(Provisional) Lecture 07: Lists and List Recursion

10:00 AM, Sep 20, 2017

Contents

1 Lists
   1.1 Compound Lists
   1.2 Teasing apart cons-lists
   1.3 Homogeneous Lists
   1.4 List types
   1.5 List Definition

2 Lists: Two Program Interpretations

3 Recursive Diagrams to the Rescue!

4 Summary

Objectives

By the end of this lecture, you will know:

- what lists are
- how to construct them
- how to “deconstruct” them (i.e., select their first and rest constituents)

By the end of this lecture, you will be able to:

- follow the design recipe to write procedures that recur on lists
- draw a recursive diagram to help you write the recursive portion of a recursive procedure.

1 Lists

A Racket list is an abstraction of the kind of list you’re familiar with, like a shopping list or a class list, but with some restrictions. Intuitively, a list is a bunch of things in a row. We can abstract out essential components of a list to make it more applicable in more situations. Take Charlie Brown
for example. As Scott McCloud said in his book “Understanding Comics,” Charlie Brown is drawn so generic so that more people can relate to him. The purpose of abstraction is to generalize.

In Racket, the items in a list can be of many different types (e.g. a boolean, then an int, then a boolean, then a string, ...). But in CS17, our lists will always be homogeneous: we’ll have a list of booleans, or a list of integers, or a list of strings, etc. Because of this, we can describe such a list, once we build one, with a new notation (for use in our type signatures):

- (int list)
- (bool list)
- (str list)

There are exactly two kinds of lists: an empty list and a cons list. The empty list is just what it sounds like: it has no contents. A cons-list (“cons” is short for “construct”) actually has some contents. The empty list is used as a starting point for building up cons-lists.

There’s exactly one empty list, and it’s called empty (which is a keyword). Note, an empty list is the abstraction for your grocery list before you add anything to it. As such, an empty list of numbers is your list before you add numbers to it and an empty list of strings is your list before you add the strings to it.

There’s an associated predicate, empty?, which is true for the empty list and false for everything else.

- (empty? empty) => true
- (empty? 15) => false
- (empty? jar) => reference to an identifier before its definition: jar

The printed representation of the (sole) empty list is empty.

**1.1 Compound Lists**

The other kind of list is a nonempty, or compound list or cons list.

Here’s an example of building a “cons-list” with the builtin procedure cons:

- (cons 31 empty)

As you can see from the example, the cons procedure is used in this format: (cons item list). In this example, the number 31 is in the first field and the list empty is in the second. That is to say, a cons-list structure has two parts: a datum and a list. The datum represents a new item to be added to the list that is the second item, and the value is (intuitively) a new list that contains that item followed by all the items of the second list in their original order.

To be useful, such a cons-list must not only be able to be constructed, but should be able to be deconstructed as well: we should be able to somehow discover what items are in the list. We’ll come to that in just a moment.
Terminology: Compound lists or cons lists are constructed by “cons”-ing the first field onto the second.

Here’s another example:

```scheme
(cons 18 (cons 31 empty))
```

One essential feature of a compound list is: the second field is required to be a list: either an atomic list (i.e., `empty`), or another compound list. So, this is a compound list: `(cons 17 (cons 18 empty))`. But this is not: `(cons 17 18)`. In fact, in CS17, that last expression is not even allowed. (In some versions of Racket, it is allowed; read about those on your own if you’re interested, but never use them in CS17.)

You can use the predicate `cons?` to test whether a datum is a compound list.

```scheme
(cons? (cons 17 empty)) => true
(cons? 15) => false
(cons? empty) => false
(cons? empty?) => false
```

We can also see the predicate `empty?` return `false` now.

```scheme
(empty? (cons 3 empty)) => false
```

### 1.2 Teasing apart cons-lists

The two things that constitute a compound list can be accessed using the `first` and `rest` selectors.

```scheme
(first (cons 17 empty)) => 17
(rest (cons 17 empty)) => empty
```

More generally, the `first` and `rest` selectors satisfy the following algebraic identities:

---

1. `cons` is short for “construct.”
2. In the “olden days,” `first` and `rest` were named `car` and `cdr`, respectively. CAR stood for Contents of Address Register, and CDR, for Contents of Decrement Register. Although you may encounter these terms from time to time in your CS career (indeed DrRacket recognizes these names), the corresponding descriptions are no longer accurate, and we’ll never use them in CS17.
(first (cons x y))
=> x

(rest (cons x y))
=> y

These selectors produce an error if their argument is not a compound list:

(first 17)
=> first: expected argument of type <non-empty list>; given 17

(rest empty)
=> rest: expected argument of type <non-empty list>; given empty

**Question:** Can the first field store a list as well?

**Answer:** Sure! The first field can store any kind of data. If a list’s first field is a list of numbers, for example, then the list itself is a list of list of numbers.

