a big win, as it saves a lot of time both in building proof scripts and in subsequently maintaining those proof scripts.

21.1.1 Strength of Proof Search

We are going to study four proof-search tactics: `auto`, `eauto`, *iauto* and *jauto*. The tactics `auto` and `eauto` are builtin in Coq. The tactic *iauto* is a shorthand for the builtin tactic `try solve [intuition eauto]`. The tactic *jauto* is defined in the library LibTactics, and simply performs some preprocessing of the goal before calling `eauto`. The goal of this chapter is to explain the general principles of proof search and to give rule of thumbs for guessing which of the four tactics mentioned above is best suited for solving a given goal.

Proof search is a compromise between efficiency and expressiveness, that is, a tradeoff between how complex goals the tactic can solve and how much time the tactic requires for terminating. The tactic `auto` builds proofs only by using the basic tactics `reflexivity`, `assumption`, and `apply`. The tactic `eauto` can also exploit `eapply`. The tactic *jauto* extends `eauto` by being able to open conjunctions and existentials that occur in the context. The tactic *iauto* is able to deal with conjunctions, disjunctions, and negation in a quite clever way; however it is not able to open existentials from the context. Also, *iauto* usually becomes very slow when the goal involves several disjunctions.

Note that proof search tactics never perform any rewriting step (tactics `rewrite`, `subst`), nor any case analysis on an arbitrary data structure or property (tactics `destruct` and `inversion`), nor any proof by induction (tactic `induction`). So, proof search is really intended to automate the final steps from the various branches of a proof. It is not able to discover the overall structure of a proof.

21.1.2 Basics

The tactic `auto` is able to solve a goal that can be proved using a sequence of `intros`, `apply`, `assumption`, and `reflexivity`. Two examples follow. The first one shows the ability for `auto` to call `reflexivity` at any time. In fact, calling `reflexivity` is always the first thing that `auto` tries to do.

```
Lemma solving_by_reflexivity :
  2 + 3 = 5.
Proof. auto. Qed.
```

The second example illustrates a proof where a sequence of two calls to `apply` are needed. The goal is to prove that if Q n implies P n for any n and if Q n holds for any n, then P 2 holds.

```
Lemma solving_by_apply : ∀ (P Q : nat→Prop),
  (∀ n, Q n → P n) →
  (∀ n, Q n) →
  P 2.
Proof. auto. Qed.
```

If we are interested to see which proof `auto` came up with, one possibility is to look at the generated proof-term, using the command:

`Print solving_by_apply.`

The proof term is:

`fun (P Q : nat → Prop) (H : ∀ n : nat, Q n → P n) (H0 : ∀ n : nat, Q n) ⇒ H 2 (H0 2)`

This essentially means that `auto` applied the hypothesis *H* (the first one), and then applied the hypothesis *H0* (the second one).

The tactic `auto` can invoke `apply` but not `eapply`. So, `auto` cannot exploit lemmas whose instantiation cannot be directly deduced from the proof goal. To exploit such lemmas, one needs to invoke the tactic `eauto`, which is able to call `eapply`.

In the following example, the first hypothesis asserts that P n is true when Q m is true for some m, and the goal is to prove that Q 1 implies P 2. This implication follows directly from the hypothesis by instantiating m as the value 1. The following proof script shows that `eauto` successfully solves the goal, whereas `auto` is not able to do so.

```
Lemma solving_by_eapply : ∀ (P Q : nat→Prop),
  (∀ n m, Q m → P n) →
  Q 1 →
  P 2.
Proof. auto. eauto. Qed.
```

21.1.3 Conjunctions

So far, we've seen that `eauto` is stronger than `auto` in the sense that it can deal with `eapply`. In the same way, we are going to see how *jauto* and *iauto* are stronger than `auto` and `eauto` in the sense that they provide better support for conjunctions.

The tactics `auto` and `eauto` can prove a goal of the form $F \wedge F'$, where F and F' are two propositions, as soon as both F and F' can be proved in the current context. An example follows.

```
Lemma solving_conj_goal : ∀ (P : nat→Prop) (F : Prop),
  (∀ n, P n) →
  F →
  F ∧ P 2.
```

```
Proof. auto. Qed.
```

However, when an assumption is a conjunction, `auto` and `eauto` are not able to exploit this conjunction. It can be quite surprising at first that `eauto` can prove very complex goals but that it fails to prove that $F \wedge F'$ implies F. The tactics *iauto* and *jauto* are able to decompose conjunctions from the context. Here is an example.

```
Lemma solving_conj_hyp : ∀ (F F' : Prop),
  F ∧ F' →
  F.
Proof. auto. eauto. jauto. Qed.
```

The tactic *jauto* is implemented by first calling a pre-processing tactic called *jauto_set*, and then calling `eauto`. So, to understand how *jauto* works, one can directly call the tactic *jauto_set*.

```
Lemma solving_conj_hyp' : ∀ (F F' : Prop),
  F ∧ F' →
  F.
Proof. intros. jauto_set. eauto. Qed.
```

Next is a more involved goal that can be solved by *iauto* and *jauto*.

```
Lemma solving_conj_more : ∀ (P Q R : nat→Prop) (F : Prop),
  (F ∧ (∀ n m, (Q m ∧ R n) → P n)) →
  (F → R 2) →
  Q 1 →
  P 2 ∧ F.
Proof. jauto. Qed.
```

The strategy of *iauto* and *jauto* is to run a global analysis of the top-level conjunctions, and then call `eauto`. For this reason, those tactics are not good at dealing with conjunctions that occur as the conclusion of some universally quantified hypothesis. The following example illustrates a general weakness of Coq proof search mechanisms.

```
Lemma solving_conj_hyp_forall : ∀ (P Q : nat→Prop),
  (∀ n, P n ∧ Q n) →
  P 2.
Proof.
  auto. eauto. iauto. jauto.
  intros. destruct (H 2). auto.
Qed.
```

This situation is slightly disappointing, since automation is able to prove the following goal, which is very similar. The only difference is that the universal quantification has been distributed over the conjunction.

```
Lemma solved_by_jauto : ∀ (P Q : nat→Prop) (F : Prop),
  (∀ n, P n) ∧ (∀ n, Q n) →
  P 2.
```

Proof. *jauto*. Qed.

21.1.4 Disjunctions

The tactics `auto` and `eauto` can handle disjunctions that occur in the goal.

```
Lemma solving_disj_goal : ∀ (F F' : Prop),
  F →
  F ∨ F'.
Proof. auto. Qed.
```

However, only *iauto* is able to automate reasoning on the disjunctions that appear in the context. For example, *iauto* can prove that $F \lor F'$ entails $F' \lor F$.

```
Lemma solving_disj_hyp : ∀ (F F' : Prop),
  F ∨ F' →
  F' ∨ F.
Proof. auto. eauto. jauto. iauto. Qed.
```

More generally, *iauto* can deal with complex combinations of conjunctions, disjunctions, and negations. Here is an example.

```
Lemma solving_tauto : ∀ (F1 F2 F3 : Prop),
  ((¬F1 ∧ F3) ∨ (F2 ∧ ¬F3)) →
  (F2 → F1) →
  (F2 → F3) →
  ¬F2.
Proof. iauto. Qed.
```

However, the ability of *iauto* to automatically perform a case analysis on disjunctions comes with a downside: *iauto* may be very slow. If the context involves several hypotheses with disjunctions, *iauto* typically generates an exponential number of subgoals on which `eauto` is called. One major advantage of *jauto* compared with *iauto* is that it never spends time performing this kind of case analyses.

21.1.5 Existentials

The tactics `eauto`, *iauto*, and *jauto* can prove goals whose conclusion is an existential. For example, if the goal is ∃ x, f x, the tactic `eauto` introduces an existential variable, say ?25, in place of x. The remaining goal is f ?25, and `eauto` tries to solve this goal, allowing itself to instantiate ?25 with any appropriate value. For example, if an assumption f 2 is available, then the variable ?25 gets instantiated with 2 and the goal is solved, as shown below.

```
Lemma solving_exists_goal : ∀ (f : nat→Prop),
  f 2 →
  ∃ x, f x.
Proof.
  auto.   eauto. Qed.
```

A major strength of *jauto* over the other proof search tactics is that it is able to exploit the existentially-quantified hypotheses, i.e., those of the form ∃ x, P.

```
Lemma solving_exists_hyp : ∀ (f g : nat→Prop),
  (∀ x, f x → g x) →
  (∃ a, f a) →
  (∃ a, g a).
Proof.
  auto. eauto. iauto.   jauto. Qed.
```

21.1.6 Negation

The tactics `auto` and `eauto` suffer from some limitations with respect to the manipulation of negations, mostly related to the fact that negation, written ¬ P, is defined as P → **False** but that the unfolding of this definition is not performed automatically. Consider the following example.

```
Lemma negation_study_1 : ∀ (P : nat→Prop),
  P 0 →
  (∀ x, ¬ P x) →
  False.
Proof.
  intros P H0 HX.
  eauto.   unfold not in *. eauto.
Qed.
```

For this reason, the tactics *iauto* and *jauto* systematically invoke `unfold not in *` as part of their pre-processing. So, they are able to solve the previous goal right away.

```
Lemma negation_study_2 : ∀ (P : nat→Prop),
  P 0 →
  (∀ x, ¬ P x) →
  False.
Proof. jauto. Qed.
```

We will come back later on to the behavior of proof search with respect to the unfolding of definitions.

21.1.7 Equalities

Coq's proof-search feature is not good at exploiting equalities. It can do very basic operations, like exploiting reflexivity and symmetry, but that's about it. Here is a simple example that `auto` can solve, by first calling `symmetry` and then applying the hypothesis.

```
Lemma equality_by_auto : ∀ (f g : nat→Prop),
  (∀ x, f x = g x) →
  g 2 = f 2.
```

```
Proof. auto. Qed.
```

To automate more advanced reasoning on equalities, one should rather try to use the tactic `congruence`, which is presented at the end of this chapter in the "Decision Procedures" section.

21.2 How Proof Search Works

21.2.1 Search Depth

The tactic `auto` works as follows. It first tries to call `reflexivity` and `assumption`. If one of these calls solves the goal, the job is done. Otherwise `auto` tries to apply the most recently introduced assumption that can be applied to the goal without producing and error. This application produces subgoals. There are two possible cases. If the sugboals produced can be solved by a recursive call to `auto`, then the job is done. Otherwise, if this application produces at least one subgoal that `auto` cannot solve, then `auto` starts over by trying to apply the second most recently introduced assumption. It continues in a similar fashion until it finds a proof or until no assumption remains to be tried.

It is very important to have a clear idea of the backtracking process involved in the execution of the `auto` tactic; otherwise its behavior can be quite puzzling. For example, `auto` is not able to solve the following triviality.

```
Lemma search_depth_0 :
  True ∧ True ∧ True ∧ True ∧ True ∧ True.
Proof.
  auto.
Abort.
```

The reason `auto` fails to solve the goal is because there are too many conjunctions. If there had been only five of them, `auto` would have successfully solved the proof, but six is too many. The tactic `auto` limits the number of lemmas and hypotheses that can be applied in a proof, so as to ensure that the proof search eventually terminates. By default, the maximal number of steps is five. One can specify a different bound, writing for example `auto` 6 to search for a proof involving at most six steps. For example, `auto` 6 would solve the previous lemma. (Similarly, one can invoke `eauto` 6 or `intuition eauto` 6.) The argument `n` of `auto n` is called the "search depth." The tactic `auto` is simply defined as a shorthand for `auto` 5.

The behavior of `auto n` can be summarized as follows. It first tries to solve the goal using `reflexivity` and `assumption`. If this fails, it tries to apply a hypothesis (or a lemma that has been registered in the hint database), and this application produces a number of sugoals. The tactic `auto` (`n`-1) is then called on each of those subgoals. If all the subgoals are solved, the job is completed, otherwise `auto n` tries to apply a different hypothesis.

During the process, `auto n` calls `auto` (`n`-1), which in turn might call `auto` (`n`-2), and so on. The tactic `auto` 0 only tries `reflexivity` and `assumption`, and does not try to apply any lemma. Overall, this means that when the maximal number of steps allowed has been exceeded, the auto tactic stops searching and backtracks to try and investigate other paths.

The following lemma admits a unique proof that involves exactly three steps. So, auto n proves this goal iff n is greater than three.

```
Lemma search_depth_1 : ∀ (P : nat→Prop),
  P 0 →
  (P 0 → P 1) →
  (P 1 → P 2) →
  (P 2).
Proof.
  auto 0.   auto 1.   auto 2.   auto 3. Qed.
```

We can generalize the example by introducing an assumption asserting that P k is derivable from P (k-1) for all k, and keep the assumption P 0. The tactic auto, which is the same as auto 5, is able to derive P k for all values of k less than 5. For example, it can prove P 4.

```
Lemma search_depth_3 : ∀ (P : nat→Prop),
  (P 0) →
  (∀ k, P (k-1) → P k) →
  (P 4).
Proof. auto. Qed.
```

However, to prove P 5, one needs to call at least auto 6.

```
Lemma search_depth_4 : ∀ (P : nat→Prop),
  (P 0) →
  (∀ k, P (k-1) → P k) →
  (P 5).
Proof. auto. auto 6. Qed.
```

Because auto looks for proofs at a limited depth, there are cases where auto can prove a goal F and can prove a goal F' but cannot prove $F \wedge F$'. In the following example, auto can prove P 4 but it is not able to prove P 4 ∧ P 4, because the splitting of the conjunction consumes one proof step. To prove the conjunction, one needs to increase the search depth, using at least auto 6.

```
Lemma search_depth_5 : ∀ (P : nat→Prop),
  (P 0) →
  (∀ k, P (k-1) → P k) →
  (P 4 ∧ P 4).
Proof. auto. auto 6. Qed.
```

21.2.2 Backtracking

In the previous section, we have considered proofs where at each step there was a unique assumption that auto could apply. In general, auto can have several choices at every step.

The strategy of `auto` consists of trying all of the possibilities (using a depth-first search exploration).

To illustrate how automation works, we are going to extend the previous example with an additional assumption asserting that P k is also derivable from P (k+1). Adding this hypothesis offers a new possibility that `auto` could consider at every step.

There exists a special command that one can use for tracing all the steps that proof-search considers. To view such a trace, one should write `debug eauto`. (For some reason, the command `debug auto` does not exist, so we have to use the command `debug eauto` instead.)

```
Lemma working_of_auto_1 : ∀ (P : nat→Prop),
  (P 0) →
  (∀ k, P (k-1) → P k) →
  (∀ k, P (k+1) → P k) →
  (P 2).
Proof. intros P H1 H2 H3. eauto. Qed.
```

The output message produced by `debug eauto` is as follows.

1 depth=5 1.1 depth=4 simple apply H2 1.1.1 depth=3 simple apply H2 1.1.1.1 depth=3 exact H1

The depth indicates the value of `n` with which `eauto n` is called. The tactics shown in the message indicate that the first thing that `eauto` has tried to do is to apply *H2*. The effect of applying *H2* is to replace the goal P 2 with the goal P 1. Then, again, *H2* has been applied, changing the goal P 1 into P 0. At that point, the goal was exactly the hypothesis *H1*.

It seems that `eauto` was quite lucky there, as it never even tried to use the hypothesis *H3* at any time. The reason is that `auto` always tried to use the *H2* first. So, let's permute the hypotheses *H2* and *H3* and see what happens.

```
Lemma working_of_auto_2 : ∀ (P : nat→Prop),
  (P 0) →
  (∀ k, P (k+1) → P k) →
  (∀ k, P (k-1) → P k) →
  (P 2).
Proof. intros P H1 H3 H2. eauto. Qed.
```

This time, the output message suggests that the proof search investigates many possibilities. If we print the proof term:

`Print working_of_auto_2.`

we observe that the proof term refers to *H3*. Thus the proof is not the simplest one, since only *H2* and *H1* are needed.

In turns out that the proof goes through the proof obligation P 3, even though it is not required to do so. The following tree drawing describes all the goals that `eauto` has been going through.

|5||4||3||2||1||0| – below, tabulation indicates the depth

P 2

- > P 3

- > *P* 4
  - > *P* 5
    - > *P* 6
      - > *P* 7
      - > *P* 5
    - > *P* 4
      - > *P* 5
      - > *P* 3
  - -> *P* 3
    - > *P* 4
      - > *P* 5
      - > *P* 3
    - > *P* 2
      - > *P* 3
      - > *P* 1
- > *P* 2
  - > *P* 3
    - > *P* 4
      - > *P* 5
      - > *P* 3
    - > *P* 2
      - > *P* 3
      - > *P* 1
  - > *P* 1
    - > *P* 2
      - > *P* 3
      - > *P* 1
    - > *P* 0
      - > !! Done !!

