What is generative recursion?

Structural recursion, which we have been using so far, is a way of deriving code whose form parallels a data definition.

Generative recursion is more general: the recursive cases are generated based on the problem to be solved.

The non-recursive cases also do not follow from a data definition.

It is much harder to come up with such solutions to problems.

It often requires deeper analysis and domain-specific knowledge.

Example revisited: GCD

;; (euclid-gcd n m) computes gcd(n,m) using Euclidean algorithm
;; euclid-gcd: Nat Nat → Nat
(define (euclid-gcd n m)
  (cond [(zero? m) n]
        [else (euclid-gcd m (remainder n m))])))
Why does this work?

**Correctness:** Follows from Math 135 proof of the identity.

**Termination:** An application terminates if it can be reduced to a value in finite time.

All of our functions so far have terminated. But why?

For a non-recursive function, it is easy to argue that it terminates, assuming all applications inside it do.

It is not clear what to do for recursive functions.

---

**Termination of recursive functions**

Why did our structurally recursive functions terminate?

A structurally recursive function always makes recursive applications on smaller instances, whose size is bounded below by the base case (e.g. the empty list).

We can thus bound the depth of recursion (the number of applications of the function before arriving at a base case).

As a result, the evaluation cannot go on forever.

---

(sum-list (list 3 6 5 4)) ⇒
(+ 3 (sum-list (list 6 5 4))) ⇒
(+ 3 (+ 6 (sum-list (list 5 4)))) ⇒ . . .

The depth of recursion of any application of sum-list is equal to the length of the list to which it is applied.

For generatively recursive functions, we need to make a similar argument.
In the case of euclid-gcd, our measure of progress is the size of the second argument.

If the first argument is smaller than the second argument, the first recursive application switches them, which makes the second argument smaller.

After that, the second argument is always smaller than the first argument in any recursive application, due to the application of the remainder modulo $m$.

The second argument always gets smaller in the recursive application (since $m > n \mod m$), but it is bounded below by 0.

Thus any application of euclid-gcd has a depth of recursion bounded by the second argument.

In fact, it is always much faster than this.

### Termination is sometimes hard

;; collatz: Nat → Nat
;; requires: n ≥ 1
(define (collatz n)
  (cond [(= n 1) 1]
        [(even? n) (collatz (/ n 2))]
        [else (collatz (+ 1 (* 3 n)))]))

It is a decades-old open research problem to discover whether or not (collatz n) terminates for all values of $n$. 

We can see better what collatz is doing by producing a list.

;; (collatz-list n) produces the list of the intermediate
;; results calculated by the collatz function.
;; collatz-list: Nat → (listof Nat)
;; requires: n >= 1
(check-expect (collatz-list 1) '(1))
(check-expect (collatz-list 4) '(4 2 1))
(define (collatz-list n)
  (cons n (cond
               [(= n 1) empty]
               [(even? n) (collatz-list (/ n 2))]
               [else (collatz-list (+ 1 (* 3 n)))])))

---

**Hoare’s Quicksort**

The Quicksort algorithm is an example of *divide and conquer*:

- divide a problem into smaller subproblems;
- recursively solve each one;
- combine the solutions to solve the original problem.

Quicksort sorts a list of numbers into non-decreasing order by first choosing a **pivot** element from the list.

The subproblems consist of the elements less than the pivot, and those greater than the pivot.

If the list is (list 9 4 15 2 12 20), and the pivot is 9, then the subproblems are (list 4 2) and (list 15 12 20).

Recursively sorting the two subproblem lists gives (list 2 4) and (list 15 20).

It is now simple to combine them with the pivot to give the answer.

(append (list 2 4) (list 9) (list 12 15 20))
The easiest pivot to select from a list \texttt{lon} is \texttt{(first lon)}.

A function which tests whether another item is less than the pivot is
\texttt{(lambda (x) (< x (first lon)))}.

The first subproblem is then
\texttt{(filter (lambda (x) (< x (first lon))) \texttt{lon})}.