**Question:** Are there any selectors that can access any element of a list (not just the first element of a list and the rest of the list)?

**Answer:** Unfortunately not. You will have to rely on first and rest to access those elements. You will learn how to do that (very) soon!

### 1.3 Homogeneous Lists

Though we sometimes fail to be rigorous about this point—in particular, in the lecture thus far, we spoke of lists generically—there are, for CS 17, only lists of numbers, lists of booleans, lists of strings, lists of structs, etc. There are also lists of lists of numbers, lists of lists of lists of numbers, and so on.

Here are formal definitions of a few representative lists:

A *list of strings* is either:

- empty, or
- (cons str alos), where str is a string and alos is a list of strings

Nothing else is a list of strings.

Similarly, a *list of numbers* is either:

- empty, or
- (cons num alon), where num is a number and alon is a list of numbers

Nothing else is a list of numbers.

As we have already noted, lists are containers—containers for data. In CS 17, we insist that the lists you define be *homogeneous*. This means that they can contain only one type of data. So lists
of strings are fine, and lists of numbers are fine, but please do not create lists that contain both strings and numbers (or any other combination of non-homogeneous data).

In particular, the following sort of list is a no-no in CS 17:

\[
\text{(define i-am-not-a-CS-17-list (cons 15 (cons "fifteen" empty)))}
\]

1.4 List types

If we have a procedure that operates on a list of integers and produces a string, we write

\[
\text{;; my-proc: (int list) -> str}
\]

That is to say: lists are described (in signatures) in the form (..., list), where the ellipsis is some data type and the parens and the word “list” are required.

We’ll return to this and make it a bit fancier soon.

1.5 List Definition

The above definitions, in which a datatype (here, a list) is defined in terms of itself, is called recursive, or inductive. Note that well-crafted recursive definitions are not circular because: (i) there is a base case, in which the type is not defined in terms of itself, and (ii) the other parts in the recursive definition define the data of that type in terms of a “smaller” datum of that type.

2 Lists: Two Program Interpretations

Let’s start using lists in a program. We’ll start with a fairly silly example, a procedure called list-size that returns the string "small" when a-list is empty and "big" otherwise. To make it concrete, we’ll have it operate on int lists.

To write this procedure, we’ll rely on our trusty design recipe, slightly revised:

2. Define the data, specifically the type of list the procedure will process.

3. Provide examples of the data the procedure will process. Also, provide examples of the output data the procedure will produce.

4. Specify the procedure’s type signature, which describes the type of lists and other data the procedure consumes, and the type it produces.

5. Following the type signature, describe the procedure’s call structure. Doing so gives names to the procedure and its arguments.

6. Write a specification for the procedure. That is, in words, not code, state the relationship between the procedure’s input and output. Make sure to refer to the inputs by their argument names.
7. Provide test cases that exemplify the procedure’s operation. These tests must follow its call structure and satisfy its specification. Be sure to test all cases in your data definition: i.e., all possible varieties of lists.

8. Write a template for the procedure based on the data definition and type signature. [More on this below.]

9. Code the procedure, using the template as a reference. Specifically, fill in the template clause by clause. For each of the possible input types, decide which fields in the input structures are relevant to the problem at hand, and figure out how to operate on them to generate the desired output.

10. Run your program on your test cases.

What’s this “template” thing?

The template for a procedure that operates on lists follows the structure of the data by using a cond expression. Note, the structure of our program will mimic the structure of our data. You will almost always use cond when writing procedures with lists. And because lists come in two forms, our cond expression will always have two cases: one for the argument being an empty list, the other for when it’s a cons list. Those will be the two “conditions” in the cond-expression.