The first few lines read as follows. To prove *P* 2, `eauto` 5 has first tried to apply *H3*, producing the subgoal *P* 3. To solve it, `eauto` 4 has tried again to apply *H3*, producing the goal *P* 4. Similarly, the search goes through *P* 5, *P* 6 and *P* 7. When reaching *P* 7, the tactic `eauto` 0 is called but as it is not allowed to try and apply any lemma, it fails. So, we come back to the goal *P* 6, and try this time to apply hypothesis *H2*, producing the subgoal *P* 5. Here again, `eauto` 0 fails to solve this goal.

The process goes on and on, until backtracking to *P* 3 and trying to apply *H3* three times in a row, going through *P* 2 and *P* 1 and *P* 0. This search tree explains why `eauto` came up with a proof term starting with an application of *H3*.

21.2.3 Adding Hints

By default, `auto` (and `eauto`) only tries to apply the hypotheses that appear in the proof context. There are two possibilities for telling `auto` to exploit a lemma that have been proved previously: either adding the lemma as an assumption just before calling `auto`, or adding the lemma as a hint, so that it can be used by every calls to `auto`.

The first possibility is useful to have `auto` exploit a lemma that only serves at this particular point. To add the lemma as hypothesis, one can type `generalize` *mylemma*; `intros`, or simply *lets*: *mylemma* (the latter requires *LibTactics.v*).

The second possibility is useful for lemmas that need to be exploited several times. The syntax for adding a lemma as a hint is `Hint Resolve` *mylemma*. For example:

```
Lemma nat_le_refl : ∀ (x:nat), x ≤ x.
Proof. apply le_n. Qed.
```

```
Hint Resolve nat_le_refl.
```

A convenient shorthand for adding all the constructors of an inductive datatype as hints is the command `Hint Constructors` *mydatatype*.

Warning: some lemmas, such as transitivity results, should not be added as hints as they would very badly affect the performance of proof search. The description of this problem and the presentation of a general work-around for transitivity lemmas appear further on.

21.2.4 Integration of Automation in Tactics

The library "LibTactics" introduces a convenient feature for invoking automation after calling a tactic. In short, it suffices to add the symbol star (×) to the name of a tactic. For example, `apply×` H is equivalent to `apply` H; *auto_star*, where *auto_star* is a tactic that can be defined as needed.

The definition of *auto_star*, which determines the meaning of the star symbol, can be modified whenever needed. Simply write:

Ltac auto_star ::= a_new_definition.

Observe the use of ::= instead of :=, which indicates that the tactic is being rebound to a new definition. So, the default definition is as follows.

```
Ltac auto_star ::= try solve [ jauto ].
```

Nearly all standard Coq tactics and all the tactics from "LibTactics" can be called with a star symbol. For example, one can invoke `subst×`, `destruct×` H, *inverts×* H, *lets×* I: H `x`, *specializes×* H `x`, and so on... There are two notable exceptions. The tactic `auto×` is just another name for the tactic *auto_star*. And the tactic `apply×` H calls `eapply` H (or the more powerful *applys* H if needed), and then calls *auto_star*. Note that there is no `eapply×` H tactic, use `apply×` H instead.

In large developments, it can be convenient to use two degrees of automation. Typically, one would use a fast tactic, like `auto`, and a slower but more powerful tactic, like *jauto*. To allow for a smooth coexistence of the two form of automation, *LibTactics.v* also defines a "tilde" version of tactics, like apply¬ *H*, destruct¬ *H*, subst¬, auto¬ and so on. The meaning of the tilde symbol is described by the *auto_tilde* tactic, whose default implementation is auto.

```
Ltac auto_tilde ::= auto.
```

In the examples that follow, only *auto_star* is needed.

An alternative, possibly more efficient version of auto_star is the following":

Ltac auto_star ::= try solve *eassumption* | auto | *jauto* .

With the above definition, *auto_star* first tries to solve the goal using the assumptions; if it fails, it tries using auto, and if this still fails, then it calls *jauto*. Even though *jauto* is strictly stronger than *eassumption* and auto, it makes sense to call these tactics first, because, when the succeed, they save a lot of time, and when they fail to prove the goal, they fail very quickly.".

21.3 Example Proofs using Automation

Let's see how to use proof search in practice on the main theorems of the "Software Foundations" course, proving in particular results such as determinism, preservation and progress.

21.3.1 Determinism

```
Module DeterministicImp.
  Import Imp.
```

Recall the original proof of the determinism lemma for the IMP language, shown below.

```
Theorem ceval_deterministic: ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  intros c st st1 st2 E1 E2.
  generalize dependent st2.
  (induction E1); intros st2 E2; inversion E2; subst.
  - reflexivity.
  - reflexivity.
  -
    assert (st' = st'0) as EQ1.
    { apply IHE1_1; assumption. }
    subst st'0.
    apply IHE1_2. assumption.
  -
    apply IHE1. assumption.
  -
```

```
    rewrite H in H5. inversion H5.
  -
    rewrite H in H5. inversion H5.
  -
      apply IHE1. assumption.
  -
    reflexivity.
  -
    rewrite H in H2. inversion H2.
  -
    rewrite H in H4. inversion H4.
  -
    assert (st' = st'0) as EQ1.
    { apply IHE1_1; assumption. }
    subst st'0.
    apply IHE1_2. assumption.
Qed.
```

Exercise: rewrite this proof using `auto` whenever possible. (The solution uses `auto` 9 times.)

```
Theorem ceval_deterministic': ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
   admit.
Admitted.
```

In fact, using automation is not just a matter of calling `auto` in place of one or two other tactics. Using automation is about rethinking the organization of sequences of tactics so as to minimize the effort involved in writing and maintaining the proof. This process is eased by the use of the tactics from *LibTactics.v*. So, before trying to optimize the way automation is used, let's first rewrite the proof of determinism:

- use *introv* H instead of `intros x` H,
- use *gen* x instead of `generalize dependent x`,
- use *inverts* H instead of `inversion` H; `subst`,
- use *tryfalse* to handle contradictions, and get rid of the cases where beval *st b1* = true and beval *st b1* = false both appear in the context.

```
Theorem ceval_deterministic'': ∀ c st st1 st2,
  st =[ c ]=> st1 →
```

```
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  introv E1 E2. gen st2.
  induction E1; intros; inverts E2; tryfalse.
  - auto.
  - auto.
  - assert (st' = st'0). auto. subst. auto.
  - auto.
  - auto.
  - auto.
  - assert (st' = st'0). auto. subst. auto.
Qed.
```

To obtain a nice clean proof script, we have to remove the calls assert (*st'* = *st'0*). Such a tactic call is not nice because it refers to some variables whose name has been automatically generated. This kind of tactics tend to be very brittle. The tactic assert (*st'* = *st'0*) is used to assert the conclusion that we want to derive from the induction hypothesis. So, rather than stating this conclusion explicitly, we are going to ask Coq to instantiate the induction hypothesis, using automation to figure out how to instantiate it. The tactic *forwards*, described in *LibTactics.v* precisely helps with instantiating a fact. So, let's see how it works out on our example.

```
Theorem ceval_deterministic''': ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  introv E1 E2. gen st2.
  induction E1; intros; inverts E2; tryfalse.
  - auto.
  - auto.
  - dup 4.
  + assert (st' = st'0). apply IHE1_1. apply H1.
    skip.
  + forwards: IHE1_1. apply H1.
    skip.
  + forwards: IHE1_1. eauto.
    skip.
  + forwards*: IHE1_1.
    skip.
Abort.
```

To polish the proof script, it remains to factorize the calls to auto, using the star symbol. The proof of determinism can then be rewritten in just 4 lines, including no more than 10 tactics.

```
Theorem ceval_deterministic'''': ∀ c st st1 st2,
  st =[ c ]=> st1 →
  st =[ c ]=> st2 →
  st1 = st2.
Proof.
  introv E1 E2. gen st2.
  induction E1; intros; inverts× E2; tryfalse.
  - forwards*: IHE1_1. subst×.
  - forwards*: IHE1_1. subst×.
Qed.
End DeterministicImp.
```

21.3.2 Preservation for STLC

To investigate how to automate the proof of the lemma preservation, let us first import the definitions required to state that lemma.

```
Set Warnings "-notation-overridden,-parsing".
From PLF Require Import StlcProp.
Module PreservationProgressStlc.
Import STLC.
Import STLCProp.
```

Consider the proof of perservation of STLC, shown below. This proof already uses eauto through the triple-dot mechanism.

```
Theorem preservation : ∀ t t' T,
  has_type empty t T →
  t -> t' →
  has_type empty t' T.
Proof with eauto.
  remember (@empty ty) as Gamma.
  intros t t' T HT. generalize dependent t'.
  (induction HT); intros t' HE; subst Gamma.
  -
    inversion HE.
  -
    inversion HE.
  -
    inversion HE; subst...
    +
```

```
      apply substitution_preserves_typing with T11...
      inversion HT1...
  -
    inversion HE.
  -
    inversion HE.
  -
    inversion HE; subst...
Qed.
```

Exercise: rewrite this proof using tactics from LibTactics and calling automation using the star symbol rather than the triple-dot notation. More precisely, make use of the tactics *inverts*× and *applys*× to call auto× after a call to *inverts* or to *applys*. The solution is three lines long.

```
Theorem preservation' : ∀ t t' T,
  has_type empty t T →
  t -> t' →
  has_type empty t' T.
Proof.
   admit.
Admitted.
```

21.3.3 Progress for STLC

Consider the proof of the progress theorem.

```
Theorem progress : ∀ t T,
  has_type empty t T →
  value t ∨ ∃ t', t -> t'.
Proof with eauto.
  intros t T Ht.
  remember (@empty ty) as Gamma.
  (induction Ht); subst Gamma...
  -
    inversion H.
  -
    right. destruct IHHt1...
    +
      destruct IHHt2...
      ×
        inversion H; subst; try solve_by_invert.
        ∃ ([x0:=t2]t)...
      ×
       destruct H0 as [t2' Hstp]. ∃ (app t1 t2')...
```

```
    +
      destruct H as [t1' Hstp]. ∃ (app t1' t2)...
  -
    right. destruct IHHt1...
    destruct t1; try solve_by_invert...
    inversion H. ∃ (test x0 t2 t3)...
Qed.
```

Exercise: optimize the above proof. Hint: make use of destruct× and *inverts*×. The solution fits on 10 short lines.

```
Theorem progress' : ∀ t T,
  has_type empty t T →
  value t ∨ ∃ t', t -> t'.
Proof.
   admit.
Admitted.
End PreservationProgressStlc.
```

21.3.4 BigStep and SmallStep

```
From PLF Require Import Smallstep.
Require Import Program.
Module Semantics.
```

Consider the proof relating a small-step reduction judgment to a big-step reduction judgment.

```
Theorem multistep__eval : ∀ t v,
  normal_form_of t v → ∃ n, v = C n ∧ t ==> n.
Proof.
  intros t v Hnorm.
  unfold normal_form_of in Hnorm.
  inversion Hnorm as [Hs Hnf]; clear Hnorm.
  rewrite nf_same_as_value in Hnf. inversion Hnf. clear Hnf.
  ∃ n. split. reflexivity.
  induction Hs; subst.
  -
    apply E_Const.
  -
    eapply step__eval. eassumption. apply IHHs. reflexivity.
Qed.
```

Our goal is to optimize the above proof. It is generally easier to isolate inductions into separate lemmas. So, we are going to first prove an intermediate result that consists of the judgment over which the induction is being performed.

Exercise: prove the following result, using tactics *introv*, `induction` and `subst`, and `apply×`. The solution fits on 3 short lines.

```
Theorem multistep_eval_ind : ∀ t v,
  t ->* v → ∀ n, C n = v → t ==> n.
Proof.
  admit.
Admitted.
```

Exercise: using the lemma above, simplify the proof of the result multistep__eval. You should use the tactics *introv*, *inverts*, `split×` and `apply×`. The solution fits on 2 lines.

```
Theorem multistep__eval' : ∀ t v,
  normal_form_of t v → ∃ n, v = C n ∧ t ==> n.
Proof.
  admit.
Admitted.
```

If we try to combine the two proofs into a single one, we will likely fail, because of a limitation of the `induction` tactic. Indeed, this tactic loses information when applied to a property whose arguments are not reduced to variables, such as t –>* (C n). You will thus need to use the more powerful tactic called `dependent induction`. (This tactic is available only after importing the Program library, as we did above.)

Exercise: prove the lemma multistep__eval without invoking the lemma multistep_eval_ind, that is, by inlining the proof by induction involved in multistep_eval_ind, using the tactic `dependent induction` instead of `induction`. The solution fits on 6 lines.

```
Theorem multistep__eval'' : ∀ t v,
  normal_form_of t v → ∃ n, v = C n ∧ t ==> n.
Proof.
  admit.
Admitted.

End Semantics.
```

21.3.5 Preservation for STLCRef

```
From Coq Require Import omega.Omega.
From PLF Require Import References.
Import STLCRef.
Require Import Program.
Module PreservationProgressReferences.
Hint Resolve store_weakening extends_refl.
```

The proof of preservation for STLCRef can be found in chapter References. The optimized proof script is more than twice shorter. The following material explains how to build the optimized proof script. The resulting optimized proof script for the preservation theorem appears afterwards.

```
Theorem preservation : ∀ ST t t' T st st',
  has_type empty ST t T →
  store_well_typed ST st →
  t / st -> t' / st' →
  ∃ ST',
    (extends ST' ST ∧
     has_type empty ST' t' T ∧
     store_well_typed ST' st').
Proof.
  remember (@empty ty) as Gamma. introv Ht. gen t'.
  (induction Ht); introv HST Hstep;

   subst Gamma; inverts Hstep; eauto.

  -


  ∃ ST. inverts Ht1. splits×. applys× substitution_preserves_typing.

  -


  forwards: IHHt1. eauto. eauto. eauto.
  jauto_set_hyps; intros.
  jauto_set_goal; intros.
  eauto. eauto. eauto.

  -


  forwards*: IHHt2.

  - forwards*: IHHt.
  - forwards*: IHHt.
  - forwards*: IHHt1.
  - forwards*: IHHt2.
  - forwards*: IHHt1.

  -
  +


    ∃ (ST ++ T1::nil). inverts keep HST. splits.
      apply extends_app.
      applys_eq T_Loc 1.
        rewrite app_length. simpl. omega.
```

```
    unfold store_Tlookup. rewrite ← H. rewrite× app_nth2.
  rewrite minus_diag. simpl. reflexivity.
  apply× store_well_typed_app.
- forwards*: IHHt.
-
+
```

```
∃ ST. splits×.
lets [_ Hsty]: HST.
applys_eq× Hsty 1.
inverts× Ht.
- forwards*: IHHt.
-
+
```

```
∃ ST. splits×. applys× assign_pres_store_typing. inverts× Ht1.
- forwards*: IHHt1.
- forwards*: IHHt2.
Qed.
```

Let's come back to the proof case that was hard to optimize. The difficulty comes from the statement of nth_eq_last, which takes the form nth (length l) (l ++ x::nil) d = x. This lemma is hard to exploit because its first argument, length l, mentions a list l that has to be exactly the same as the l occuring in l ++ x::nil. In practice, the first argument is often a natural number n that is provably equal to length l yet that is not syntactically equal to length l. There is a simple fix for making nth_eq_last easy to apply: introduce the intermediate variable n explicitly, so that the goal becomes nth n (l ++ x::nil) d = x, with a premise asserting n = length l.

```
Lemma nth_eq_last' : ∀ (A : Type) (l : list A) (x d : A) (n : nat),
  n = length l → nth n (l ++ x::nil) d = x.
Proof. intros. subst. apply nth_eq_last. Qed.
```

The proof case for ref from the preservation theorem then becomes much easier to prove, because rewrite nth_eq_last' now succeeds.

```
Lemma preservation_ref : ∀ (st:store) (ST : store_ty) T1,
  length ST = length st →
  Ref T1 = Ref (store_Tlookup (length st) (ST ++ T1::nil)).
Proof.
  intros. dup.
  unfold store_Tlookup. rewrite× nth_eq_last'.
```

```
  fequal. symmetry. apply× nth_eq_last'.
Qed.
```

The optimized proof of preservation is summarized next.

```
Theorem preservation' : ∀ ST t t' T st st',
  has_type empty ST t T →
  store_well_typed ST st →
  t / st -> t' / st' →
  ∃ ST',
    (extends ST' ST ∧
     has_type empty ST' t' T ∧
     store_well_typed ST' st').
Proof.
  remember (@empty ty) as Gamma. introv Ht. gen t'.
  induction Ht; introv HST Hstep; subst Gamma; inverts Hstep; eauto.
  - ∃ ST. inverts Ht1. splits×. applys× substitution_preserves_typing.
  - forwards*: IHHt1.
  - forwards*: IHHt2.
  - forwards*: IHHt.
  - forwards*: IHHt.
  - forwards*: IHHt1.
  - forwards*: IHHt2.
  - forwards*: IHHt1.
  - ∃ (ST ++ T1::nil). inverts keep HST. splits.
    apply extends_app.
    applys_eq T_Loc 1.
      rewrite app_length. simpl. omega.
      unfold store_Tlookup. rewrite× nth_eq_last'.
    apply× store_well_typed_app.
  - forwards*: IHHt.
  - ∃ ST. splits×. lets [_ Hsty]: HST.
    applys_eq× Hsty 1. inverts× Ht.
  - forwards*: IHHt.
  - ∃ ST. splits×. applys× assign_pres_store_typing. inverts× Ht1.
  - forwards*: IHHt1.
  - forwards*: IHHt2.
Qed.
```

