Understanding Recursion

sk, rob and dbtucker (modified for CS 536 by kfisler)

2002-09-20

1 Writing a Recursive Function

Can we write the factorial function in *AFunExp*? Well, we currently don't have multiplication, or a way of making choices in our *AFunExp* code. But those two are easy to address. Multiplication is really very similar to addition and subtraction; to make choices, let's add a simple construct called `if0` with the following BNF:

```
<AFunIfExp> ::= ... all the terms in <AFunExp>, using <AFunIfExp> instead ...
          | {* <AFunIfExp> <AFunIfExp>}
          | {if0 <AFunIfExp> <AFunIfExp> <AFunIfExp>}
```

where *AFunIfExp* names a language with **c**onditionals in addition to the features we have already studied. An `if0` evaluates its fi rst sub-expression. If this yields the value `0` it evaluates the second, otherwise it evaluates the third. For example,

```
  {if0 {+ 5 -5}
       1
       2}
```

evaluates to 1.

Given *AFunIfExp*, we're ready to write factorial:

```
  {let {fac {fun {n}
              {if0 n
                   1
                   {* n {fac {+ n -1}}}}}}
    {fac 5}}
```

What does this evaluate to? 120? No. Consider the following simpler expression, which you were asked to think about as a puzzle when we studied substitution:

```
  {let {x x} x}
```

In this program, the `x` in the named expression position of the `let` has no binding. Similarly, the environment of the function bound to `fac` above has no bound identifi ers. Therefore, only the environment of the fi rst invocation (the body of the `let`) has a binding for `fac`. When `fac` is applied to `5`, the interpreter evaluates the body of `fac` in the closure environment, which has no bound identifi ers, so interpretation stops on the intended recursive call to `fac` with an unbound identifi er error.

Before you continue reading, please pause for a moment, study the program carefully, write down the environments at each stage, step by hand through the interpreter, even run the program if you wish, to convince yourself that this error will occur. Understanding the error thoroughly will be essential to following the remainder of these notes.

2 A Recursion Construct

It's clear that the problem arises with the scope of `let`: it makes the new binding available only in its body. In contrast, we need a construct that will make the new binding available to the named expression also. Different intents, so different names: Rather than change `let`, let's add a new construct to our language, `rec`.

```
<AFunRecExp> ::= ... all the terms in <AFunIfExp>, using <AFunRecExp> instead ...
           | {rec {<id> <AFunRecExp>} <AFunRecExp}
```

AFunRecExp is *AFunIfExp* extended with a construct for recursive binding. We can use `rec` to write a description of factorial as follows:

```
{rec {fac {fun {n}
            {if0 n
                 1
                 {* n {fac {+ n -1}}}}}}
  {fac 5}}
```

Simply defining a new bit of syntax isn't enough; we must also describe what it means. Indeed, notice that syntactically, nothing about `rec` seems to inherently demarcate it as a generator of recursive bindings, since it has the exact same "shape" as `let`. The interesting work lies in the interpreter. But before we get there, we'll first need to think hard about the semantics at a more abstract level.

3 Environments for Recursion

It's clear from the analysis of our failed attempt at writing *fac* above that the problem has something to do with the environments. Let's try to make this intuition more precise.

We can refer to constructs such as `let` as *environment transformers*. That is, they are functions that consume an environment and transform it into one for each of their sub-expressions. We will call the environment they consume—which is the one active outside the use of the construct—the *ambient* environment.

There are two transformers associated with `let`: one for its named expression, and the other for its body. Let's write them both explicitly.

$\rho_{\texttt{let},named}(e) = e$

That is, whatever the ambient environment for a `let`, that's the environment used for the named expression. In contrast,

```
ρ_{let,body}(e) = (new-sub bound-id
                           bound-value
                           e)
```

where *bound-id* and *bound-value* have to be replaced with the corresponding identifier name and value, respectively.

Now let's try to construct the intended transformers for `rec` in the factorial definition above. Since `rec` has two sub-expressions, just like `let`, we will need to describe two transformers. The body seems easier to tackle, so let's try it first. At first blush, we might assume that the body transformer is the same as it was for `let`, so:

```
ρ_{rec,body}(e) = (new-sub 'fac
                           (funV ···)
                           e)
```

Actually, we should be a bit more specific than that: we must specify the environment contained in the closure. Once again, if we had a `let` instead of a `rec`, the closure would close over the ambient environment:

```
ρ_{rec,body}(e) = (new-sub 'fac
                           (funV 'n
                                 (if0 ···) ;; body
                                 e)
                          e)
```

But this is no good! When the `fac` procedure is invoked, the interpreter is going to evaluate its body in the environment bound to *e*, which doesn't have a binding for `fac`. So this environment is only good for the first invocation of `fac`; it leads to an error on subsequent invocations.