What about the two results? For the empty list, there’s a single result to be written down, and the empty list itself has no associated data except that it’s empty. So we write . . . .

For a cons-list, the result can depend on the contents of the list, which we can get by using the first and rest procedures to access that content. So the result, for the cons case of the cons expression, looks something like

( ... (first my-list) ... (rest my-list) ... )

Also of particular interest in the case of lists (and trees, and any recursive data type) is a possible recursive call to the procedure itself in the answer to the cons? question. We’ll see that in a moment.

Now let’s get to writing that program.

Step 2: The first step in the design recipe is to give an abstract definition of the data. As you know, our plan is to write a list-size procedure that operates on lists. But what sort of lists? For concreteness and homogeneity, let’s work with lists of numbers. In this case, we have:

;; (int list)
;; - empty
;; - (cons int (int list))

Step 3 Here are some examples of lists of numbers:

(define 10 empty)
(define 11 (cons 19 empty))
(define 12 (cons 18 (cons 19 empty)))
(define 13 (cons 17 (cons 18 (cons 19 empty))))

Here are some examples of strings:
Step 4: Write a type signature. As already mentioned, our list-size procedure will take as input a list of integers. It will output a string that is either "small" or "big".

```scheme
;; list-size : (int list) → string
```

Step 5: The length procedure has the following call structure:

```scheme
(define (list-size aloi)
  ...
)
```

Note that aloi is an abbreviation for “a list of integers”.

Step 6: The specification of the list-size procedure is straightforward:

```scheme
;; Input: a list of integers, aloi
;; Output: a string, "small" if the input list is empty, "big" otherwise.
```

Step 7: Following the data definition and examples provided in Step 2 construct test cases:

```scheme
;; Test cases for list-size
(check-expect (list-size 10) "small")
(check-expect (list-size 11) "big")
(check-expect (list-size 12) "big")
(check-expect (list-size 13) "big")
```

It is essential that you test all cases that arise in your data definition: i.e., all possible varieties of input data. Here we’ve checked every variety of input data, both empty and cons lists!

Step 2: The next step in the design recipe is to write a template for our procedure based on the data definition and type signature. Here’s how to do that for lists:

```scheme
(define (list-size aloi)
  (cond
   [(empty? aloi) ... ]
   [(cons? aloi) ( ... (first aloi) ... (list-size (rest aloi)) ... )]))
```

This abstract template describes how procedures typically operate on lists. Following the recursive data definition, the template considers two cases, one for each case of the data definition, in this case empty? and cons?. In the cons? case, where the input is non-atomic, the procedure uses the data’s selection operators to get at its constituents, namely the first element and the rest of the list.

At this point, we have completed six of the eight steps in our design recipe. We have a data definition and examples of the data, a type signature, a call structure, a specification, some test cases, and a template.
Step 9: The next (and certainly not least) step is to fill in our template. To do so, we must determine two things: first, what should our procedure do in the empty? case; second, what should it do in the cons? case? To determine what it should do in the empty? case is simple: If the list is empty, the procedure should produce "small". So, we fill in "small" in the appropriate location in our template.

```lisp
(define (list-size aloi)
  (cond
    [(empty? aloi) "small"]
    [(cons? aloi) ... (first aloi) ... (list-size (rest aloi)) ... ]))
```

The cons? case for this example is equally simple. If the list is a compound list, return "big". Hence, our code looks like this:

```lisp
(define (list-size aloi)
  (cond
    [(empty? aloi) "small"]
    [(cons? aloi) "big"]))
```

**Question:** What happened to `(first aloi)` and `(rest aloi)`? It was part of the template, but it isn’t part of the final procedure.

**Answer:** As it happens, the list-size procedure doesn’t make use of the first element of the input list nor the rest of the list. In general, these the only things you can use, but you don’t have to use all of them! Other procedures however, as you will see later, do use first and rest. All constituents of the data must appear as part of the template, even though not all procedures need access to all constituents, because you may not know beforehand exactly which constituents are needed by which procedures.