21.3.6 Progress for STLCRef

The proof of progress for STLCRef can be found in chapter References. The optimized proof script is, here again, about half the length.

```
Theorem progress : ∀ ST t T st,
```

```
  has_type empty ST t T →
  store_well_typed ST st →
  (value t ∨ ∃ t' st', t / st -> t' / st').
Proof.
  introv Ht HST. remember (@empty ty) as Gamma.
  induction Ht; subst Gamma; tryfalse; try solve [left*].
  - right. destruct× IHHt1 as [K||].
    inverts K; inverts Ht1.
     destruct× IHHt2.
  - right. destruct× IHHt as [K||].
    inverts K; try solve [inverts Ht]. eauto.
  - right. destruct× IHHt as [K||].
    inverts K; try solve [inverts Ht]. eauto.
  - right. destruct× IHHt1 as [K||].
    inverts K; try solve [inverts Ht1].
     destruct× IHHt2 as [M||].
      inverts M; try solve [inverts Ht2]. eauto.
  - right. destruct× IHHt1 as [K||].
    inverts K; try solve [inverts Ht1]. destruct× n.
  - right. destruct× IHHt.
  - right. destruct× IHHt as [K||].
    inverts K; inverts Ht as M.
      inverts HST as N. rewrite× N in M.
  - right. destruct× IHHt1 as [K||].
    destruct× IHHt2.
     inverts K; inverts Ht1 as M.
     inverts HST as N. rewrite× N in M.
Qed.
End PreservationProgressReferences.
```

21.3.7 Subtyping

```
From PLF Require Sub.
Module SubtypingInversion.
Import Sub.
```

Consider the inversion lemma for typing judgment of abstractions in a type system with subtyping.

```
Lemma abs_arrow : ∀ x S1 s2 T1 T2,
  has_type empty (abs x S1 s2) (Arrow T1 T2) →
     subtype T1 S1
  ∧ has_type (update empty x S1) s2 T2.
Proof with eauto.
```

```
  intros x S1 s2 T1 T2 Hty.
  apply typing_inversion_abs in Hty.
  destruct Hty as [S2 [Hsub Hty]].
  apply sub_inversion_arrow in Hsub.
  destruct Hsub as [U1 [U2 [Heq [Hsub1 Hsub2]]]].
  inversion Heq; subst...
Qed.
```

Exercise: optimize the proof script, using *introv*, *lets* and *inverts*×. In particular, you will find it useful to replace the pattern apply K in H. destruct H as I with *lets* I: K H. The solution fits on 3 lines.

```
Lemma abs_arrow' : ∀ x S1 s2 T1 T2,
  has_type empty (abs x S1 s2) (Arrow T1 T2) →
     subtype T1 S1
  ∧ has_type (update empty x S1) s2 T2.
Proof.
   admit.
Admitted.
End SubtypingInversion.
```

21.4 Advanced Topics in Proof Search

21.4.1 Stating Lemmas in the Right Way

Due to its depth-first strategy, eauto can get exponentially slower as the depth search increases, even when a short proof exists. In general, to make proof search run reasonably fast, one should avoid using a depth search greater than 5 or 6. Moreover, one should try to minimize the number of applicable lemmas, and usually put first the hypotheses whose proof usefully instantiates the existential variables.

In fact, the ability for eauto to solve certain goals actually depends on the order in which the hypotheses are stated. This point is illustrated through the following example, in which P is a property of natural numbers. This property is such that P n holds for any n as soon as P m holds for at least one m different from zero. The goal is to prove that P 2 implies P 1. When the hypothesis about P is stated in the form ∀ n m, P m → m ≠ 0 → P n, then eauto works. However, with ∀ n m, m ≠ 0 → P m → P n, the tactic eauto fails.

```
Lemma order_matters_1 : ∀ (P : nat→Prop),
  (∀ n m, P m → m ≠ 0 → P n) →
  P 2 →
  P 1.
Proof.
  eauto. Qed.
Lemma order_matters_2 : ∀ (P : nat→Prop),
```

```
  (∀ n m, m ≠ 0 → P m → P n) →
  P 5 →
  P 1.
Proof.
  eauto.
  intros P H K.
  eapply H.
  eauto.
Abort.
```

It is very important to understand that the hypothesis ∀ n m, P m → m ≠ 0 → P n is eauto-friendly, whereas ∀ n m, m ≠ 0 → P m → P n really isn't. Guessing a value of m for which P m holds and then checking that m ≠ 0 holds works well because there are few values of m for which P m holds. So, it is likely that `eauto` comes up with the right one. On the other hand, guessing a value of m for which m ≠ 0 and then checking that P m holds does not work well, because there are many values of m that satisfy m ≠ 0 but not P m.

21.4.2 Unfolding of Definitions During Proof-Search

The use of intermediate definitions is generally encouraged in a formal development as it usually leads to more concise and more readable statements. Yet, definitions can make it a little harder to automate proofs. The problem is that it is not obvious for a proof search mechanism to know when definitions need to be unfolded. Note that a naive strategy that consists in unfolding all definitions before calling proof search does not scale up to large proofs, so we avoid it. This section introduces a few techniques for avoiding to manually unfold definitions before calling proof search.

To illustrate the treatment of definitions, let P be an abstract property on natural numbers, and let myFact be a definition denoting the proposition P x holds for any x less than or equal to 3.

```
Axiom P : nat → Prop.
Definition myFact := ∀ x, x ≤ 3 → P x.
```

Proving that myFact under the assumption that P x holds for any x should be trivial. Yet, `auto` fails to prove it unless we unfold the definition of myFact explicitly.

```
Lemma demo_hint_unfold_goal_1 :
  (∀ x, P x) →
  myFact.
Proof.
  auto.   unfold myFact. auto. Qed.
```

To automate the unfolding of definitions that appear as proof obligation, one can use the command `Hint Unfold` myFact to tell Coq that it should always try to unfold myFact when myFact appears in the goal.

```
Hint Unfold myFact.
```

This time, automation is able to see through the definition of myFact.

```
Lemma demo_hint_unfold_goal_2 :
  (∀ x, P x) →
  myFact.
Proof. auto. Qed.
```

However, the `Hint Unfold` mechanism only works for unfolding definitions that appear in the goal. In general, proof search does not unfold definitions from the context. For example, assume we want to prove that P 3 holds under the assumption that **True** → myFact.

```
Lemma demo_hint_unfold_context_1 :
  (True → myFact) →
  P 3.
Proof.
  intros.
  auto.   unfold myFact in *. auto. Qed.
```

There is actually one exception to the previous rule: a constant occuring in an hypothesis is automatically unfolded if the hypothesis can be directly applied to the current goal. For example, `auto` can prove myFact → P 3, as illustrated below.

```
Lemma demo_hint_unfold_context_2 :
  myFact →
  P 3.
Proof. auto. Qed.
```

21.4.3 Automation for Proving Absurd Goals

In this section, we'll see that lemmas concluding on a negation are generally not useful as hints, and that lemmas whose conclusion is **False** can be useful hints but having too many of them makes proof search inefficient. We'll also see a practical work-around to the efficiency issue.

Consider the following lemma, which asserts that a number less than or equal to 3 is not greater than 3.

```
Parameter le_not_gt : ∀ x,
  (x ≤ 3) →
  ¬ (x > 3).
```

Equivalently, one could state that a number greater than three is not less than or equal to 3.

```
Parameter gt_not_le : ∀ x,
  (x > 3) →
  ¬ (x ≤ 3).
```

In fact, both statements are equivalent to a third one stating that $x \leq 3$ and $x > 3$ are contradictory, in the sense that they imply False.

```
Parameter le_gt_false : ∀ x,
  (x ≤ 3) →
  (x > 3) →
  False.
```

The following investigation aim at figuring out which of the three statments is the most convenient with respect to proof automation. The following material is enclosed inside a `Section`, so as to restrict the scope of the hints that we are adding. In other words, after the end of the section, the hints added within the section will no longer be active.

```
Section DemoAbsurd1.
```

Let's try to add the first lemma, *le_not_gt*, as hint, and see whether we can prove that the proposition $\exists$ x, $x \leq 3 \wedge x > 3$ is absurd.

```
Hint Resolve le_not_gt.

Lemma demo_auto_absurd_1 :
  (∃ x, x ≤ 3 ∧ x > 3) →
  False.
Proof.
  intros. jauto_set. eauto.   eapply le_not_gt. eauto. eauto.
Qed.
```

The lemma *gt_not_le* is symmetric to *le_not_gt*, so it will not be any better. The third lemma, *le_gt_false*, is a more useful hint, because it concludes on False, so proof search will try to apply it when the current goal is False.

```
Hint Resolve le_gt_false.

Lemma demo_auto_absurd_2 :
  (∃ x, x ≤ 3 ∧ x > 3) →
  False.
Proof.
  dup.

  intros. jauto_set. eauto.

  jauto.
Qed.
```

In summary, a lemma of the form $H1 \to H2 \to$ False is a much more effective hint than $H1 \to \neg H2$, even though the two statments are equivalent up to the definition of the negation symbol $\neg$.

That said, one should be careful with adding lemmas whose conclusion is False as hint. The reason is that whenever reaching the goal False, the proof search mechanism will potentially try to apply all the hints whose conclusion is False before applying the appropriate one.

End DemoAbsurd1.

Adding lemmas whose conclusion is **False** as hint can be, locally, a very effective solution. However, this approach does not scale up for global hints. For most practical applications, it is reasonable to give the name of the lemmas to be exploited for deriving a contradiction. The tactic false H, provided by LibTactics serves that purpose: false H replaces the goal with **False** and calls `eapply` H. Its behavior is described next. Observe that any of the three statements *le_not_gt*, *gt_not_le* or *le_gt_false* can be used.

```
Lemma demo_false : ∀ x,
  (x ≤ 3) →
  (x > 3) →
  4 = 5.
Proof.
  intros. dup 4.
  - false. eapply le_gt_false.
    + auto.      + skip.
  - false. eapply le_gt_false.
    + eauto.     + eauto.
  - false le_gt_false. eauto. eauto.
  - false le_not_gt. eauto. eauto.
Abort.
```

In the above example, false *le_gt_false*; `eauto` proves the goal, but false *le_gt_false*; `auto` does not, because `auto` does not correctly instantiate the existential variable. Note that false× *le_gt_false* would not work either, because the star symbol tries to call `auto` first. So, there are two possibilities for completing the proof: either call false *le_gt_false*; `eauto`, or call false× (*le_gt_false* 3).

21.4.4 Automation for Transitivity Lemmas

Some lemmas should never be added as hints, because they would very badly slow down proof search. The typical example is that of transitivity results. This section describes the problem and presents a general workaround.

Consider a subtyping relation, written *subtype* S T, that relates two object S and T of type *typ*. Assume that this relation has been proved reflexive and transitive. The corresponding lemmas are named *subtype_refl* and *subtype_trans*.

```
Parameter typ : Type.

Parameter subtype : typ → typ → Prop.

Parameter subtype_refl : ∀ T,
  subtype T T.

Parameter subtype_trans : ∀ S T U,
```

subtype S T → *subtype* T U → *subtype* S U.

Adding reflexivity as hint is generally a good idea, so let's add reflexivity of subtyping as hint.

```
Hint Resolve subtype_refl.
```

Adding transitivity as hint is generally a bad idea. To understand why, let's add it as hint and see what happens. Because we cannot remove hints once we've added them, we are going to open a "Section," so as to restrict the scope of the transitivity hint to that section.

```
Section HintsTransitivity.

Hint Resolve subtype_trans.
```

Now, consider the goal ∀ S T, *subtype* S T, which clearly has no hope of being solved. Let's call `eauto` on this goal.

```
Lemma transitivity_bad_hint_1 : ∀ S T,
  subtype S T.
Proof.
  intros. eauto. Abort.
```

Note that after closing the section, the hint *subtype_trans* is no longer active.

```
End HintsTransitivity.
```

In the previous example, the proof search has spent a lot of time trying to apply transitivity and reflexivity in every possible way. Its process can be summarized as follows. The first goal is *subtype* S T. Since reflexivity does not apply, `eauto` invokes transitivity, which produces two subgoals, *subtype* S ?X and *subtype* ?X T. Solving the first subgoal, *subtype* S ?X, is straightforward, it suffices to apply reflexivity. This unifies ?X with S. So, the second sugoal, *subtype* ?X T, becomes *subtype* S T, which is exactly what we started from...

The problem with the transitivity lemma is that it is applicable to any goal concluding on a subtyping relation. Because of this, `eauto` keeps trying to apply it even though it most often doesn't help to solve the goal. So, one should never add a transitivity lemma as a hint for proof search.

There is a general workaround for having automation to exploit transitivity lemmas without giving up on efficiency. This workaround relies on a powerful mechanism called "external hint." This mechanism allows to manually describe the condition under which a particular lemma should be tried out during proof search.

For the case of transitivity of subtyping, we are going to tell Coq to try and apply the transitivity lemma on a goal of the form *subtype* S U only when the proof context already contains an assumption either of the form *subtype* S T or of the form *subtype* T U. In other words, we only apply the transitivity lemma when there is some evidence that this application might help. To set up this "external hint," one has to write the following.

```
Hint Extern 1 (subtype ?S ?U) ⇒
  match goal with
  | H: subtype S ?T ⊢ _ ⇒ apply (@subtype_trans S T U)
```

```
| H: subtype ?T U ⊢ _ ⇒ apply (@subtype_trans S T U)
end.
```

This hint declaration can be understood as follows.

- "Hint Extern" introduces the hint.
- The number "1" corresponds to a priority for proof search. It doesn't matter so much what priority is used in practice.
- The pattern *subtype* ?S ?U describes the kind of goal on which the pattern should apply. The question marks are used to indicate that the variables ?S and ?U should be bound to some value in the rest of the hint description.
- The construction `match goal with` ... `end` tries to recognize patterns in the goal, or in the proof context, or both.
- The first pattern is H: *subtype* S ?T ⊢ _. It indices that the context should contain an hypothesis H of type *subtype* S ?T, where S has to be the same as in the goal, and where ?T can have any value.
- The symbol ⊢ _ at the end of H: *subtype* S ?T ⊢ _ indicates that we do not impose further condition on how the proof obligation has to look like.
- The branch ⇒ `apply` (@*subtype_trans* S T U) that follows indicates that if the goal has the form *subtype* S U and if there exists an hypothesis of the form *subtype* S T, then we should try and apply transitivity lemma instantiated on the arguments S, T and U. (Note: the symbol @ in front of *subtype_trans* is only actually needed when the "Implicit Arguments" feature is activated.)
- The other branch, which corresponds to an hypothesis of the form H: *subtype* ?T U is symmetrical.

Note: the same external hint can be reused for any other transitive relation, simply by renaming *subtype* into the name of that relation.

Let us see an example illustrating how the hint works.

```
Lemma transitivity_workaround_1 : ∀ T1 T2 T3 T4,
  subtype T1 T2 →
  subtype T2 T3 →
  subtype T3 T4 →
  subtype T1 T4.
Proof.
  intros. eauto. Qed.
```

We may also check that the new external hint does not suffer from the complexity blow up.

```
Lemma transitivity_workaround_2 : ∀ S T,
  subtype S T.
Proof.
  intros. eauto. Abort.
```

21.5 Decision Procedures

A decision procedure is able to solve proof obligations whose statement admits a particular form. This section describes three useful decision procedures. The tactic `omega` handles goals involving arithmetic and inequalities, but not general multiplications. The tactic `ring` handles goals involving arithmetic, including multiplications, but does not support inequalities. The tactic `congruence` is able to prove equalities and inequalities by exploiting equalities available in the proof context.