Let's consider how this closure came to have that environment. The closure simply closes over whatever environment was active at the point of the procedure's definition. Therefore, the real problem is making sure we have the right environment for the named-expression portion of the `rec`. If we can do that, then the procedure in that position would close over the right environment, and everything would be set up right when we get to the body.

Let's therefore shift our attention to the environment transformer for the named expression. If we evaluate an invocation of `fac` in the ambient environment, it fails immediately because `fac` isn't bound. If we evaluate it in (*new-sub* 'fac (*funV* $\cdots$ *e*) *e*), we can perform one function application before we halt with an error. What if we wanted to be able to perform two function calls (i.e., one to initiate the computation, and one recursive call)? Then the following environment would suffice:

```
ρ_rec,named(e) =
  (new-sub 'fac
        (funV 'n
              (if0 ···) ;; body
              (new-sub 'fac
                    (funV 'n
                          (if0 ···) ;; body
                          e)
                    e))
        e)
```

That is, when the body of `fac` begins to evaluate, it does so in the environment

```
(new-sub 'fac
      (funV 'n
            (if0 ···) ;; body
            e)
      e)
```

which contains the "seeds" for one more invocation of `fac`. That second invocation, however, evaluates its body in the environment bound to *e*, which has no bindings for `fac`, so any further invocations would halt with an error.

Let's try this one more time: the following environment will suffice for *three* recursive invocations of `fac`:

```
ρ_rec,named(e) =
  (new-sub 'fac
        (funV 'n
              (if0 ···) ;; body
              (new-sub 'fac
                    (funV 'n
                          (if0 ···) ;; body
                          (new-sub 'fac
                                (funV 'n
                                      (if0 ···) ;; body
                                      e)
                                e))
                    e))
        e)
```

There's a pattern forming here. To get true recursion, we need to not "bottom out", which happens when we run out of extended environments. This would not happen if the "bottom-most" environment were somehow to refer back to the one enclosing it. If it could do so, we wouldn't even need to go to three levels; we only need one level of environment extension. That is, the place of the boxed *e* in

```
ρ_rec,named(e) =
  (new-sub 'fac
        (funV 'n
              (if0 ···) ;; body
              $\boxed{e}$ )
        e)
```

should instead be a reference to the entire right-hand side of that definition. That is, the environment must be a *cyclic* data structure, or one that refers back to itself.

We don't seem to have an easy way to represent this environment transformer, because we can't formally just draw an arrow from the box to the right-hand side. However, in such cases we can use a variable to name the box's location, and specify the constraint externally (that is, once again, name and conquer). Concretely, we can write

```
$\rho_{\texttt{rec},named}(e) =$
    (new-sub 'fac
              (funV 'n
                        (if0 $\cdots$) ;; body
                        $E$)
              e)
```

But this has introduced an unbound (free) identifier, E. There is an easy way of making sure it's bound:

```
$\rho_{\texttt{rec},named}(e) =$
    $\lambda\, E$ .
        (new-sub 'fac
                  (funV 'n
                            (if0 $\cdots$) ;; body
                            $E$)
                  e)
```

We'll call this a *pre-transformer*, because it consumes both *e*, the ambient environment, and E, the environment to put in the closure. For some ambient environment e_0, let's set

$$F_{e_0} = \rho_{\texttt{rec},named}(e_0)$$

Remember that F_{e_0} is a procedure ready to consume an E, the environment to put in the closure. What does it return? It returns an environment that extends the ambient environment. If we feed the right environment for E, then recursion can proceed forever. What E will enable this?

Whatever we get by feeding some E_0 to F_{e_0}—that is, $F_{e_0}(E_0)$—is precisely the environment that will be bound in the closure in the named expression of the `rec`, by definition of F_{e_0}. But the environment we get back will be the *same* environment: one that extends the ambient environment with a suitable binding for `fac` (because the form of F_{e_0} is (*new-sub* 'fac $\cdots$)). In short,

$$E = F_{e_0}(E)$$

That is, the environment E we need to feed F_{e_0} needs to be the same as the environment we will get from applying F_{e_0} to it. This looks pretty tricky: we're being asked to pass in the very environment we want to get back out! Strange as it may seem, it's not an unreasonable request. We call such a value a *fixed-point* of a function. In this particular case, the fixed-point is an environment that extends the ambient environment with a binding of a name to a closure whose environment is ... itself. That is, the solution is a *cyclic* environment.

Aside: Recursiveness and Cyclicity

It's important to distinguish between *recursive* and *cyclic* data. A recursive object contains references to instances of objects of the same type as it. A cyclic object is a special case of a recursive object. A cyclic object contains references not only to objects of the same *kind* as itself, it contains an actual reference to *itself*.