A similar expression will find the second subproblem
\texttt{(items greater than the pivot)}.

```
;; (quick-sort \texttt{lon}) sorts \texttt{lon} in non-decreasing order
;; quick-sort: (listof Num) \rightarrow (listof Num)
(define (quick-sort \texttt{lon})
  (cond [(empty? \texttt{lon}) empty]
        [else (local
            [(define pivot (first \texttt{lon}))
             (define less (filter (lambda (x) (< x pivot)) (rest \texttt{lon}))
             (define greater (filter (lambda (x) (>= x pivot)) (rest \texttt{lon})))]
             (append (quick-sort less) (list pivot) (quick-sort greater))))])
```

Termination of quicksort follows from the fact that both subproblems
have fewer elements than the original list (since neither contains the
pivot).

Thus the depth of recursion of an application of \texttt{quick-sort} is
bounded above by the number of elements in the argument list.

This would not have been true if we had mistakenly written
\texttt{(filter (lambda (x) (>= x pivot)) \texttt{lon})}

instead of the correct
\texttt{(filter (lambda (x) (>= x pivot)) (rest \texttt{lon}))}
In the teaching languages, the built-in function `quicksort` (note no hyphen) consumes two arguments, a list and a comparison function.

\[
\text{quicksort '(1 5 2 4 3) <) } \Rightarrow \text{'(1 2 3 4 5)}
\]
\[
\text{quicksort '(1 5 2 4 3) >) } \Rightarrow \text{'(5 4 3 2 1)}
\]

Intuitively, quicksort works best when the two recursive function applications are on arguments about the same size.

When one recursive function application is always on an empty list (as is the case when quick-sort is applied to an already-sorted list), the pattern of recursion is similar to the worst case of insertion sort, and the number of steps is roughly proportional to the square of the length of the list.

We will go into more detail on efficiency considerations in CS 136.

Here is another way of looking at quicksort.

Consider taking a list and inserting each element into an initially-empty binary search tree as a key with some dummy value, using an accumulator so that the first element of the list is the first inserted.
(define (list→bst lon)
  (local [(define (list→bst/acc lon bst-so-far)
                  (cond [(empty? lon) bst-so-far]
                        [else (list→bst/acc
                             (rest lon)
                             (add-to-bst (first lon)
                             "dummy"
                             bst-so-far))))
             (list→bst/acc lon empty))])

Inserting the elements of the list in order uses structural recursion on the list, with an accumulator for the BST.

Each insertion uses structural recursion on the BST.

We can recover the original list, but in sorted order, by flattening the tree in such a manner that the key of the root node comes between the flattened result of the left tree and the flattened result of the right tree (instead of before both).

This is called an **inorder traversal**, and computing it also uses structural recursion on the BST.

;; (bst-inorder t) flattens a BST into a sorted list
;;     with root between left and right
;; bst-inorder: BST → (listof Num)
(define (bst-inorder t)
  (cond [(empty? t) empty]
        [else (append (bst-inorder (node-left t))
                     (list (node-key t))
                     (bst-inorder (node-right t)))]))
(define (treesort lon)
  (bst-inorder (list → bst lon)))

Quicksort can be viewed as simply eliminating the trees from treesort.

The first key inserted into the BST, which ends up at the root, is the pivot.

(bst-inorder (node-left t)) is exactly the list of items of value less than the pivot.

Quicksort takes a number of structurally-recursive computations and makes them more efficient by using generative recursion.

Modifying the design recipe

The design recipe becomes much more vague when we move away from data-directed design.

The purpose statement remains unchanged, but additional documentation is often required to describe how the function works.

Examples need to illustrate the workings of the algorithm.

We cannot apply a template, since there is no data definition.

Typically there are tests for the easy cases that don’t require recursion, followed by the formulation and recursive solution of subproblems, and then combination of the solutions.