Now here is list-size in its entirety:

```lisp
;; (int list)
;; - empty
;; - (cons number (int list))

(define 10 empty)
(define 11 (cons 19 empty))
(define 12 (cons 18 (cons 19 empty)))
(define 13 (cons 17 (cons 18 (cons 19 empty))))

;; example of strings
;; ""
;; "small"
;; "big"

;; list-size : (int list) → string
;; Input: a list of integers, aloi
;; Output: a string, "small" if the input list is empty, "big" otherwise.

(define (list-size aloi)
  (cond
```


Step[10] But wait! There’s one more step in our design recipe: run. Does our code pass all of our test cases? Normally, Racket will answer this question for you. And we showed in class that it worked fine. But in general, we’ll have to evaluate things manually just to get a grip on what this recursion thing is doing. We’ll do that in the next class.

Step[2] The first step, again, in the design recipe is to give an abstract definition of the data. As you know, our plan is to write a `list-length` procedure that operates on lists. But what sort of lists? For concreteness, let’s work with lists of integers. In this case, we have:

```racket
;; (int list)
;; - empty
;; - (cons number (int list))
```

Step[3]

```racket
(define 10 empty)
(define 11 (cons 19 empty))
(define 12 (cons 18 (cons 19 empty)))
(define 13 (cons 17 (cons 18 (cons 19 empty))))
```

Here are some examples of integers:

```racket
;; output examples
;; 0
;; 1
```

Step[4] Write a type signature. As already mentioned, our `list-length` procedure will take as input a list of integers. It will output a single integer:

```racket
;; list-length : (int list) \rightarrow int
```

Step[5] The `list-length` procedure has the following call structure:

```racket
(define (list-length aloi)
  ...
)
```

Step[6] The specification of the `list-length` procedure is straightforward:

```racket
;; Input: a list of integers, aloi
;; Output: an integer, the length of aloi
```
Step 7: Following the data definition and examples provided in Step 2, construct test cases:

;;; Test cases for length
(check-expect (list-length 10) 0)
(check-expect (list-length 11) 1)
(check-expect (list-length 12) 2)
(check-expect (list-length 13) 3)

Again note that we’re testing on all cases that arise in the list data definition.

Step 2: The next step in the design recipe is to write a template for our procedure based on the data definition and type signature. Just like last time:

(define (list-length aloi)
  (cond
    [(empty? aloi) ... ]
    [(cons? aloi) ( ... (first aloi) ... (list-length (rest aloi)) ... )]))

At this point, we have again completed six of the eight steps in our design recipe. We have a data definition and examples of the data, a type signature, a call structure, a specification, some test cases, and a template.

Step 9: The next (and certainly not least) step is to fill in our template. Just like last time, we must ask ourselves: first, what should our procedure do in the empty? case; second, what should it do in the cons? case? To determine what it should do in the empty? case is easy: check the test cases. The answer is immediate: If the list is empty (i.e., l0), the procedure should produce 0. So, we fill in 0 in the appropriate location in our template.

(define (list-length aloi)
  (cond
    [(empty? aloi) 0]
    [(cons? aloi) ... (first aloi) ... (list-length (rest aloi)) ... ]))

3 Recursive Diagrams to the Rescue!

For the cons? case, we use a peculiarly clever idea. That is, the length of a compound list (cons item lst) is always 1 greater than the length of lst. To further illustrate this idea, we rely on a neat tool—a recursive diagram—and help us break down the problem:

1. First, we write down the original input and the corresponding original output. The original input is whatever example you’re working on. The original output is whatever the procedure-specification says it should be. Filling in these two parts of the diagram doesn’t require anything but the specification.

2. Second, we figure out the recursive input (this is often already done in the template). For many many list-processing procedures, the recursive input is the “rest” of the list; it’s certainly a good thing to try. We’ll soon see other possibilities, however.
3. Third, we pose the problem on the recursive input, which yields the recursive output (again, by looking at the procedure specification).

4. Fourth, we figure out how to reconstruct the original output from the recursive output.

Here is a sample recursive diagram for the list-length procedure:

- original input: (cons 1 (cons 2 (cons 3 empty)))
  - recursive input: (cons 2 (cons 3 empty))
  - recursive output: 2

- original output: 3

And here is another:

- original input: (cons 10 (cons 20 (cons 30 (cons 40 (cons 50 empty)))))
  - recursive input: (cons 20 (cons 30 (cons 40 (cons 50 empty))))
  - recursive output: 4

- original output: 5

In both cases, I’ve left a gap between these items; that’s space in which you can write ideas about how you might transform the recursive result into the original result. In the case of this procedure, that’s fairly easy, I hope. As you can see, we simply add one to get from our recursive output to our original output. In more complex procedures, this is where the real work happens.