To easily recall the distinction, think of a typical family tree as a canonical recursive structure. Each person in the tree refers to two more family trees, one each representing the lineage of their mother and father. However, nobody is their own parent, so a family tree is never cyclic. Therefore, structural recursion over a family tree will always terminate. In contrast, the Web is not merely recursive, it's cyclic: a Web page can refer to another page which can refer back to the first one (or, a page can refer to itself). Naïve recursion over a cyclic datum will potentially not terminate: the recursor needs to either not try traversing the entire object, or must track which nodes it has already visited and accordingly prune its traversal. Web search engines face the challenge of doing this efficiently.

Aside: Fixed-Points

You've already seen fixed points in mathematics. For example, consider functions over the real numbers. The function $f(x) = 0$ has exactly one fixed point, because $f(n) = n$ only when $n = 0$. But not all functions over the reals have fixed points: consider $f(x) = x + 1$. A function can have two fixed points: $f(x) = x^2$ has fixed points at 0 and 1 (but not, say, at -1). And because a fixed point occurs whenever the graph of the function intersects the line $y = x$, the function $f(x) = x$ has infinitely many fixed points. The study of fixed points over topological spaces is fascinating, and yields many rich and surprising theorems.[1]

4 A Hazard

When programming with `rec`, we have to be extremely careful to avoid using values before their time. Specifically, consider this program:

```
{rec {f f}
  f}
```

What should this evaluate to? The `f` in the body has whatever value the `f` did in the named expression of the `rec`—whose value is unclear, because we're in the midst of a recursive definition. What you get really depends on the implementation strategy. An implementation could give you some sort of strange, internal value. It could even result in an infinite loop, as `f` tries to look up the definition of `f`, which depends on the definition of `f`, which . . .

There is a safe way out of this pickle. The problem arises because the named expression can be any complex expression, including the identifier we are trying to bind. But recall that we went to all this trouble to create recursive *procedures*. If we merely wanted to bind, say, a number, we have no need to write

```
{rec {n 5}
  {+ n 10}}
```

when we could write

```
{let {n 5}
  {+ n 10}}
```

just as well instead. Therefore, instead of the liberal syntax

```
<AFunRecExp> ::= ... all the terms in <AFunIfExp>, using <AFunRecExp> instead ...
            | {rec {<id> <AFunRecExp>} <AFunRecExp}
```

we could use the more conservative syntax

```
<AFunRecExp> ::= ... all the terms in <AFunIfExp>, using <AFunRecExp> instead ...
            | {rec {<id> <proc>} <AFunRecExp}
```

That is, we demand that the named expression *syntactically* be a procedure. Then, interpretation of the named expression immediately constructs a closure, and the closure can't be applied until we interpret the body—by which time the environment is in a stable state.

Puzzles

1. Are there any other expressions we can allow, beyond syntactic procedures, that would not compromise the safety of this conservative recursion regime?
2. Can you write a reasonable program that is permitted in the liberal syntax, and that safely evaluates to a legitimate value, that the conservative syntax prevents?

[1]One of these is the Brouwer fixed point theorem. The theorem says that every continuous function from the unit n-ball to itself must have a fixed point. A famous consequence of this theorem is the following result. Take two instances of the same map, align them, and lay them flat on a table. Now crumple the upper copy, and lay it atop the smooth map any way you like (but entirely fitting within it). No matter how you place it, at least one point of the crumpled map lies directly above its equivalent point on the smooth map!

5 Implementing Recursion

We have now reduced the problem of creating recursive functions to that of creating cyclic environments.

The interpreter's rule for `let` looks like this:

```
[with (bound-id named-expr bound-body)
      (interp bound-body
              (new-sub bound-id
                       (interp named-expr env)
                       env))]
```

It is tempting to write something similar for `rec`, perhaps making a concession for the recursive environment as follows:

```
[rec (bound-id named-expr bound-body)
     (interp bound-body
             (aRecSub bound-id
                      (interp named-expr env)
                      env))]
```

This is, however, unlikely to work correctly. The problem is that it interprets the named expression in the environment *env*. We know that the named expression must syntactically be a `fun` (the parser, presumably, enforces this), which means its value is going to be a closure. That closure is going to capture its environment, which in this case will be *env*, the ambient environment. But *env* doesn't have a binding for the identifi er being bound by the `rec` expression, which means the function won't be recursive. So this does us no good at all.

Rather than hasten to evaluate the named expression, we could pass the pieces of the function to the procedure that will create the recursive environment. When it creates the recursive environment, it can generate a closure for the named expression that closes over this recursive environment. In code,

```
[rec (bound-id named-expr bound-body)
     (cases AFunRecExp named-expr
       [fun (fun-param fun-body)
          (interp bound-body
                  (aRecSub bound-id
                           fun-param
                           fun-body
                           env))]
       [else (error 'rec "expecting a syntactic procedure")])]
```

This puts the onus on *aRecSub* (the remaining procedures, such as *new-sub* and *lookup*, can remain the same as they were before). We encourage you to ponder this problem; we will examine concrete solutions in our next class. Recall that we have two different representations of environments: as datatypes and as procedures.