21.5.1 Omega

The tactic `omega` supports natural numbers (type **nat**) as well as integers (type Z, available by including the module ZArith). It supports addition, substraction, equalities and inequalities. Before using `omega`, one needs to import the module Omega, as follows.

```
Require Import Omega.
```

Here is an example. Let x and y be two natural numbers (they cannot be negative). Assume y is less than 4, assume x+x+1 is less than y, and assume x is not zero. Then, it must be the case that x is equal to one.

```
Lemma omega_demo_1 : ∀ (x y : nat),
  (y ≤ 4) →
  (x + x + 1 ≤ y) →
  (x ≠ 0) →
  (x = 1).
Proof. intros. omega. Qed.
```

Another example: if z is the mean of x and y, and if the difference between x and y is at most 4, then the difference between x and z is at most 2.

```
Lemma omega_demo_2 : ∀ (x y z : nat),
  (x + y = z + z) →
  (x - y ≤ 4) →
  (x - z ≤ 2).
Proof. intros. omega. Qed.
```

One can proof **False** using `omega` if the mathematical facts from the context are contradictory. In the following example, the constraints on the values x and y cannot be all satisfied in the same time.

```
Lemma omega_demo_3 : ∀ (x y : nat),
```

```
  (x + 5 ≤ y) →
  (y - x < 3) →
  False.
Proof. intros. omega. Qed.
```

Note: `omega` can prove a goal by contradiction only if its conclusion reduces to **False**. The tactic `omega` always fails when the conclusion is an arbitrary proposition P, even though **False** implies any proposition P (by *ex_falso_quodlibet*).

```
Lemma omega_demo_4 : ∀ (x y : nat) (P : Prop),
  (x + 5 ≤ y) →
  (y - x < 3) →
  P.
Proof.
  intros.
  false. omega.
Qed.
```

21.5.2 Ring

Compared with `omega`, the tactic `ring` adds support for multiplications, however it gives up the ability to reason on inequations. Moreover, it supports only integers (type Z) and not natural numbers (type **nat**). Here is an example showing how to use `ring`.

```
Require Import ZArith.
Module RingDemo.
  Open Scope Z_scope.

Lemma ring_demo : ∀ (x y z : Z),
     x × (y + z) - z × 3 × x
  = x × y - 2 × x × z.
Proof. intros. ring. Qed.

End RingDemo.
```

21.5.3 Congruence

The tactic `congruence` is able to exploit equalities from the proof context in order to automatically perform the rewriting operations necessary to establish a goal. It is slightly more powerful than the tactic `subst`, which can only handle equalities of the form $x = e$ where x is a variable and e an expression.

```
Lemma congruence_demo_1 :
   ∀ (f : nat→nat→nat) (g h : nat→nat) (x y z : nat),
   f (g x) (g y) = z →
   2 = g x →
   g y = h z →
```

$f\ 2\ (h\ z) = z.$
```
Proof. intros. congruence. Qed.
```

Moreover, `congruence` is able to exploit universally quantified equalities, for example ∀ a, g a = h a.

```
Lemma congruence_demo_2 :
  ∀ (f : nat→nat→nat) (g h : nat→nat) (x y z : nat),
  (∀ a, g a = h a) →
  f (g x) (g y) = z →
  g x = 2 →
  f 2 (h y) = z.
Proof. congruence. Qed.
```

Next is an example where `congruence` is very useful.

```
Lemma congruence_demo_4 : ∀ (f g : nat→nat),
 (∀ a, f a = g a) →
 f (g (g 2)) = g (f (f 2)).
Proof. congruence. Qed.
```

The tactic `congruence` is able to prove a contradiction if the goal entails an equality that contradicts an inequality available in the proof context.

```
Lemma congruence_demo_3 :
  ∀ (f g h : nat→nat) (x : nat),
  (∀ a, f a = h a) →
  g x = f x →
  g x ≠ h x →
  False.
Proof. congruence. Qed.
```

One of the strengths of `congruence` is that it is a very fast tactic. So, one should not hesitate to invoke it wherever it might help.

21.6 Summary

Let us summarize the main automation tactics available.

- `auto` automatically applies `reflexivity`, `assumption`, and `apply`.
- `eauto` moreover tries `eapply`, and in particular can instantiate existentials in the conclusion.
- *iauto* extends `eauto` with support for negation, conjunctions, and disjunctions. However, its support for disjunction can make it exponentially slow.
- *jauto* extends `eauto` with support for negation, conjunctions, and existential at the head of hypothesis.

- `congruence` helps reasoning about equalities and inequalities.
- `omega` proves arithmetic goals with equalities and inequalities, but it does not support multiplication.
- `ring` proves arithmetic goals with multiplications, but does not support inequalities.

In order to set up automation appropriately, keep in mind the following rule of thumbs:

- automation is all about balance: not enough automation makes proofs not very robust on change, whereas too much automation makes proofs very hard to fix when they break.
- if a lemma is not goal directed (i.e., some of its variables do not occur in its conclusion), then the premises need to be ordered in such a way that proving the first premises maximizes the chances of correctly instantiating the variables that do not occur in the conclusion.
- a lemma whose conclusion is **False** should only be added as a local hint, i.e., as a hint within the current section.
- a transitivity lemma should never be considered as hint; if automation of transitivity reasoning is really necessary, an `Extern Hint` needs to be set up.
- a definition usually needs to be accompanied with a `Hint Unfold`.

Becoming a master in the black art of automation certainly requires some investment, however this investment will pay off very quickly.

Chapter 22

PE: Partial Evaluation

The Equiv chapter introduced constant folding as an example of a program transformation and proved that it preserves the meaning of programs. Constant folding operates on manifest constants such as ANum expressions. For example, it simplifies the command Y ::= 3 + 1 to the command Y ::= 4. However, it does not propagate known constants along data flow. For example, it does not simplify the sequence

X ::= 3;; Y ::= X + 1

to

X ::= 3;; Y ::= 4

because it forgets that X is 3 by the time it gets to Y.

We might naturally want to enhance constant folding so that it propagates known constants and uses them to simplify programs. Doing so constitutes a rudimentary form of *partial evaluation*. As we will see, partial evaluation is so called because it is like running a program, except only part of the program can be evaluated because only part of the input to the program is known. For example, we can only simplify the program

X ::= 3;; Y ::= (X + 1) - Y

to

X ::= 3;; Y ::= 4 - Y

without knowing the initial value of Y.

```
From PLF Require Import Maps.
From Coq Require Import Bool.Bool.
From Coq Require Import Arith.Arith.
From Coq Require Import Arith.EqNat.
From Coq Require Import Arith.PeanoNat. Import Nat.
From Coq Require Import omega.Omega.
From Coq Require Import Logic.FunctionalExtensionality.
From Coq Require Import Lists.List.
Import ListNotations.

From PLF Require Import Smallstep.
From PLF Require Import Imp.
```

22.1 Generalizing Constant Folding

The starting point of partial evaluation is to represent our partial knowledge about the state. For example, between the two assignments above, the partial evaluator may know only that X is 3 and nothing about any other variable.

22.1.1 Partial States

Conceptually speaking, we can think of such partial states as the type **string** → **option nat** (as opposed to the type **string** → **nat** of concrete, full states). However, in addition to looking up and updating the values of individual variables in a partial state, we may also want to compare two partial states to see if and where they differ, to handle conditional control flow. It is not possible to compare two arbitrary functions in this way, so we represent partial states in a more concrete format: as a list of **string** × **nat** pairs.

```
Definition pe_state := list (string × nat).
```

The idea is that a variable (of type **string**) appears in the list if and only if we know its current **nat** value. The pe_lookup function thus interprets this concrete representation. (If the same variable appears multiple times in the list, the first occurrence wins, but we will define our partial evaluator to never construct such a pe_state.)

```
Fixpoint pe_lookup (pe_st : pe_state) (V:string) : option nat :=
  match pe_st with
  | [] ⇒ None
  | (V',n')::pe_st ⇒ if eqb_string V V' then Some n'
                          else pe_lookup pe_st V
  end.
```

For example, empty_pe_state represents complete ignorance about every variable – the function that maps every identifier to None.

```
Definition empty_pe_state : pe_state := [].
```

More generally, if the **list** representing a pe_state does not contain some identifier, then that pe_state must map that identifier to None. Before we prove this fact, we first define a useful tactic for reasoning with **string** equality. The tactic

compare V V'

means to reason by cases over eqb_string V V'. In the case where $V = V'$, the tactic substitutes V for V' throughout.

```
Tactic Notation "compare" ident(i) ident(j) :=
  let H := fresh "Heq" i j in
  destruct (eqb_stringP i j);
  [ subst j | ].
Theorem pe_domain: ∀ pe_st V n,
  pe_lookup pe_st V = Some n →
```

```
  In V (map (@fst _ _) pe_st).
Proof. intros pe_st V n H. induction pe_st as [| [V' n'] pe_st].
  - inversion H.
  - simpl in H. simpl. compare V V'; auto. Qed.
```

In what follows, we will make heavy use of the In property from the standard library, also defined in *Logic.v*:

```
Print In.
```

Besides the various lemmas about In that we've already come across, the following one (taken from the standard library) will also be useful:

```
Check filter_In.
```

If a type A has an operator *eqb* for testing equality of its elements, we can compute a boolean inb *eqb* a l for testing whether In a l holds or not.

```
Fixpoint inb {A : Type} (eqb : A → A → bool) (a : A) (l : list A) :=
  match l with
  | [] ⇒ false
  | a'::l' ⇒ eqb a a' || inb eqb a l'
  end.
```

It is easy to relate inb to In with the **reflect** property:

```
Lemma inbP : ∀ A : Type, ∀ eqb : A→A→bool,
  (∀ a1 a2, reflect (a1 = a2) (eqb a1 a2)) →
  ∀ a l, reflect (In a l) (inb eqb a l).
Proof.
  intros A eqb beqP a l.
  induction l as [|a' l' IH].
  - constructor. intros [].
  - simpl. destruct (beqP a a').
    + subst. constructor. left. reflexivity.
    + simpl. destruct IH; constructor.
      × right. trivial.
      × intros [H1 | H2]; congruence.
Qed.
```

22.1.2 Arithmetic Expressions

Partial evaluation of **aexp** is straightforward – it is basically the same as constant folding, fold_constants_aexp, except that sometimes the partial state tells us the current value of a variable and we can replace it by a constant expression.

```
Fixpoint pe_aexp (pe_st : pe_state) (a : aexp) : aexp :=
  match a with
  | ANum n ⇒ ANum n
```

```
  | AId i ⇒ match pe_lookup pe_st i with
               | Some n ⇒ ANum n
               | None ⇒ AId i
               end
  | APlus a1 a2 ⇒
      match (pe_aexp pe_st a1 , pe_aexp pe_st a2) with
      | (ANum n1 , ANum n2) ⇒ ANum (n1 + n2)
      | (a1', a2') ⇒ APlus a1' a2'
      end
  | AMinus a1 a2 ⇒
      match (pe_aexp pe_st a1 , pe_aexp pe_st a2) with
      | (ANum n1 , ANum n2) ⇒ ANum (n1 - n2)
      | (a1', a2') ⇒ AMinus a1' a2'
      end
  | AMult a1 a2 ⇒
      match (pe_aexp pe_st a1 , pe_aexp pe_st a2) with
      | (ANum n1 , ANum n2) ⇒ ANum (n1 × n2)
      | (a1', a2') ⇒ AMult a1' a2'
      end
  end.
```

This partial evaluator folds constants but does not apply the associativity of addition.

```
Example test_pe_aexp1:
  pe_aexp [(X,3)] (X + 1 + Y)%imp
  = (4 + Y)%imp.
Proof. reflexivity. Qed.

Example text_pe_aexp2:
  pe_aexp [(Y,3)] (X + 1 + Y)%imp
  = (X + 1 + 3)%imp.
Proof. reflexivity. Qed.
```

Now, in what sense is pe_aexp correct? It is reasonable to define the correctness of pe_aexp as follows: whenever a full state *st*:state is *consistent* with a partial state *pe_st*:pe_state (in other words, every variable to which *pe_st* assigns a value is assigned the same value by *st*), evaluating a and evaluating pe_aexp *pe_st* a in *st* yields the same result. This statement is indeed true.

```
Definition pe_consistent (st:state) (pe_st:pe_state) :=
  ∀ V n, Some n = pe_lookup pe_st V → st V = n.

Theorem pe_aexp_correct_weak: ∀ st pe_st, pe_consistent st pe_st →
  ∀ a, aeval st a = aeval st (pe_aexp pe_st a).
Proof. unfold pe_consistent. intros st pe_st H a.
  induction a; simpl;
    try reflexivity;
```

```
    try (destruct (pe_aexp pe_st a1);
         destruct (pe_aexp pe_st a2);
         rewrite IHa1; rewrite IHa2; reflexivity).
  -
    remember (pe_lookup pe_st x) as l. destruct l.
    + rewrite H with (n:=n) by apply Heql. reflexivity.
    + reflexivity.
Qed.
```

However, we will soon want our partial evaluator to remove assignments. For example, it will simplify

X ::= 3;; Y ::= X - Y;; X ::= 4

to just

Y ::= 3 - Y;; X ::= 4

by delaying the assignment to X until the end. To accomplish this simplification, we need the result of partial evaluating

pe_aexp (X,3) (X - Y)

to be equal to 3 - Y and *not* the original expression X - Y. After all, it would be incorrect, not just inefficient, to transform

X ::= 3;; Y ::= X - Y;; X ::= 4

to

Y ::= X - Y;; X ::= 4

even though the output expressions 3 - Y and X - Y both satisfy the correctness criterion that we just proved. Indeed, if we were to just define pe_aexp *pe_st* a = a then the theorem pe_aexp_correct_weak would already trivially hold.

Instead, we want to prove that the pe_aexp is correct in a stronger sense: evaluating the expression produced by partial evaluation (aeval *st* (pe_aexp *pe_st* a)) must not depend on those parts of the full state *st* that are already specified in the partial state *pe_st*. To be more precise, let us define a function *pe_override*, which updates *st* with the contents of *pe_st*. In other words, *pe_override* carries out the assignments listed in *pe_st* on top of *st*.

```
Fixpoint pe_update (st:state) (pe_st:pe_state) : state :=
  match pe_st with
  | [] ⇒ st
  | (V,n)::pe_st ⇒ t_update (pe_update st pe_st) V n
  end.

Example test_pe_update:
  pe_update (Y !-> 1) [(X,3);(Z,2)]
  = (X !-> 3 ; Z !-> 2 ; Y !-> 1).
Proof. reflexivity. Qed.
```

Although pe_update operates on a concrete **list** representing a pe_state, its behavior is defined entirely by the pe_lookup interpretation of the pe_state.

```
Theorem pe_update_correct: ∀ st pe_st V0,
```

```
  pe_update st pe_st V0 =
  match pe_lookup pe_st V0 with
  | Some n ⇒ n
  | None ⇒ st V0
  end.
Proof. intros. induction pe_st as [| [V n] pe_st]. reflexivity.
  simpl in *. unfold t_update.
  compare V0 V; auto. rewrite ← eqb_string_refl; auto. rewrite false_eqb_string; auto.
Qed.
```

We can relate pe_consistent to pe_update in two ways. First, overriding a state with a partial state always gives a state that is consistent with the partial state. Second, if a state is already consistent with a partial state, then overriding the state with the partial state gives the same state.

```
Theorem pe_update_consistent: ∀ st pe_st,
  pe_consistent (pe_update st pe_st) pe_st.
Proof. intros st pe_st V n H. rewrite pe_update_correct.
  destruct (pe_lookup pe_st V); inversion H. reflexivity. Qed.
```

```
Theorem pe_consistent_update: ∀ st pe_st,
  pe_consistent st pe_st → ∀ V, st V = pe_update st pe_st V.
Proof. intros st pe_st H V. rewrite pe_update_correct.
  remember (pe_lookup pe_st V) as l. destruct l; auto. Qed.
```

Now we can state and prove that pe_aexp is correct in the stronger sense that will help us define the rest of the partial evaluator.