So this approach only really requires that we figure out two key things:

1. how to derive the recursive input from the original input
2. how to derive the original output from the recursive output
Let us address each of these questions in turn. First, to derive the recursive input from the original input, we will take a closer look at the recursive data definition for lists. The recursive input is simply the smaller list of numbers, from which the original input is constructed. And how can we get at this smaller list? Using the rest selector, of course. In general, the most important feature of the recursive input is that it must be in some way smaller than the original input. We’ll talk about what “smaller” can mean in many ways over the next few weeks. But for many of the procedures that you’ll write for the next week, the way to get the recursive input is to compute the rest of the original input.

In the first diagram, we see that the length of the recursive input is 2, while the length of the original is 3. In the second diagram, we see that the length of the recursive input is 4, while the length of the original is 5. How can we derive the original output from the recursive output? Easy. The length of the original output is 1 more than the length of the recursive output. (So we might, in that open space between the two outputs, write something like “Just add one!”)

Hence, our code looks like this:

```racket
(define (list-length aloi)
  (cond
    [(empty? aloi) 0]
    [(cons? aloi) (+ 1 (list-length (rest aloi)))]))
```

**Question:** Does it really make sense to call `list-length` on `rest aloi`? The `list-length` procedure operates on `aloi`.

**Answer:** In the call structure `(list-length aloi)`, `aloi` is just a name for `list-length`’s formal argument. When the `list-length` procedure is evaluated, the actual argument with which the procedure is called (in this case, `rest aloi`) is bound to the formal argument (in this case, `aloi`). This binding is then added to the environment under which the procedure is evaluated.

**Question:** What would happen if we were to recur on the whole of the input list instead of just the rest of that list?

**Answer:** Your input would never get smaller, so your procedure would never terminate. That would violate the advice that the recursive input should be smaller than the original input.

And we’ve done it! We’ve written our first recursive procedure in Racket! Here it is, in its entirety:

```racket
;; (int list)
;; - empty
;; - (cons int (int list))

(define 10 empty)
(define 11 (cons 17 empty))
(define 12 (cons 18 (cons 17 empty)))
(define 13 (cons 19 (cons 18 (cons 17 empty))))

;; example ints
;; 0
;; 1

;; list-length : (int list) → int
```
;; Input: a list of integers, aloi
;; Output: an integer, the length of aloi

(define (list-length aloi)
  (cond
   [(empty? aloi) 0]
   [(cons? aloi) (+ 1 (list-length (rest aloi)))]))

;; Test cases for length
(check-expect (list-length l0) 0)
(check-expect (list-length l1) 1)
(check-expect (list-length l2) 2)
(check-expect (list-length l3) 3)

Step 10: But wait! There’s one more step in our design recipe: run. Does our code pass all of our test cases? Yes! But why? For that, you can work through the rules of processing in the particular cases of 10 and 11 to see what happens, and to stand back and admire the cleverness of the rules. If you don’t want to do that, I’ll be doing it for you on Monday.

4 Summary

Ideas

- We’ve introduced the concept of lists in programming. Lists are an important data structure in Racket. In CS17, we will use homogeneous lists - lists of one type - in quite a few procedures. As such, we’ve also introduced a specialized design recipe when working with lists.
- We’ve also introduced the concept of recursion. Recursive procedures work by calling themselves. In functional programming, recursion offers an elegant way to solve problems.

Skills

- We’ve introduced the first and rest selector which access the first item in the list and a list containing all but the first item, respectively.
- We’ve also introduced two predicates. empty? checks if a list is empty or not, and cons? checks if a list contains any data or not.
- We’ve learned how to use recursive diagrams to break down and solve a problem. While the example of recursive diagrams for list size is primitive, you will soon see how to use recursive diagrams to solve complex problems.

Please let us know if you find any mistakes, inconsistencies, or confusing language in this or any other CS17 document by filling out the anonymous feedback form: [http://cs.brown.edu/courses/cs017/feedback](http://cs.brown.edu/courses/cs017/feedback)