Example: breaking strings into lines

Traditionally, the character set used in computers has included not only alphanumeric characters and punctuation, but “control” characters as well.

An example in Racket is #\newline, which signals the start of a new line of text.

The characters ‘\’ and ‘n’ appearing consecutively in a string constant are interpreted as a single newline character.

For example, the string "ab\ncd" is a five-character string with a newline as the third character.
A line is a maximal sequence of non-newline characters.

Text files often contain embedded newlines to split the file into lines.
The same can be true of strings entered by a user or displayed.

We will develop a function to break a string into lines.

\[
\text{(check-expect (string→strlines "abc\ndef") (list "abc" "def"))}
\]

What should the following produce?

\[
\begin{align*}
\text{(string→strlines "abc\ndef\nghi")} \\
\text{(string→strlines "\nabc")} \\
\text{(string→strlines "abc")} \\
\text{(string→strlines "c")} \\
\text{(string→strlines "")} \\
\text{(string→strlines "\n")} \\
\text{(string→strlines "abc\n\ngghi")} \\
\text{(string→strlines "\n\n")}
\end{align*}
\]

Internally, \text{string→strlines} will need to convert its argument into a list
of characters using \text{string→list}. We'll apply a helper function, 
\text{list→lines}, to that list.

What should \text{list→lines} produce?

One natural solution is a list, where each item is a list of characters
that make up one line in \text{string→strlines}'s result.

\[
\begin{align*}
\text{;; (list→lines loc) turns loc into list of char lists based on newlines} \\
\text{;; list→lines: (listof Char) → (listof (listof Char))} \\
\text{;; Example:} \\
\text{(check-expect (list→lines '(#\a #\b #\newline #\c #\d))} \\
\text{ '((#\a #\b) (#\c #\d)))}
\end{align*}
\]
Thus string→strlines can be implemented as

\[\text{string} \rightarrow \text{strlines} : \text{Str} \rightarrow (\text{listof Str})\]

\[
\begin{align*}
\text{(define (string→strlines str)} & \text{)} \\
\text{ (local} & \text{)} \\
\text{ ;; list→lines: (listof Char) } \rightarrow (\text{listof (listof Char)}) \\
\text{ [(define (list→lines loc)} & \text{)} \\
\text{ . . . )]} \\
\text{ (map list→string)} & \text{)} \\
\text{ (list→lines (string→list str)))})
\end{align*}
\]

We can use structural recursion on lists to write list→lines, but as we will see, the resulting code is awkward.

Rewriting the usual list template just a bit so we can refer to the recursive application several times without recomputing it gives us:

\[
\begin{align*}
\text{(define (list→lines loc)} & \text{)} \\
\text{ (cond} [(\text{empty? loc)} . . . ] \\
\text{ [else} \\
\text{ (local} [(\text{define r (list→lines (rest loc))}) \\
\text{ . . . (first loc) . . . r . . . )})
\end{align*}
\]

There are two base cases to consider: the first is where the list is empty, and the second is where there is a newline character at the very end.

We can then add this additional base case to our template.

\[
\begin{align*}
\text{(define (list→lines loc)} & \text{)} \\
\text{ (cond} [(\text{empty? loc)} . . . ] \\
\text{ [(and (empty? (rest loc)) (char=? \#\newline (first loc)) . . . ]} \\
\text{ [else} \\
\text{ (local} [(\text{define r (list→lines (rest loc))}) \\
\text{ . . . (first loc) . . . r . . . )})
\end{align*}
\]
In both of these base cases, we start our list of lines with an empty line.

\[
(\text{define (list→lines loc)} \\
(\text{cond} [(\text{empty? loc)} \ (\text{list empty})] \\
\quad [(\text{and} \ (\text{empty? (rest loc)}) \ (\text{char=} \ #\newline) \ (\text{first loc}))] \\
\quad \ (\text{list empty})] \\
\quad \ [\text{else} \\
\quad \ (\text{local} [(\text{define r} \ (\text{list→lines (rest loc)}))] \\
\quad \ (\ldots \ (\text{first loc}) \ldots \ r \ldots)))]))
\]

CS 135 Winter 2019 11: Generative recursion 31

If \(\text{loc}\) is nonempty, and has \#\newline as its first character, we \text{cons} an empty line onto the recursive result \(r\).