Intuitively, running a program using partial evaluation is a two-stage process. In the first, *static* stage, we partially evaluate the given program with respect to some partial state to get a *residual* program. In the second, *dynamic* stage, we evaluate the residual program with respect to the rest of the state. This dynamic state provides values for those variables that are unknown in the static (partial) state. Thus, the residual program should be equivalent to *prepending* the assignments listed in the partial state to the original program.

```
Theorem pe_aexp_correct: ∀ (pe_st:pe_state) (a:aexp) (st:state),
  aeval (pe_update st pe_st) a = aeval st (pe_aexp pe_st a).
Proof.
  intros pe_st a st.
  induction a; simpl;
    try reflexivity;
    try (destruct (pe_aexp pe_st a1);
         destruct (pe_aexp pe_st a2);
         rewrite IHa1; rewrite IHa2; reflexivity).
  rewrite pe_update_correct. destruct (pe_lookup pe_st x); reflexivity.
Qed.
```

22.1.3 Boolean Expressions

The partial evaluation of boolean expressions is similar. In fact, it is entirely analogous to the constant folding of boolean expressions, because our language has no boolean variables.

```
Fixpoint pe_bexp (pe_st : pe_state) (b : bexp) : bexp :=
  match b with
  | BTrue ⇒ BTrue
  | BFalse ⇒ BFalse
  | BEq a1 a2 ⇒
      match (pe_aexp pe_st a1, pe_aexp pe_st a2) with
      | (ANum n1, ANum n2) ⇒ if n1 =? n2 then BTrue else BFalse
      | (a1', a2') ⇒ BEq a1' a2'
      end
  | BLe a1 a2 ⇒
      match (pe_aexp pe_st a1, pe_aexp pe_st a2) with
      | (ANum n1, ANum n2) ⇒ if n1 <=? n2 then BTrue else BFalse
      | (a1', a2') ⇒ BLe a1' a2'
      end
  | BNot b1 ⇒
      match (pe_bexp pe_st b1) with
      | BTrue ⇒ BFalse
      | BFalse ⇒ BTrue
      | b1' ⇒ BNot b1'
      end
  | BAnd b1 b2 ⇒
      match (pe_bexp pe_st b1, pe_bexp pe_st b2) with
      | (BTrue, BTrue) ⇒ BTrue
      | (BTrue, BFalse) ⇒ BFalse
      | (BFalse, BTrue) ⇒ BFalse
      | (BFalse, BFalse) ⇒ BFalse
      | (b1', b2') ⇒ BAnd b1' b2'
      end
  end.

Example test_pe_bexp1:
  pe_bexp [(X,3)] (~(X ≤ 3))%imp
  = false.
Proof. reflexivity. Qed.

Example test_pe_bexp2: ∀ b:bexp,
  b = (~(X ≤ (X + 1)))%imp →
  pe_bexp [] b = b.
Proof. intros b H. rewrite → H. reflexivity. Qed.
```

The correctness of pe_bexp is analogous to the correctness of pe_aexp above.

```
Theorem pe_bexp_correct: ∀ (pe_st:pe_state) (b:bexp) (st:state),
  beval (pe_update st pe_st) b = beval st (pe_bexp pe_st b).
Proof.
  intros pe_st b st.
  induction b; simpl;
    try reflexivity;
    try (remember (pe_aexp pe_st a1) as a1';
         remember (pe_aexp pe_st a2) as a2';
         assert (H1: aeval (pe_update st pe_st) a1 = aeval st a1');
         assert (H2: aeval (pe_update st pe_st) a2 = aeval st a2');
           try (subst; apply pe_aexp_correct);
         destruct a1'; destruct a2'; rewrite H1; rewrite H2;
         simpl; try destruct (n =? n0);
         try destruct (n <=? n0); reflexivity);
    try (destruct (pe_bexp pe_st b); rewrite IHb; reflexivity);
    try (destruct (pe_bexp pe_st b1);
         destruct (pe_bexp pe_st b2);
         rewrite IHb1; rewrite IHb2; reflexivity).
Qed.
```

22.2 Partial Evaluation of Commands, Without Loops

What about the partial evaluation of commands? The analogy between partial evaluation and full evaluation continues: Just as full evaluation of a command turns an initial state into a final state, partial evaluation of a command turns an initial partial state into a final partial state. The difference is that, because the state is partial, some parts of the command may not be executable at the static stage. Therefore, just as pe_aexp returns a residual **aexp** and pe_bexp returns a residual **bexp** above, partially evaluating a command yields a residual command.

Another way in which our partial evaluator is similar to a full evaluator is that it does not terminate on all commands. It is not hard to build a partial evaluator that terminates on all commands; what is hard is building a partial evaluator that terminates on all commands yet automatically performs desired optimizations such as unrolling loops. Often a partial evaluator can be coaxed into terminating more often and performing more optimizations by writing the source program differently so that the separation between static and dynamic information becomes more apparent. Such coaxing is the art of *binding-time improvement*. The binding time of a variable tells when its value is known – either "static", or "dynamic."

Anyway, for now we will just live with the fact that our partial evaluator is not a total function from the source command and the initial partial state to the residual command and the final partial state. To model this non-termination, just as with the full evaluation of commands, we use an inductively defined relation. We write

c1 / st \\ c1' / st'

to mean that partially evaluating the source command *c1* in the initial partial state *st* yields the residual command *c1'* and the final partial state *st'*. For example, we want something like

□ / (X ::= 3 ;; Y ::= Z * (X + X)) \\ (Y ::= Z * 6) / (X,3)

to hold. The assignment to X appears in the final partial state, not the residual command.

(Writing something like *st* =[*c1*]=> *c1'* / *st'* would be closer to the notation used in Imp; perhaps this should be changed!)

22.2.1 Assignment

Let's start by considering how to partially evaluate an assignment. The two assignments in the source program above needs to be treated differently. The first assignment X ::= 3, is *static*: its right-hand-side is a constant (more generally, simplifies to a constant), so we should update our partial state at X to 3 and produce no residual code. (Actually, we produce a residual *SKIP*.) The second assignment Y ::= Z × (X + X) is *dynamic*: its right-hand-side does not simplify to a constant, so we should leave it in the residual code and remove Y, if present, from our partial state. To implement these two cases, we define the functions pe_add and pe_remove. Like pe_update above, these functions operate on a concrete **list** representing a pe_state, but the theorems pe_add_correct and pe_remove_correct specify their behavior by the pe_lookup interpretation of the pe_state.

```
Fixpoint pe_remove (pe_st:pe_state) (V:string) : pe_state :=
  match pe_st with
  | [] ⇒ []
  | (V',n')::pe_st ⇒ if eqb_string V V' then pe_remove pe_st V
                          else (V',n') :: pe_remove pe_st V
  end.
Theorem pe_remove_correct: ∀ pe_st V V0,
  pe_lookup (pe_remove pe_st V) V0
  = if eqb_string V V0 then None else pe_lookup pe_st V0.
Proof. intros pe_st V V0. induction pe_st as [| [V' n'] pe_st].
  - destruct (eqb_string V V0); reflexivity.
  - simpl. compare V V'.
    + rewrite IHpe_st.
      destruct (eqb_stringP V V0). reflexivity.
      rewrite false_eqb_string; auto.
    + simpl. compare V0 V'.
      × rewrite false_eqb_string; auto.
      × rewrite IHpe_st. reflexivity.
Qed.
Definition pe_add (pe_st:pe_state) (V:string) (n:nat) : pe_state :=
  (V,n) :: pe_remove pe_st V.
Theorem pe_add_correct: ∀ pe_st V n V0,
```

```
  pe_lookup (pe_add pe_st V n) V0
  = if eqb_string V V0 then Some n else pe_lookup pe_st V0.
Proof. intros pe_st V n V0. unfold pe_add. simpl.
  compare V V0.
  - rewrite <- eqb_string_refl; auto.
  - rewrite pe_remove_correct.
    repeat rewrite false_eqb_string; auto.
Qed.
```

We will use the two theorems below to show that our partial evaluator correctly deals with dynamic assignments and static assignments, respectively.

```
Theorem pe_update_update_remove: ∀ st pe_st V n,
  t_update (pe_update st pe_st) V n =
  pe_update (t_update st V n) (pe_remove pe_st V).
Proof. intros st pe_st V n. apply functional_extensionality.
  intros V0. unfold t_update. rewrite !pe_update_correct.
  rewrite pe_remove_correct. destruct (eqb_string V V0); reflexivity.
  Qed.

Theorem pe_update_update_add: ∀ st pe_st V n,
  t_update (pe_update st pe_st) V n =
  pe_update st (pe_add pe_st V n).
Proof. intros st pe_st V n. apply functional_extensionality. intros V0.
  unfold t_update. rewrite !pe_update_correct. rewrite pe_add_correct.
  destruct (eqb_string V V0); reflexivity. Qed.
```

22.2.2 Conditional

Trickier than assignments to partially evaluate is the conditional, *TEST b1 THEN c1 ELSE c2 FI*. If *b1* simplifies to BTrue or BFalse then it's easy: we know which branch will be taken, so just take that branch. If *b1* does not simplify to a constant, then we need to take both branches, and the final partial state may differ between the two branches!

The following program illustrates the difficulty:

X ::= 3;; TEST Y <= 4 THEN Y ::= 4;; TEST X <= Y THEN Y ::= 999 ELSE SKIP FI ELSE SKIP FI

Suppose the initial partial state is empty. We don't know statically how Y compares to 4, so we must partially evaluate both branches of the (outer) conditional. On the *THEN* branch, we know that Y is set to 4 and can even use that knowledge to simplify the code somewhat. On the *ELSE* branch, we still don't know the exact value of Y at the end. What should the final partial state and residual program be?

One way to handle such a dynamic conditional is to take the intersection of the final partial states of the two branches. In this example, we take the intersection of (Y,4),(X,3) and (X,3), so the overall final partial state is (X,3). To compensate for forgetting that Y is

4, we need to add an assignment Y ::= 4 to the end of the *THEN* branch. So, the residual program will be something like

SKIP;; TEST Y <= 4 THEN SKIP;; SKIP;; Y ::= 4 ELSE SKIP FI

Programming this case in Coq calls for several auxiliary functions: we need to compute the intersection of two pe_states and turn their difference into sequences of assignments.

First, we show how to compute whether two pe_states to disagree at a given variable. In the theorem pe_disagree_domain, we prove that two pe_states can only disagree at variables that appear in at least one of them.

```
Definition pe_disagree_at (pe_st1 pe_st2 : pe_state) (V:string) : bool :=
  match pe_lookup pe_st1 V, pe_lookup pe_st2 V with
  | Some x, Some y ⇒ negb (x =? y)
  | None, None ⇒ false
  | _, _ ⇒ true
  end.

Theorem pe_disagree_domain: ∀ (pe_st1 pe_st2 : pe_state) (V:string),
  true = pe_disagree_at pe_st1 pe_st2 V →
  In V (map (@fst _ _) pe_st1 ++ map (@fst _ _) pe_st2).
Proof. unfold pe_disagree_at. intros pe_st1 pe_st2 V H.
  apply in_app_iff.
  remember (pe_lookup pe_st1 V) as lookup1.
  destruct lookup1 as [n1|]. left. apply pe_domain with n1. auto.
  remember (pe_lookup pe_st2 V) as lookup2.
  destruct lookup2 as [n2|]. right. apply pe_domain with n2. auto.
  inversion H. Qed.
```

We define the pe_compare function to list the variables where two given pe_states disagree. This list is exact, according to the theorem pe_compare_correct: a variable appears on the list if and only if the two given pe_states disagree at that variable. Furthermore, we use the pe_unique function to eliminate duplicates from the list.

```
Fixpoint pe_unique (l : list string) : list string :=
  match l with
  | [] ⇒ []
  | x::l ⇒
      x :: filter (fun y ⇒ if eqb_string x y then false else true) (pe_unique l)
  end.

Theorem pe_unique_correct: ∀ l x,
  In x l ↔ In x (pe_unique l).
Proof. intros l x. induction l as [| h t]. reflexivity.
  simpl in *. split.
  -
    intros. inversion H; clear H.
      left. assumption.
```

```
      destruct (eqb_stringP h x).
         left. assumption.
         right. apply filter_In. split.
           apply IHt. assumption.
           rewrite false_eqb_string; auto.
  -
    intros. inversion H; clear H.
       left. assumption.
       apply filter_In in H0. inversion H0. right. apply IHt. assumption.
Qed.
Definition pe_compare (pe_st1 pe_st2 : pe_state) : list string :=
  pe_unique (filter (pe_disagree_at pe_st1 pe_st2)
    (map (@fst _ _) pe_st1 ++ map (@fst _ _) pe_st2)).
Theorem pe_compare_correct: ∀ pe_st1 pe_st2 V,
  pe_lookup pe_st1 V = pe_lookup pe_st2 V ↔
  ¬ In V (pe_compare pe_st1 pe_st2).
Proof. intros pe_st1 pe_st2 V.
  unfold pe_compare. rewrite ← pe_unique_correct. rewrite filter_In.
  split; intros Heq.
  -
    intro. destruct H. unfold pe_disagree_at in H0. rewrite Heq in H0.
    destruct (pe_lookup pe_st2 V).
    rewrite ← beq_nat_refl in H0. inversion H0.
    inversion H0.
  -
    assert (Hagree: pe_disagree_at pe_st1 pe_st2 V = false).
    {
      remember (pe_disagree_at pe_st1 pe_st2 V) as disagree.
      destruct disagree; [| reflexivity].
      apply pe_disagree_domain in Heqdisagree.
      exfalso. apply Heq. split. assumption. reflexivity. }
    unfold pe_disagree_at in Hagree.
    destruct (pe_lookup pe_st1 V) as [n1||];
    destruct (pe_lookup pe_st2 V) as [n2||];
      try reflexivity; try solve_by_invert.
    rewrite negb_false_iff in Hagree.
    apply eqb_eq in Hagree. subst. reflexivity. Qed.
```

The intersection of two partial states is the result of removing from one of them all the variables where the two disagree. We define the function pe_removes, in terms of pe_remove above, to perform such a removal of a whole list of variables at once.

The theorem pe_compare_removes testifies that the pe_lookup interpretation of the result of this intersection operation is the same no matter which of the two partial states we remove the variables from. Because pe_update only depends on the pe_lookup interpretation of partial states, pe_update also does not care which of the two partial states we remove the variables from; that theorem pe_compare_update is used in the correctness proof shortly.

```
Fixpoint pe_removes (pe_st:pe_state) (ids : list string) : pe_state :=
  match ids with
  | [] ⇒ pe_st
  | V::ids ⇒ pe_remove (pe_removes pe_st ids) V
  end.

Theorem pe_removes_correct: ∀ pe_st ids V,
  pe_lookup (pe_removes pe_st ids) V =
  if inb eqb_string V ids then None else pe_lookup pe_st V.
Proof. intros pe_st ids V. induction ids as [| V' ids]. reflexivity.
  simpl. rewrite pe_remove_correct. rewrite IHids.
  compare V' V.
  - rewrite ← eqb_string_refl. reflexivity.
  - rewrite false_eqb_string; try congruence. reflexivity.
Qed.

Theorem pe_compare_removes: ∀ pe_st1 pe_st2 V,
  pe_lookup (pe_removes pe_st1 (pe_compare pe_st1 pe_st2)) V =
  pe_lookup (pe_removes pe_st2 (pe_compare pe_st1 pe_st2)) V.
Proof.
  intros pe_st1 pe_st2 V. rewrite !pe_removes_correct.
  destruct (inbP _ _ eqb_stringP V (pe_compare pe_st1 pe_st2)).
  - reflexivity.
  - apply pe_compare_correct. auto. Qed.

Theorem pe_compare_update: ∀ pe_st1 pe_st2 st,
  pe_update st (pe_removes pe_st1 (pe_compare pe_st1 pe_st2)) =
  pe_update st (pe_removes pe_st2 (pe_compare pe_st1 pe_st2)).
Proof. intros. apply functional_extensionality. intros V.
  rewrite !pe_update_correct. rewrite pe_compare_removes. reflexivity.
Qed.
```

Finally, we define an assign function to turn the difference between two partial states into a sequence of assignment commands. More precisely, assign *pe_st* *ids* generates an assignment command for each variable listed in *ids*.

```
Fixpoint assign (pe_st : pe_state) (ids : list string) : com :=
  match ids with
  | [] ⇒ SKIP
  | V::ids ⇒ match pe_lookup pe_st V with
              | Some n ⇒ (assign pe_st ids;; V ::= ANum n)
              | None ⇒ assign pe_st ids
              end
```

```
  end.
```

The command generated by assign always terminates, because it is just a sequence of assignments. The (total) function assigned below computes the effect of the command on the (dynamic state). The theorem assign_removes then confirms that the generated assignments perfectly compensate for removing the variables from the partial state.

```
Definition assigned (pe_st:pe_state) (ids : list string) (st:state) : state :=
  fun V ⇒ if inb eqb_string V ids then
                match pe_lookup pe_st V with
                | Some n ⇒ n
                | None ⇒ st V
                end
          else st V.

Theorem assign_removes: ∀ pe_st ids st,
  pe_update st pe_st =
  pe_update (assigned pe_st ids st) (pe_removes pe_st ids).
Proof. intros pe_st ids st. apply functional_extensionality. intros V.
  rewrite !pe_update_correct. rewrite pe_removes_correct. unfold assigned.
  destruct (inbP _ _ eqb_stringP V ids); destruct (pe_lookup pe_st V); reflexivity.
Qed.

Lemma ceval_extensionality: ∀ c st st1 st2,
  st =[ c ]=> st1 → (∀ V, st1 V = st2 V) → st =[ c ]=> st2.
Proof. intros c st st1 st2 H Heq.
  apply functional_extensionality in Heq. rewrite ← Heq. apply H. Qed.

Theorem eval_assign: ∀ pe_st ids st,
  st =[ assign pe_st ids ]=> assigned pe_st ids st.
Proof. intros pe_st ids st. induction ids as [| V ids]; simpl.
  - eapply ceval_extensionality. apply E_Skip. reflexivity.
  -
    remember (pe_lookup pe_st V) as lookup. destruct lookup.
    + eapply E_Seq. apply IHids. unfold assigned. simpl.
      eapply ceval_extensionality. apply E_Ass. simpl. reflexivity.
      intros V0. unfold t_update. compare V V0.
      × rewrite ← Heqlookup. rewrite ← eqb_string_refl. reflexivity.
      × rewrite false_eqb_string; simpl; congruence.
    + eapply ceval_extensionality. apply IHids.
      unfold assigned. intros V0. simpl. compare V V0.
      × rewrite ← Heqlookup.
        rewrite ← eqb_string_refl.
        destruct (inbP _ _ eqb_stringP V ids); reflexivity.
      × rewrite false_eqb_string; simpl; congruence.
Qed.
```

22.2.3 The Partial Evaluation Relation

At long last, we can define a partial evaluator for commands without loops, as an inductive relation! The inequality conditions in PE_AssDynamic and PE_If are just to keep the partial evaluator deterministic; they are not required for correctness.

```
Reserved Notation "c1 '/' st '\\' c1' '/' st'"
  (at level 40, st at level 39, c1' at level 39).

Inductive pe_com : com → pe_state → com → pe_state → Prop :=
  | PE_Skip : ∀ pe_st,
      SKIP / pe_st \\ SKIP / pe_st
  | PE_AssStatic : ∀ pe_st a1 n1 l,
      pe_aexp pe_st a1 = ANum n1 →
      (l ::= a1) / pe_st \\ SKIP / pe_add pe_st l n1
  | PE_AssDynamic : ∀ pe_st a1 a1' l,
      pe_aexp pe_st a1 = a1' →
      (∀ n, a1' ≠ ANum n) →
      (l ::= a1) / pe_st \\ (l ::= a1') / pe_remove pe_st l
  | PE_Seq : ∀ pe_st pe_st' pe_st'' c1 c2 c1' c2',
      c1 / pe_st \\ c1' / pe_st' →
      c2 / pe_st' \\ c2' / pe_st'' →
      (c1 ;; c2) / pe_st \\ (c1' ;; c2') / pe_st''
  | PE_IfTrue : ∀ pe_st pe_st' b1 c1 c2 c1',
      pe_bexp pe_st b1 = BTrue →
      c1 / pe_st \\ c1' / pe_st' →
      (TEST b1 THEN c1 ELSE c2 FI) / pe_st \\ c1' / pe_st'
  | PE_IfFalse : ∀ pe_st pe_st' b1 c1 c2 c2',
      pe_bexp pe_st b1 = BFalse →
      c2 / pe_st \\ c2' / pe_st' →
      (TEST b1 THEN c1 ELSE c2 FI) / pe_st \\ c2' / pe_st'
  | PE_If : ∀ pe_st pe_st1 pe_st2 b1 c1 c2 c1' c2',
      pe_bexp pe_st b1 ≠ BTrue →
      pe_bexp pe_st b1 ≠ BFalse →
      c1 / pe_st \\ c1' / pe_st1 →
      c2 / pe_st \\ c2' / pe_st2 →
      (TEST b1 THEN c1 ELSE c2 FI) / pe_st
        \\ (TEST pe_bexp pe_st b1
             THEN c1' ;; assign pe_st1 (pe_compare pe_st1 pe_st2)
             ELSE c2' ;; assign pe_st2 (pe_compare pe_st1 pe_st2) FI)
            / pe_removes pe_st1 (pe_compare pe_st1 pe_st2)

  where "c1 '/' st '\\' c1' '/' st'" := (pe_com c1 st c1' st').

Hint Constructors pe_com.
```

```
Hint Constructors ceval.
```

22.2.4 Examples

Below are some examples of using the partial evaluator. To make the **pe_com** relation actually usable for automatic partial evaluation, we would need to define more automation tactics in Coq. That is not hard to do, but it is not needed here.

```
Example pe_example1:
  (X ::= 3 ;; Y ::= Z × (X + X))%imp
  / [] \\ (SKIP;; Y ::= Z × 6)%imp / [(X,3)].
Proof. eapply PE_Seq. eapply PE_AssStatic. reflexivity.
  eapply PE_AssDynamic. reflexivity. intros n H. inversion H. Qed.

Example pe_example2:
  (X ::= 3 ;; TEST X ≤ 4 THEN X ::= 4 ELSE SKIP FI)%imp
  / [] \\ (SKIP;; SKIP)%imp / [(X,4)].
Proof. eapply PE_Seq. eapply PE_AssStatic. reflexivity.
  eapply PE_IfTrue. reflexivity.
  eapply PE_AssStatic. reflexivity. Qed.

Example pe_example3:
  (X ::= 3;;
   TEST Y ≤ 4 THEN
     Y ::= 4;;
     TEST X = Y THEN Y ::= 999 ELSE SKIP FI
   ELSE SKIP FI)%imp / []
  \\ (SKIP;;
       TEST Y ≤ 4 THEN
         (SKIP;; SKIP);; (SKIP;; Y ::= 4)
       ELSE SKIP;; SKIP FI)%imp
      / [(X,3)].
Proof. erewrite f_equal2 with (f := fun c st ⇒ _ / _ \\ c / st).
  eapply PE_Seq. eapply PE_AssStatic. reflexivity.
  eapply PE_If; intuition eauto; try solve_by_invert.
  econstructor. eapply PE_AssStatic. reflexivity.
  eapply PE_IfFalse. reflexivity. econstructor.
  reflexivity. reflexivity. Qed.
```

22.2.5 Correctness of Partial Evaluation

Finally let's prove that this partial evaluator is correct!

```
Reserved Notation "c' '/' pe_st' '/' st '\\' st''"
  (at level 40, pe_st' at level 39, st at level 39).
```

```
Inductive pe_ceval
  (c':com) (pe_st':pe_state) (st:state) (st'':state) : Prop :=
  | pe_ceval_intro : ∀ st',
    st =[ c' ]=> st' →
    pe_update st' pe_st' = st'' →
    c' / pe_st' / st \\ st''
  where "c' '/' pe_st' '/' st '\\' st''" := (pe_ceval c' pe_st' st st'').

Hint Constructors pe_ceval.

Theorem pe_com_complete:
  ∀ c pe_st pe_st' c', c / pe_st \\ c' / pe_st' →
  ∀ st st'',
  (pe_update st pe_st =[ c ]=> st'') →
  (c' / pe_st' / st \\ st'').
Proof. intros c pe_st pe_st' c' Hpe.
  induction Hpe; intros st st'' Heval;
  try (inversion Heval; subst;
       try (rewrite → pe_bexp_correct, → H in *; solve_by_invert);
       []);
  eauto.
  - econstructor. econstructor.
    rewrite → pe_aexp_correct. rewrite ← pe_update_update_add.
    rewrite → H. reflexivity.
  - econstructor. econstructor. reflexivity.
    rewrite → pe_aexp_correct. rewrite ← pe_update_update_remove.
    reflexivity.
  -
    edestruct IHHpe1. eassumption. subst.
    edestruct IHHpe2. eassumption.
    eauto.
  - inversion Heval; subst.
    + edestruct IHHpe1. eassumption.
      econstructor. apply E_IfTrue. rewrite ← pe_bexp_correct. assumption.
      eapply E_Seq. eassumption. apply eval_assign.
      rewrite ← assign_removes. eassumption.
    + edestruct IHHpe2. eassumption.
      econstructor. apply E_IfFalse. rewrite ← pe_bexp_correct. assumption.
      eapply E_Seq. eassumption. apply eval_assign.
      rewrite → pe_compare_update.
      rewrite ← assign_removes. eassumption.
Qed.

Theorem pe_com_sound:
  ∀ c pe_st pe_st' c', c / pe_st \\ c' / pe_st' →
```

```
  ∀ st st'',
  (c' / pe_st' / st \\ st'') →
  (pe_update st pe_st =[ c ]=> st'').
Proof. intros c pe_st pe_st' c' Hpe.
  induction Hpe;
    intros st st'' [st' Heval Heq];
    try (inversion Heval; []; subst); auto.
  - rewrite ← pe_update_update_add. apply E_Ass.
    rewrite → pe_aexp_correct. rewrite → H. reflexivity.
  - rewrite ← pe_update_update_remove. apply E_Ass.
    rewrite ← pe_aexp_correct. reflexivity.
  - eapply E_Seq; eauto.
  - apply E_IfTrue.
    rewrite → pe_bexp_correct. rewrite → H. reflexivity. eauto.
  - apply E_IfFalse.
    rewrite → pe_bexp_correct. rewrite → H. reflexivity. eauto.
  -
    inversion Heval; subst; inversion H7;
      (eapply ceval_deterministic in H8; [| apply eval_assign]); subst.
    +
      apply E_IfTrue. rewrite → pe_bexp_correct. assumption.
      rewrite ← assign_removes. eauto.
    +
      rewrite → pe_compare_update.
      apply E_IfFalse. rewrite → pe_bexp_correct. assumption.
      rewrite ← assign_removes. eauto.
Qed.
```

The main theorem. Thanks to David Menendez for this formulation!

```
Corollary pe_com_correct:
  ∀ c pe_st pe_st' c', c / pe_st \\ c' / pe_st' →
  ∀ st st'',
  (pe_update st pe_st =[ c ]=> st'') ↔
  (c' / pe_st' / st \\ st'').
Proof. intros c pe_st pe_st' c' H st st''. split.
  - apply pe_com_complete. apply H.
  - apply pe_com_sound. apply H.
Qed.
```

22.3 Partial Evaluation of Loops

It may seem straightforward at first glance to extend the partial evaluation relation **pe_com** above to loops. Indeed, many loops are easy to deal with. Considered this repeated-squaring loop, for example:

WHILE 1 <= X DO Y ::= Y * Y;; X ::= X - 1 END

If we know neither X nor Y statically, then the entire loop is dynamic and the residual command should be the same. If we know X but not Y, then the loop can be unrolled all the way and the residual command should be, for example,

Y ::= Y * Y;; Y ::= Y * Y;; Y ::= Y * Y

if X is initially 3 (and finally 0). In general, a loop is easy to partially evaluate if the final partial state of the loop body is equal to the initial state, or if its guard condition is static.

But there are other loops for which it is hard to express the residual program we want in Imp. For example, take this program for checking whether Y is even or odd:

X ::= 0;; WHILE 1 <= Y DO Y ::= Y - 1 ;; X ::= 1 - X END

The value of X alternates between 0 and 1 during the loop. Ideally, we would like to unroll this loop, not all the way but *two-fold*, into something like

WHILE 1 <= Y DO Y ::= Y - 1;; IF 1 <= Y THEN Y ::= Y - 1 ELSE X ::= 1;; EXIT FI END;; X ::= 0

Unfortunately, there is no *EXIT* command in Imp. Without extending the range of control structures available in our language, the best we can do is to repeat loop-guard tests or add flag variables. Neither option is terribly attractive.

Still, as a digression, below is an attempt at performing partial evaluation on Imp commands. We add one more command argument *c''* to the **pe_com** relation, which keeps track of a loop to roll up.

```
Module Loop.

Reserved Notation "c1 '/' st '\\' c1' '/' st' '/' c''"
  (at level 40, st at level 39, c1' at level 39, st' at level 39).

Inductive pe_com : com → pe_state → com → pe_state → com → Prop :=
  | PE_Skip : ∀ pe_st,
      SKIP / pe_st \\ SKIP / pe_st / SKIP
  | PE_AssStatic : ∀ pe_st a1 n1 l,
      pe_aexp pe_st a1 = ANum n1 →
      (l ::= a1) / pe_st \\ SKIP / pe_add pe_st l n1 / SKIP
  | PE_AssDynamic : ∀ pe_st a1 a1' l,
      pe_aexp pe_st a1 = a1' →
      (∀ n, a1' ≠ ANum n) →
      (l ::= a1) / pe_st \\ (l ::= a1') / pe_remove pe_st l / SKIP
  | PE_Seq : ∀ pe_st pe_st' pe_st'' c1 c2 c1' c2' c'',
      c1 / pe_st \\ c1' / pe_st' / SKIP →
      c2 / pe_st' \\ c2' / pe_st'' / c'' →
      (c1 ;; c2) / pe_st \\ (c1' ;; c2') / pe_st'' / c''
```

```
| PE_IfTrue : ∀ pe_st pe_st' b1 c1 c2 c1' c'',
    pe_bexp pe_st b1 = BTrue →
    c1 / pe_st \\ c1' / pe_st' / c'' →
    (TEST b1 THEN c1 ELSE c2 FI) / pe_st \\ c1' / pe_st' / c''
| PE_IfFalse : ∀ pe_st pe_st' b1 c1 c2 c2' c'',
    pe_bexp pe_st b1 = BFalse →
    c2 / pe_st \\ c2' / pe_st' / c'' →
    (TEST b1 THEN c1 ELSE c2 FI) / pe_st \\ c2' / pe_st' / c''
| PE_If : ∀ pe_st pe_st1 pe_st2 b1 c1 c2 c1' c2' c'',
    pe_bexp pe_st b1 ≠ BTrue →
    pe_bexp pe_st b1 ≠ BFalse →
    c1 / pe_st \\ c1' / pe_st1 / c'' →
    c2 / pe_st \\ c2' / pe_st2 / c'' →
    (TEST b1 THEN c1 ELSE c2 FI) / pe_st
      \\ (TEST pe_bexp pe_st b1
           THEN c1' ;; assign pe_st1 (pe_compare pe_st1 pe_st2)
           ELSE c2' ;; assign pe_st2 (pe_compare pe_st1 pe_st2) FI)
          / pe_removes pe_st1 (pe_compare pe_st1 pe_st2)
          / c''
| PE_WhileFalse : ∀ pe_st b1 c1,
    pe_bexp pe_st b1 = BFalse →
    (WHILE b1 DO c1 END) / pe_st \\ SKIP / pe_st / SKIP
| PE_WhileTrue : ∀ pe_st pe_st' pe_st'' b1 c1 c1' c2' c2'',
    pe_bexp pe_st b1 = BTrue →
    c1 / pe_st \\ c1' / pe_st' / SKIP →
    (WHILE b1 DO c1 END) / pe_st' \\ c2' / pe_st'' / c2'' →
    pe_compare pe_st pe_st'' ≠ [] →
    (WHILE b1 DO c1 END) / pe_st \\ (c1';;c2') / pe_st'' / c2''
| PE_While : ∀ pe_st pe_st' pe_st'' b1 c1 c1' c2' c2'',
    pe_bexp pe_st b1 ≠ BFalse →
    pe_bexp pe_st b1 ≠ BTrue →
    c1 / pe_st \\ c1' / pe_st' / SKIP →
    (WHILE b1 DO c1 END) / pe_st' \\ c2' / pe_st'' / c2'' →
    pe_compare pe_st pe_st'' ≠ [] →
    (c2'' = SKIP%imp ∨ c2'' = WHILE b1 DO c1 END%imp) →
    (WHILE b1 DO c1 END) / pe_st
      \\ (TEST pe_bexp pe_st b1
           THEN c1';; c2';; assign pe_st'' (pe_compare pe_st pe_st'')
           ELSE assign pe_st (pe_compare pe_st pe_st'') FI)%imp
          / pe_removes pe_st (pe_compare pe_st pe_st'')
          / c2''
| PE_WhileFixedEnd : ∀ pe_st b1 c1,
```

```
      pe_bexp pe_st b1 ≠ BFalse →
      (WHILE b1 DO c1 END) / pe_st \\ SKIP / pe_st / (WHILE b1 DO c1 END)
  | PE_WhileFixedLoop : ∀ pe_st pe_st' pe_st'' b1 c1 c1' c2',
      pe_bexp pe_st b1 = BTrue →
      c1 / pe_st \\ c1' / pe_st' / SKIP →
      (WHILE b1 DO c1 END) / pe_st'
        \\ c2' / pe_st'' / (WHILE b1 DO c1 END) →
      pe_compare pe_st pe_st'' = [] →
      (WHILE b1 DO c1 END) / pe_st
        \\ (WHILE BTrue DO SKIP END) / pe_st / SKIP

  | PE_WhileFixed : ∀ pe_st pe_st' pe_st'' b1 c1 c1' c2',
      pe_bexp pe_st b1 ≠ BFalse →
      pe_bexp pe_st b1 ≠ BTrue →
      c1 / pe_st \\ c1' / pe_st' / SKIP →
      (WHILE b1 DO c1 END) / pe_st'
        \\ c2' / pe_st'' / (WHILE b1 DO c1 END) →
      pe_compare pe_st pe_st'' = [] →
      (WHILE b1 DO c1 END) / pe_st
        \\ (WHILE pe_bexp pe_st b1 DO c1';; c2' END) / pe_st / SKIP

  where "c1 '/' st '\\' c1' '/' st' '/' c''" := (pe_com c1 st c1' st' c'').
Hint Constructors pe_com.
```