If \(\text{loc}\) has, say, \#\a as its first character, then this should be put at the front of the first line in the recursive result \(r\).
This approach works, but is pretty low-level and hard to reason about. If we reformulate the problem at a higher level, it turns out to be easier to reason about – but requires generative recursion to implement.

A generative approach

View it as sequence of lines instead of a sequence of characters.

```
;; string → strlines: Str → (listof Str)
(define (string→strlines/gen str)
  (local
    ;; list→lines/gen: (listof Char) → (listof (listof Char))
    [(define (list→lines/gen loc)
       . . . )]
    (map list→string
     (list→lines/gen (string→list str))))
```

We can create a “sequence of lines template”.

```
;; list→lines: (listof Char) → (listof (listof Char))
(define (list→lines/gen loc)
  (local
    ;; list→lines (rest-of-lines loc))]
    (cond [(empty? rlines) . . . fline . . .]
         [else . . . fline . . . (list→lines/gen rlines) . . .]))
```
We can easily fill in this template.

;; list→lines: (listof Char) → (listof (listof Char))
(define (list→lines/gen loc)
  (local
    [(define fline (first-line loc))
     (define rlines (rest-of-lines loc))
     (cond 
       [(empty? rlines) (list fline)]
       [else (cons fline (list→lines/gen rlines))])))

The two helper functions are also simpler, using structural recursion on lists.

;; (first-line loc) produces longest newline-free prefix of loc
;; first-line: (listof Char) → (listof Char)
;; Examples:
  (check-expect (first-line empty) empty)
  (check-expect (first-line '(#\newline)) empty)
  (check-expect (first-line '(#\a #\newline)) '(#\a))

(define (first-line loc)
  (cond 
    [(empty? loc) empty]
    [(char=? (first loc) #\newline) empty]
    [else (cons (first loc) (first-line (rest loc)))]))

;; (rest-of-lines loc) produces loc with everything
;; up to and including the first newline removed
;; rest-of-lines: (listof Char) → (listof Char)
;; Examples:
  (check-expect (rest-of-lines empty) empty)
  (check-expect (rest-of-lines '(#\a #\newline)) '(#\a))
  (check-expect (rest-of-lines '(#\a #\newline #\b)) '(#\b))

(define (rest-of-lines loc)
  (cond 
    [(empty? loc) empty]
    [(char=? (first loc) #\newline) (rest loc)]
    [else (rest-of-lines (rest loc))])))
Generalizing string → strlines

We can generalize string → strlines to accept a predicate that specifies when to break the string. We will call the pieces tokens instead of lines.

;; string → tokens: (Char → Bool) Str → (listof Str)
(define (string→tokens break? str)
  (map list→string
       (list→tokens break? (string→list str))))

(check-expect
  (string→tokens char-whitespace? "One two
three")
  (list "One" "two" "three"))
Tokenizing is especially important in interpreting or parsing a computer program or other specification. For instance,

\[
\text{(define (square x)} \newline
\quad (\times x x))
\]

\text{can be broken into twelve tokens.}

Here, our simple method of using a character predicate will not work.

In later courses, we will discuss tokenizing in order to deal with situations such as a string representing a computer program, a Web page, or some other form of structured data.

\textbf{Goals of this module}

You should understand the idea of generative recursion, why termination is not assured, and how a quantitative notion of a measure of progress towards a solution can be used to justify that such a function will return a result.

You should understand the examples given, particularly quicksort, line breaking, and tokenizing using a general break predicate.

You should be comfortable with accumulative recursion, and able to write accumulatively recursive helper functions with associated wrapper functions to hide the accumulator(s) from the user.