22.3.1 Examples

```
Ltac step i :=
  (eapply i; intuition eauto; try solve_by_invert);
  repeat (try eapply PE_Seq;
          try (eapply PE_AssStatic; simpl; reflexivity);
          try (eapply PE_AssDynamic;
                [ simpl; reflexivity
                | intuition eauto; solve_by_invert])).
Definition square_loop: com :=
  (WHILE 1 ≤ X DO
    Y ::= Y × Y;;
    X ::= X - 1
  END)%imp.
Example pe_loop_example1:
  square_loop / []
  \\ (WHILE 1 ≤ X DO
         (Y ::= Y × Y;;
```

```
        X ::= X - 1);; SKIP
      END)%imp / [] / SKIP.
Proof. erewrite f_equal2 with (f := fun c st ⇒ _ / _ \\ c / st / SKIP).
  step PE_WhileFixed. step PE_WhileFixedEnd. reflexivity.
  reflexivity. reflexivity. Qed.

Example pe_loop_example2:
  (X ::= 3;; square_loop)%imp / []
  \\ (SKIP;;
      (Y ::= Y × Y;; SKIP);;
      (Y ::= Y × Y;; SKIP);;
      (Y ::= Y × Y;; SKIP);;
      SKIP)%imp / [(X,0)] / SKIP%imp.
Proof. erewrite f_equal2 with (f := fun c st ⇒ _ / _ \\ c / st / SKIP).
  eapply PE_Seq. eapply PE_AssStatic. reflexivity.
  step PE_WhileTrue.
  step PE_WhileTrue.
  step PE_WhileTrue.
  step PE_WhileFalse.
  inversion H. inversion H. inversion H.
  reflexivity. reflexivity. Qed.

Example pe_loop_example3:
  (Z ::= 3;; subtract_slowly) / []
  \\ (SKIP;;
      TEST ~(X = 0) THEN
        (SKIP;; X ::= X - 1);;
        TEST ~(X = 0) THEN
          (SKIP;; X ::= X - 1);;
          TEST ~(X = 0) THEN
            (SKIP;; X ::= X - 1);;
            WHILE ~(X = 0) DO
              (SKIP;; X ::= X - 1);; SKIP
            END;;
            SKIP;; Z ::= 0
          ELSE SKIP;; Z ::= 1 FI;; SKIP
        ELSE SKIP;; Z ::= 2 FI;; SKIP
      ELSE SKIP;; Z ::= 3 FI)%imp / [] / SKIP.
Proof. erewrite f_equal2 with (f := fun c st ⇒ _ / _ \\ c / st / SKIP).
  eapply PE_Seq. eapply PE_AssStatic. reflexivity.
  step PE_While.
  step PE_While.
  step PE_While.
  step PE_WhileFixed.
```

```
  step PE_WhileFixedEnd.
  reflexivity. inversion H. inversion H. inversion H.
  reflexivity. reflexivity. Qed.
Example pe_loop_example4:
  (X ::= 0;;
   WHILE X ≤ 2 DO
     X ::= 1 - X
   END)%imp / [] \\ (SKIP;; WHILE true DO SKIP END)%imp / [(X,0)] / SKIP.
Proof. erewrite f_equal2 with (f := fun c st ⇒ _ / _ \\ c / st / SKIP).
  eapply PE_Seq. eapply PE_AssStatic. reflexivity.
  step PE_WhileFixedLoop.
  step PE_WhileTrue.
  step PE_WhileFixedEnd.
  inversion H. reflexivity. reflexivity. reflexivity. Qed.
```

22.3.2 Correctness

Because this partial evaluator can unroll a loop n-fold where n is a (finite) integer greater than one, in order to show it correct we need to perform induction not structurally on dynamic evaluation but on the number of times dynamic evaluation enters a loop body.

```
Reserved Notation "c1 '/' st '\\' st' '#' n"
  (at level 40, st at level 39, st' at level 39).
Inductive ceval_count : com → state → state → nat → Prop :=
  | E'Skip : ∀ st,
      SKIP / st \\ st # 0
  | E'Ass : ∀ st a1 n l,
      aeval st a1 = n →
      (l ::= a1) / st \\ (t_update st l n) # 0
  | E'Seq : ∀ c1 c2 st st' st'' n1 n2,
      c1 / st \\ st' # n1 →
      c2 / st' \\ st'' # n2 →
      (c1 ;; c2) / st \\ st'' # (n1 + n2)
  | E'IfTrue : ∀ st st' b1 c1 c2 n,
      beval st b1 = true →
      c1 / st \\ st' # n →
      (TEST b1 THEN c1 ELSE c2 FI) / st \\ st' # n
  | E'IfFalse : ∀ st st' b1 c1 c2 n,
      beval st b1 = false →
      c2 / st \\ st' # n →
      (TEST b1 THEN c1 ELSE c2 FI) / st \\ st' # n
  | E'WhileFalse : ∀ b1 st c1,
      beval st b1 = false →
```

```
      (WHILE b1 DO c1 END) / st \\ st # 0
  | E'WhileTrue : ∀ st st' st'' b1 c1 n1 n2,
      beval st b1 = true →
      c1 / st \\ st' # n1 →
      (WHILE b1 DO c1 END) / st' \\ st'' # n2 →
      (WHILE b1 DO c1 END) / st \\ st'' # S (n1 + n2)

  where "c1 '/' st '\\' st' # n" := (ceval_count c1 st st' n).
Hint Constructors ceval_count.
Theorem ceval_count_complete: ∀ c st st',
  st =[ c ]=> st' → ∃ n, c / st \\ st' # n.
Proof. intros c st st' Heval.
  induction Heval;
    try inversion IHHeval1;
    try inversion IHHeval2;
    try inversion IHHeval;
    eauto. Qed.
Theorem ceval_count_sound: ∀ c st st' n,
  c / st \\ st' # n → st =[ c ]=> st'.
Proof. intros c st st' n Heval. induction Heval; eauto. Qed.
Theorem pe_compare_nil_lookup: ∀ pe_st1 pe_st2,
  pe_compare pe_st1 pe_st2 = [] →
  ∀ V, pe_lookup pe_st1 V = pe_lookup pe_st2 V.
Proof. intros pe_st1 pe_st2 H V.
  apply (pe_compare_correct pe_st1 pe_st2 V).
  rewrite H. intro. inversion H0. Qed.
Theorem pe_compare_nil_update: ∀ pe_st1 pe_st2,
  pe_compare pe_st1 pe_st2 = [] →
  ∀ st, pe_update st pe_st1 = pe_update st pe_st2.
Proof. intros pe_st1 pe_st2 H st.
  apply functional_extensionality. intros V.
  rewrite !pe_update_correct.
  apply pe_compare_nil_lookup with (V:=V) in H.
  rewrite H. reflexivity. Qed.
Reserved Notation "c' '/' pe_st' '/' c'' '/' st '\\' st'' '#' n"
  (at level 40, pe_st' at level 39, c'' at level 39,
   st at level 39, st'' at level 39).
Inductive pe_ceval_count (c':com) (pe_st':pe_state) (c'':com)
                              (st:state) (st'':state) (n:nat) : Prop :=
  | pe_ceval_count_intro : ∀ st' n',
    st =[ c' ]=> st' →
```

```
    c'' / pe_update st' pe_st' \\ st'' # n' →
    n' ≤ n →
    c' / pe_st' / c'' / st \\ st'' # n
  where "c' '/' pe_st' '/' c'' '/' st '\\' st'' '#' n" :=
        (pe_ceval_count c' pe_st' c'' st st'' n).

Hint Constructors pe_ceval_count.

Lemma pe_ceval_count_le: ∀ c' pe_st' c'' st st'' n n',
  n' ≤ n →
  c' / pe_st' / c'' / st \\ st'' # n' →
  c' / pe_st' / c'' / st \\ st'' # n.
Proof. intros c' pe_st' c'' st st'' n n' Hle H. inversion H.
  econstructor; try eassumption. omega. Qed.

Theorem pe_com_complete:
  ∀ c pe_st pe_st' c' c'', c / pe_st \\ c' / pe_st' / c'' →
  ∀ st st'' n,
  (c / pe_update st pe_st \\ st'' # n) →
  (c' / pe_st' / c'' / st \\ st'' # n).
Proof. intros c pe_st pe_st' c' c'' Hpe.
  induction Hpe; intros st st'' n Heval;
  try (inversion Heval; subst;
        try (rewrite → pe_bexp_correct, → H in *; solve_by_invert);
        []);
  eauto.
  - econstructor. econstructor.
    rewrite → pe_aexp_correct. rewrite ← pe_update_update_add.
    rewrite → H. apply E'Skip. auto.
  - econstructor. econstructor. reflexivity.
    rewrite → pe_aexp_correct. rewrite ← pe_update_update_remove.
    apply E'Skip. auto.
  -
    edestruct IHHpe1 as [? ? ? Hskip ?]. eassumption.
    inversion Hskip. subst.
    edestruct IHHpe2. eassumption.
    econstructor; eauto. omega.
  - inversion Heval; subst.
    + edestruct IHHpe1. eassumption.
      econstructor. apply E_IfTrue. rewrite ← pe_bexp_correct. assumption.
      eapply E_Seq. eassumption. apply eval_assign.
      rewrite ← assign_removes. eassumption. eassumption.
    + edestruct IHHpe2. eassumption.
      econstructor. apply E_IfFalse. rewrite ← pe_bexp_correct. assumption.
      eapply E_Seq. eassumption. apply eval_assign.
```

```
    rewrite -> pe_compare_update.
    rewrite <- assign_removes. eassumption. eassumption.
-
  edestruct IHHpe1 as [? ? ? Hskip ?]. eassumption.
  inversion Hskip. subst.
  edestruct IHHpe2. eassumption.
  econstructor; eauto. omega.
- inversion Heval; subst.
  + econstructor. apply E_IfFalse.
    rewrite <- pe_bexp_correct. assumption.
    apply eval_assign.
    rewrite <- assign_removes. inversion H2; subst; auto.
    auto.
  +
    edestruct IHHpe1 as [? ? ? Hskip ?]. eassumption.
    inversion Hskip. subst.
    edestruct IHHpe2. eassumption.
    econstructor. apply E_IfTrue.
    rewrite <- pe_bexp_correct. assumption.
    repeat eapply E_Seq; eauto. apply eval_assign.
    rewrite -> pe_compare_update, <- assign_removes. eassumption.
    omega.
- exfalso.
  generalize dependent (S (n1 + n2)). intros n.
  clear - H H0 IHHpe1 IHHpe2. generalize dependent st.
  induction n using lt_wf_ind; intros st Heval. inversion Heval; subst.
  + rewrite pe_bexp_correct, H in H7. inversion H7.
  +
    edestruct IHHpe1 as [? ? ? Hskip ?]. eassumption.
    inversion Hskip. subst.
    edestruct IHHpe2. eassumption.
    rewrite <- (pe_compare_nil_update _ _ H0) in H7.
    apply H1 in H7; [| omega]. inversion H7.
- generalize dependent st.
  induction n using lt_wf_ind; intros st Heval. inversion Heval; subst.
  + rewrite pe_bexp_correct in H8. eauto.
  + rewrite pe_bexp_correct in H5.
    edestruct IHHpe1 as [? ? ? Hskip ?]. eassumption.
    inversion Hskip. subst.
    edestruct IHHpe2. eassumption.
    rewrite <- (pe_compare_nil_update _ _ H1) in H8.
    apply H2 in H8; [| omega]. inversion H8.
```

```
      econstructor; [ eapply E_WhileTrue; eauto | eassumption | omega].
Qed.

Theorem pe_com_sound:
  ∀ c pe_st pe_st' c' c'', c / pe_st \\ c' / pe_st' / c'' →
  ∀ st st'' n,
  (c' / pe_st' / c'' / st \\ st'' # n) →
  (pe_update st pe_st =[ c ]=> st'').
Proof. intros c pe_st pe_st' c' c'' Hpe.
  induction Hpe;
    intros st st'' n [st' n' Heval Heval' Hle];
    try (inversion Heval; []; subst);
    try (inversion Heval'; []; subst); eauto.
  - rewrite ← pe_update_update_add. apply E_Ass.
    rewrite → pe_aexp_correct. rewrite → H. reflexivity.
  - rewrite ← pe_update_update_remove. apply E_Ass.
    rewrite ← pe_aexp_correct. reflexivity.
  - eapply E_Seq; eauto.
  - apply E_IfTrue.
    rewrite → pe_bexp_correct. rewrite → H. reflexivity.
    eapply IHHpe. eauto.
  - apply E_IfFalse.
    rewrite → pe_bexp_correct. rewrite → H. reflexivity.
    eapply IHHpe. eauto.
  - inversion Heval; subst; inversion H7; subst; clear H7.
    +
      eapply ceval_deterministic in H8; [| apply eval_assign]. subst.
      rewrite ← assign_removes in Heval'.
      apply E_IfTrue. rewrite → pe_bexp_correct. assumption.
      eapply IHHpe1. eauto.
    +
      eapply ceval_deterministic in H8; [| apply eval_assign]. subst.
      rewrite → pe_compare_update in Heval'.
      rewrite ← assign_removes in Heval'.
      apply E_IfFalse. rewrite → pe_bexp_correct. assumption.
      eapply IHHpe2. eauto.
  - apply E_WhileFalse.
    rewrite → pe_bexp_correct. rewrite → H. reflexivity.
  - eapply E_WhileTrue.
    rewrite → pe_bexp_correct. rewrite → H. reflexivity.
    eapply IHHpe1. eauto. eapply IHHpe2. eauto.
  - inversion Heval; subst.
    +
```

```
      inversion H9. subst. clear H9.
      inversion H10. subst. clear H10.
      eapply ceval_deterministic in H11; [| apply eval_assign]. subst.
      rewrite → pe_compare_update in Heval'.
      rewrite ← assign_removes in Heval'.
      eapply E_WhileTrue. rewrite → pe_bexp_correct. assumption.
      eapply IHHpe1. eauto.
      eapply IHHpe2. eauto.
    + apply ceval_count_sound in Heval'.
      eapply ceval_deterministic in H9; [| apply eval_assign]. subst.
      rewrite ← assign_removes in Heval'.
      inversion H2; subst.
      × inversion Heval'. subst. apply E_WhileFalse.
        rewrite → pe_bexp_correct. assumption.
      × assumption.
  - eapply ceval_count_sound. apply Heval'.
  -
    apply loop_never_stops in Heval. inversion Heval.
  -
    clear - H1 IHHpe1 IHHpe2 Heval.
    remember (WHILE pe_bexp pe_st b1 DO c1';; c2' END)%imp as c'.
    induction Heval;
      inversion Heqc'; subst; clear Heqc'.
    + apply E_WhileFalse.
      rewrite pe_bexp_correct. assumption.
    +
      assert (IHHeval2' := IHHeval2 (refl_equal _)).
      apply ceval_count_complete in IHHeval2'. inversion IHHeval2'.
      clear IHHeval1 IHHeval2 IHHeval2'.
      inversion Heval1. subst.
      eapply E_WhileTrue. rewrite pe_bexp_correct. assumption. eauto.
      eapply IHHpe2. econstructor. eassumption.
      rewrite ← (pe_compare_nil_update _ _ H1). eassumption. apply le_n.
Qed.

Corollary pe_com_correct:
  ∀ c pe_st pe_st' c', c / pe_st \\ c' / pe_st' / SKIP →
  ∀ st st'',
  (pe_update st pe_st =[ c ]=> st'') ↔
  (∃ st', st =[ c' ]=> st' ∧ pe_update st' pe_st' = st'').
Proof. intros c pe_st pe_st' c' H st st''. split.
  - intros Heval.
    apply ceval_count_complete in Heval. inversion Heval as [n Heval'].
```

```
    apply pe_com_complete with (st:=st) (st'':=st'') (n:=n) in H.
    inversion H as [? ? ? Hskip ?]. inversion Hskip. subst. eauto.
    assumption.
  - intros [st' [Heval Heq]]. subst st''.
    eapply pe_com_sound in H. apply H.
    econstructor. apply Heval. apply E'Skip. apply le_n.
Qed.
```

End LOOP.

22.4 Partial Evaluation of Flowchart Programs

Instead of partially evaluating *WHILE* loops directly, the standard approach to partially evaluating imperative programs is to convert them into *flowcharts*. In other words, it turns out that adding labels and jumps to our language makes it much easier to partially evaluate. The result of partially evaluating a flowchart is a residual flowchart. If we are lucky, the jumps in the residual flowchart can be converted back to *WHILE* loops, but that is not possible in general; we do not pursue it here.

22.4.1 Basic blocks

A flowchart is made of *basic blocks*, which we represent with the inductive type **block**. A basic block is a sequence of assignments (the constructor Assign), concluding with a conditional jump (the constructor If) or an unconditional jump (the constructor Goto). The destinations of the jumps are specified by *labels*, which can be of any type. Therefore, we parameterize the **block** type by the type of labels.

```
Inductive block (Label:Type) : Type :=
  | Goto : Label → block Label
  | If : bexp → Label → Label → block Label
  | Assign : string → aexp → block Label → block Label.
Arguments Goto {Label} _.
Arguments If {Label} _ _ _.
Arguments Assign {Label} _ _ _.
```

We use the "even or odd" program, expressed above in Imp, as our running example. Converting this program into a flowchart turns out to require 4 labels, so we define the following type.

```
Inductive parity_label : Type :=
  | entry : parity_label
  | loop : parity_label
  | body : parity_label
  | done : parity_label.
```

The following **block** is the basic block found at the body label of the example program.

```
Definition parity_body : block parity_label :=
  Assign Y (Y - 1)
   (Assign X (1 - X)
     (Goto loop)).
```

To evaluate a basic block, given an initial state, is to compute the final state and the label to jump to next. Because basic blocks do not *contain* loops or other control structures, evaluation of basic blocks is a total function – we don't need to worry about non-termination.

```
Fixpoint keval {L:Type} (st:state) (k : block L) : state × L :=
  match k with
  | Goto l ⇒ (st , l)
  | If b l1 l2 ⇒ (st , if beval st b then l1 else l2)
  | Assign i a k ⇒ keval (t_update st i (aeval st a)) k
  end.
```

```
Example keval_example:
  keval empty_st parity_body
  = ((X !-> 1 ; Y !-> 0) , loop).
Proof. reflexivity. Qed.
```

22.4.2 Flowchart programs

A flowchart program is simply a lookup function that maps labels to basic blocks. Actually, some labels are *halting states* and do not map to any basic block. So, more precisely, a flowchart program whose labels are of type L is a function from L to **option** (**block** L).

```
Definition program (L:Type) : Type := L → option (block L).
```

```
Definition parity : program parity_label := fun l ⇒
  match l with
  | entry ⇒ Some (Assign X 0 (Goto loop))
  | loop ⇒ Some (If (1 ≤ Y) body done)
  | body ⇒ Some parity_body
  | done ⇒ None
  end.
```

Unlike a basic block, a program may not terminate, so we model the evaluation of programs by an inductive relation **peval** rather than a recursive function.

```
Inductive peval {L:Type} (p : program L)
  : state → L → state → L → Prop :=
  | E_None: ∀ st l,
     p l = None →
     peval p st l st l
  | E_Some: ∀ st l k st' l' st'' l'',
```

```
    p l = Some k →
    keval st k = (st', l') →
    peval p st' l' st'' l'' →
    peval p st l st'' l''.
Example parity_eval: peval parity empty_st entry empty_st done.
Proof. erewrite f_equal with (f := fun st ⇒ peval _ _ _ st _).
  eapply E_Some. reflexivity. reflexivity.
  eapply E_Some. reflexivity. reflexivity.
  apply E_None. reflexivity.
  apply functional_extensionality. intros i. rewrite t_update_same; auto.
Qed.
```

22.4.3 Partial Evaluation of Basic Blocks and Flowchart Programs

Partial evaluation changes the label type in a systematic way: if the label type used to be L, it becomes pe_state × L. So the same label in the original program may be unfolded, or blown up, into multiple labels by being paired with different partial states. For example, the label loop in the parity program will become two labels: ([(X,0)], loop) and ([(X,1)], loop). This change of label type is reflected in the types of pe_block and pe_program defined presently.

```
Fixpoint pe_block {L:Type} (pe_st:pe_state) (k : block L)
  : block (pe_state × L) :=
  match k with
  | Goto l ⇒ Goto (pe_st, l)
  | If b l1 l2 ⇒
    match pe_bexp pe_st b with
    | BTrue ⇒ Goto (pe_st, l1)
    | BFalse ⇒ Goto (pe_st, l2)
    | b' ⇒ If b' (pe_st, l1) (pe_st, l2)
    end
  | Assign i a k ⇒
    match pe_aexp pe_st a with
    | ANum n ⇒ pe_block (pe_add pe_st i n) k
    | a' ⇒ Assign i a' (pe_block (pe_remove pe_st i) k)
    end
  end.
Example pe_block_example:
  pe_block [(X,0)] parity_body
  = Assign Y (Y - 1) (Goto ([(X,1)], loop)).
Proof. reflexivity. Qed.
Theorem pe_block_correct: ∀ (L:Type) st pe_st k st' pe_st' (l':L),
  keval st (pe_block pe_st k) = (st', (pe_st', l')) →
  keval (pe_update st pe_st) k = (pe_update st' pe_st', l').
```

```
Proof. intros. generalize dependent pe_st. generalize dependent st.
  induction k as [l | b l1 l2 | i a k];
    intros st pe_st H.
  - inversion H; reflexivity.
  -
    replace (keval st (pe_block pe_st (If b l1 l2)))
       with (keval st (If (pe_bexp pe_st b) (pe_st, l1) (pe_st, l2)))
       in H by (simpl; destruct (pe_bexp pe_st b); reflexivity).
    simpl in *. rewrite pe_bexp_correct.
    destruct (beval st (pe_bexp pe_st b)); inversion H; reflexivity.
  -
    simpl in *. rewrite pe_aexp_correct.
    destruct (pe_aexp pe_st a); simpl;
      try solve [rewrite pe_update_update_add; apply IHk; apply H];
      solve [rewrite pe_update_update_remove; apply IHk; apply H].
Qed.

Definition pe_program {L:Type} (p : program L)
  : program (pe_state × L) :=
  fun pe_l ⇒ match pe_l with | (pe_st, l) ⇒
                  option_map (pe_block pe_st) (p l)
                end.

Inductive pe_peval {L:Type} (p : program L)
  (st:state) (pe_st:pe_state) (l:L) (st'o:state) (l':L) : Prop :=
  | pe_peval_intro : ∀ st' pe_st',
    peval (pe_program p) st (pe_st, l) st' (pe_st', l') →
    pe_update st' pe_st' = st'o →
    pe_peval p st pe_st l st'o l'.

Theorem pe_program_correct:
  ∀ (L:Type) (p : program L) st pe_st l st'o l',
  peval p (pe_update st pe_st) l st'o l' ↔
  pe_peval p st pe_st l st'o l'.
Proof. intros.
  split.
  - intros Heval.
    remember (pe_update st pe_st) as sto.
    generalize dependent pe_st. generalize dependent st.
    induction Heval as
      [ sto l Hlookup | sto l k st'o l' st''o l'' Hlookup Hkeval Heval ];
      intros st pe_st Heqsto; subst sto.
    + eapply pe_peval_intro. apply E_None.
      simpl. rewrite Hlookup. reflexivity. reflexivity.
    +
```

```
      remember (keval st (pe_block pe_st k)) as x.
      destruct x as [st' [pe_st' l'_]].
      symmetry in Heqx. erewrite pe_block_correct in Hkeval by apply Heqx.
      inversion Hkeval. subst st'o l'_. clear Hkeval.
      edestruct IHHeval. reflexivity. subst st''o. clear IHHeval.
      eapply pe_peval_intro; [| reflexivity]. eapply E_Some; eauto.
      simpl. rewrite Hlookup. reflexivity.
  - intros [st' pe_st' Heval Heqst'o].
    remember (pe_st, l) as pe_st_l.
    remember (pe_st', l') as pe_st'_l'.
    generalize dependent pe_st. generalize dependent l.
    induction Heval as
      [ st [pe_st_ l_] Hlookup
      | st [pe_st_ l_] pe_k st' [pe_st'_ l'_] st'' [pe_st'' l'']
        Hlookup Hkeval Heval ];
      intros l pe_st Heqpe_st_l;
      inversion Heqpe_st_l; inversion Heqpe_st'_l'; repeat subst.
    + apply E_None. simpl in Hlookup.
      destruct (p l'); [ solve [ inversion Hlookup ] | reflexivity ].
    +
      simpl in Hlookup. remember (p l) as k.
      destruct k as [k|]; inversion Hlookup; subst.
      eapply E_Some; eauto. apply pe_block_correct. apply Hkeval.
Qed.
```

Chapter 23

Postscript

Congratulations: We've made it to the end!

23.1 Looking Back

We've covered a lot of ground. Here's a quick review of the whole trajectory we've followed, starting at the beginning of *Logical Foundations*...

- *Functional programming*:

  - "declarative" programming style (recursion over immutable data structures, rather than looping over mutable arrays or pointer structures)
  - higher-order functions
  - polymorphism

- *Logic*, the mathematical basis for software engineering:

  logic calculus

  - ————————- ˜ ——————————-

  software engineering mechanical/civil engineering

  - inductively defined sets and relations
  - inductive proofs
  - proof objects

- *Coq*, an industrial-strength proof assistant

  - functional core language

    - core tactics
    - automation

- *Foundations of programming languages*

    - notations and definitional techniques for precisely specifying
        - abstract syntax
        - operational semantics
            - big-step style
            - small-step style
        - type systems
    - program equivalence
    - Hoare logic
    - fundamental metatheory of type systems
        - progress and preservation
    - theory of subtyping

23.2 Looking Around

Large-scale applications of these core topics can be found everywhere, both in ongoing research projects and in real-world software systems. Here are a few recent examples involving formal, machine-checked verification of real-world software and hardware systems, to give a sense of what is being done today...

CompCert

CompCert is a fully verified optimizing compiler for almost all of the ISO C90 / ANSI C language, generating code for x86, ARM, and PowerPC processors. The whole of CompCert is is written in Gallina and extracted to an efficient OCaml program using Coq's extraction facilities.

"The CompCert project investigates the formal verification of realistic compilers usable for critical embedded software. Such verified compilers come with a mathematical, machine-checked proof that the generated executable code behaves exactly as prescribed by the semantics of the source program. By ruling out the possibility of compiler-introduced bugs, verified compilers strengthen the guarantees that can be obtained by applying formal methods to source programs."

In 2011, CompCert was included in a landmark study on fuzz-testing a large number of real-world C compilers using the CSmith tool. The CSmith authors wrote:

- The striking thing about our CompCert results is that the middle-end bugs we found in all other compilers are absent. As of early 2011, the under-development version of CompCert is the only compiler we have tested for which Csmith cannot find wrong-code errors. This is not for lack of trying: we have devoted about six CPU-years to the task. The apparent unbreakability of CompCert supports a strong argument that developing compiler optimizations within a proof framework, where safety checks are explicit and machine-checked, has tangible benefits for compiler users.

http://compcert.inria.fr

seL4

seL4 is a fully verified microkernel, considered to be the world's first OS kernel with an end-to-end proof of implementation correctness and security enforcement. It is implemented in C and ARM assembly and specified and verified using Isabelle. The code is available as open source.

"seL4 has been comprehensively formally verified: a rigorous process to prove mathematically that its executable code, as it runs on hardware, correctly implements the behaviour allowed by the specification, and no others. Furthermore, we have proved that the specification has the desired safety and security properties (integrity and confidentiality)... The verification was achieved at a cost that is significantly less than that of traditional high-assurance development approaches, while giving guarantees traditional approaches cannot provide."

https://sel4.systems.

CertiKOS

CertiKOS is a clean-slate, fully verified hypervisor, written in CompCert C and verified in Coq.

"The CertiKOS project aims to develop a novel and practical programming infrastructure for constructing large-scale certified system software. By combining recent advances in programming languages, operating systems, and formal methods, we hope to attack the following research questions: (1) what OS kernel structure can offer the best support for extensibility, security, and resilience? (2) which semantic models and program logics can best capture these abstractions? (3) what are the right programming languages and environments for developing such certified kernels? and (4) how to build automation facilities to make certified software development really scale?"

http://flint.cs.yale.edu/certikos/

Ironclad

Ironclad Apps is a collection of fully verified web applications, including a "notary" for securely signing statements, a password hasher, a multi-user trusted counter, and a differentially-private database.

The system is coded in the verification-oriented programming language Dafny and verified using Boogie, a verification tool based on Hoare logic.

"An Ironclad App lets a user securely transmit her data to a remote machine with the guarantee that every instruction executed on that machine adheres to a formal abstract specification of the app's behavior. This does more than eliminate implementation vulnerabilities such as buffer overflows, parsing errors, or data leaks; it tells the user exactly how the app will behave at all times. We provide these guarantees via complete, low-level software verification. We then use cryptography and secure hardware to enable secure channels from the verified software to remote users."

https://github.com/Microsoft/Ironclad/tree/master/ironclad-apps

Verdi

Verdi is a framework for implementing and formally verifying distributed systems.

"Verdi supports several different fault models ranging from idealistic to realistic. Verdi's verified system transformers (VSTs) encapsulate common fault tolerance techniques. Developers can verify an application in an idealized fault model, and then apply a VST to obtain an application that is guaranteed to have analogous properties in a more adversarial environment. Verdi is developed using the Coq proof assistant, and systems are extracted to OCaml for execution. Verdi systems, including a fault-tolerant key-value store, achieve comparable performance to unverified counterparts."

http://verdi.uwplse.org

DeepSpec

The Science of Deep Specification is an NSF "Expedition" project (running from 2016 to 2020) that focuses on the specification and verification of full functional correctness of both software and hardware. It also sponsors workshops and summer schools.

- Website: http://deepspec.org/
- Overview presentations:
  - http://deepspec.org/about/
  - https://www.youtube.com/watch?v=IPNdsnRWBkk

REMS

REMS is a european project on Rigorous Engineering of Mainstream Systems. It has produced detailed formal specifications of a wide range of critical real-world interfaces, protocols, and APIs, including the C language, the ELF linker format, the ARM, Power, MIPS, CHERI, and RISC-V instruction sets, the weak memory models of ARM and Power processors, and POSIX filesystems.

"The project is focussed on lightweight rigorous methods: precise specification (post hoc and during design) and testing against specifications, with full verification only in some cases. The project emphasises building useful (and reusable) semantics and tools. We are building accurate full-scale mathematical models of some of the key computational abstractions (processor architectures, programming languages, concurrent OS interfaces, and network protocols), studying how this can be done, and investigating how such models can be used for new verification research and in new systems and programming language research. Supporting all this, we are also working on new specification tools and their foundations."

http://www.cl.cam.ac.uk/~pes20/rems/

Others

There's much more. Other projects worth checking out include:

- Vellvm (formal specification and verification of LLVM optimization passes)
- Zach Tatlock's formally certified browser
- Tobias Nipkow's formalization of most of Java
- The CakeML verified ML compiler
- Greg Morrisett's formal specification of the x86 instruction set and the RockSalt Software Fault Isolation tool (a better, faster, more secure version of Google's Native Client)
- Ur/Web, a programming language for verified web applications embedded in Coq
- the Princeton Verified Software Toolchain

23.3 Looking Forward

Some good places to learn more...

- This book includes several optional chapters covering topics that you may find useful. Take a look at the table of contents and the chapter dependency diagram to find them.
- More on Hoare logic and program verification
  - The Formal Semantics of Programming Languages: An Introduction, by Glynn Winskel *Winskel* 1993 (in Bib.v).
  - Many practical verification tools, e.g. Microsoft's Boogie system, Java Extended Static Checking, etc.

- More on the foundations of programming languages:
  - Concrete Semantics with Isabelle/HOL, by Tobias Nipkow and Gerwin Klein *Nipkow* 2014 (in Bib.v)
  - Types and Programming Languages, by Benjamin C. Pierce *Pierce* 2002 (in Bib.v).
  - Practical Foundations for Programming Languages, by Robert Harper *Harper* 2016 (in Bib.v).
  - Foundations for Programming Languages, by John C. Mitchell *Mitchell* 1996 (in Bib.v).
- Iron Lambda (http://iron.ouroborus.net/) is a collection of Coq formalisations for functional languages of increasing complexity. It fills part of the gap between the end of the Software Foundations course and what appears in current research papers. The collection has at least Progress and Preservation theorems for a number of variants of STLC and the polymorphic lambda-calculus (System F).
- Finally, here are some of the main conferences on programming languages and formal verification:
  - Principles of Programming Langauges (POPL)
  - Programming Language Design and Implementation (PLDI)
  - International Conference on Functional Programming (ICFP)
  - Computer Aided Verification (CAV)
  - Interactive Theorem Proving (ITP)
  - Certified Programs and Proofs (CPP)
  - SPLASH/OOPSLA conferences
  - Principles in Practice workshop (PiP)
  - CoqPL workshop

Date

Chapter 24

Bib: Bibliography

24.1 Resources cited in this volume

Aydemir 2008 Brian Aydemir, Arthur Charguéraud, Benjamin C. Pierce, Randy Pollack, and Stephanie Weirich. Engineering Formal Metatheory. In ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL), San Francisco, California, pages 3-15. ACM, January 2008. http://www.cis.upenn.edu/~bcpierce/papers/binders.pdf

Harper 2015 Practical Foundations for Programming Languages, by Robert Harper. Cambridge University Press. Second edition, 2016. http://tinyurl.com/z82xwta

Mitchell 1996 Foundations for Programming Languages, by John C. Mitchell. MIT Press, 1996. http://tinyurl.com/zkosavw

Nipkow 2014 Concrete Semantics with Isabelle/HOL, by Tobias Nipkow and Gerwin Klein. Springer 2014. http://www.concrete-semantics.org

Pierce 2002 Types and Programming Languages, by Benjamin C. Pierce. MIT Press, 2002. http://tinyurl.com/gtnudmu

Pugh 1991 Pugh, William. "The Omega test: a fast and practical integer programming algorithm for dependence analysis." Proceedings of the 1991 ACM/IEEE conference on Supercomputing. ACM, 1991. http://dl.acm.org/citation.cfm?id=125848

Winskel 1993 The Formal Semantics of Programming Languages: An Introduction, by Glynn Winskel. MIT Press, 1993. http://tinyurl.com/j2k6ev7

Date