15-150 Lecture 23:
Lazy functional programming

Lectures by Stefan Muller∗
June 22, 2018

1 Topics

• Lazy programming (in a call-by-value language)
• Delaying computation
• Demand-driven manipulation of infinite data

2 Introduction

SML is a call-by-value language: functions evaluate their arguments. Haskell, another functional
programming language, is not: the whole argument gets substituted in as an expression and is
evaluated only when it’s needed. For example, if you add 3 + 5, the addition will only actually
be done when someone asks for the answer (e.g. by checking if the result is greater than 7). This
behavior is referred to as call-by-need or lazy.

It turns out that while SML does not do this by default1, we can still write programs with this
behavior. We can do this by using functions to delay computation. Functions are values: SML
doesn’t evaluate inside a function body until, if and when, the function is applied to a value.

We can use this property to delay a computation by making it be the body of a function known
as a suspension, sometimes also called a thunk. When the function is called on its argument, it
computes the head of the stream, along with another thunk for computing the next element of the
stream, and so forth.

Definition: A suspension of a type τ is a function of type unit -> τ. We say a suspension is
forced when it is applied to argument ()

The reason we care about suspensions is that we want to delay evaluation. In an eager language
like SML, writing e (e.g., in the REPL) causes evaluation of e. Writing (fn () => e) prevents
immediate evaluation of e since functions are already values. Only when forced will the suspension
cause evaluation of e.

We’ll focus on one particular application of laziness: representing infinite data structures.

∗based on notes by Michael Erdmann and Stephen Brookes for 15-150
1For the purposes of this class, it has been important for us to use a call-by-value language, as laziness makes it
much more difficult to reason about cost and correctness.
3 Representing Infinite Data

Because SML is call-by-value, an attempt to build an infinite list, such as the following code fragment, is doomed to failure (actually, will loop forever):

(* repeat : 'a list -> 'a list *)
fun repeat L = L @ repeat L

Even though SML tells us this function is well-typed, if we apply it to a list we won’t get a value. For example

repeat [0] ==> [0] @ repeat [0]
===> [0] @ ([0] @ repeat [0])
===> . . . forever

Although this is a silly example, it illustrates a problem. We may need to build and manipulate infinite data structures, and deal with such data in a demand-driven manner: just evaluate as much as we need, while leaving the rest for later if we need it. The call-by-value behavior of functions needs to be side-stepped somehow, if we want to prevent eager evaluation of the rest of the data.

In a functional programming language, one way to represent an infinite list of data of some type t is as a function int -> t. However, this kind of representation is not well-tailored for lazy demand-driven computation, particularly when future data may depend on external input. Moreover, such a representation may create a lot of tedious index manipulation.

Today we will introduce another representation for potentially infinite lists, which we call streams. We will implement operations on streams that facilitate solving non-trivial demand-driven problems.

The main idea is that a t stream of values of type t is really a delayed computation.

We define streams as follows:

datatype 'a stream = Stream of unit -> 'a front
and 'a front = Empty | Cons of 'a * 'a stream

Observe that for any type t, a t stream is basically a suspension of a t front, which looks much like a t list, only now as deferred computation. Specifically, a (nonempty) t front is a “cons” of an element of type t and another t stream.

We need a function for turning a suspension of a front into a stream, and for exposing the first element of a stream by forcing the underlying suspension:

(* delay : (unit -> 'a front) -> 'a stream *)
fun delay d = Stream(d)

(* expose : 'a stream -> 'a front *)
fun expose (Stream(d)) = d( )

We also have an empty stream, which is essentially just a suspension of an empty front:

val empty : 'a stream = Stream(fn () => Empty)

Finally, we have a function for adding an element to the beginning of a stream, returning the resulting stream:

(* cons : 'a * 'a stream -> 'a stream *)
fun cons (x, s) = Stream(fn () => Cons(x, s))

2
### 3.1 Examples

Here are some examples that construct values of type int stream:

\begin{verbatim}
(* zeros_f : unit -> int front *)
fun zeros_f ( ) = Cons(0, delay zeros_f)

(* zeros : int stream *)
val zeros = delay zeros_f

zeros is an infinite stream of 0s.

(* ones_f : unit -> int front *)
fun ones_f ( ) = Cons(1, delay ones_f)

(* ones : int stream *)
val ones = delay ones_f

Ones is an infinite stream of 1s.

(* streamify : 'a list -> 'a stream *)
fun streamify [ ] = empty
| streamify (x::L) = cons(x, streamify L)

The function streamify turns a list into a finite stream with the same elements as in the list (and in the same order).

(* natsFrom : int -> (unit -> int front) *)
fun natsFrom n ( ) = Cons(n, delay (natsFrom (n+1)))

(* nats : int stream *)
val nats = delay (natsFrom 0)

The function natsFrom constructs a suspension of a front, that represents the natural numbers from some number upward. Nats is the stream of all natural numbers, in ascending order.

(* lazyappend : 'a list * 'a stream -> 'a stream *)
fun lazyappend ([], s) = s
| lazyappend (x::xs, s) = cons(x, lazyappend(xs,s))

The function lazyappend appends a list to a stream, returning a stream of all the list elements followed by all the original stream elements.

(* repeat : 'a list -> 'a stream *)
(* REQUIRES: xs is not nil *)
(* ENSURES: take(expose(xs), n) terminates for all n >= 0 *)
(* (see definition of take below) *)
fun repeat xs = lazyappend(xs, delay(fn () => expose(repeat xs)))
\end{verbatim}
The function **repeat** expects a list and returns a stream that represents infinite appendes of the list to itself. For example, **repeat** \([1,2,3]\) is the stream consisting of \(1,2,3,1,2,3,\ldots\), repeating forever.

What happens if we evaluate **repeat** \([\ ]\)?

What happens if we evaluate **expose**(repeat \([\ ]\))? 

### 3.2 Viewing a Stream

Here is a useful function that shows us a “view” of a stream: we can look at the first \(n\) items in the stream, as an ordinary finite list.

\[
\begin{align*}
(* \ \text{take} & : 'a \ \text{stream} \ \ast \ \text{int} \ \rightarrow \ 'a \ \text{list} \\
\text{take'} & : 'a \ \text{front} \ \ast \ \text{int} \ \rightarrow \ 'a \ \text{list} \\
\text{REQUIRES:} & \ \text{n} \ \geq \ 0 \\
\text{ENSURES:} & \ \text{take(s,n)} \ \text{returns a list of the first n elements of s in order, unless s has fewer than n elements in which case take(s,n) raises Subscript. If computation of any of these n elements loops forever or raises an exception, then so will take(s,n).} \\
(*)
\end{align*}
\]

\[
\begin{align*}
\text{fun} & \ \text{take} \ (s, \ 0) = [] \\
\mid & \ \text{take} \ (s, \ \text{n}) = \text{take'} \ (\text{expose} \ s, \ \text{n}) \\
\text{and} & \ \text{take'} \ (\text{Empty}, \ _) = \text{raise} \ \text{Subscript} \\
\mid & \ \text{take'} \ (\text{Cons}(x,s), \ \text{n}) = x::\text{take}(s, \ \text{n}-1)
\end{align*}
\]

Example: \(\text{take(nats, 5)} = [0,1,2,3,4].\)

Note: “taking” the first few elements of a stream forces the first few suspensions to be applied to their argument ( ) to extract data from the tail of the stream. This is all purely functional, without side-effect. If we evaluate \(\text{take(nats, 2)}\) after \(\text{take(Nats, 5)}\) we will get the result \([0,1]\), not \([5,6]\). The first take doesn’t change the value of the stream.

Observe: The mutual recursion between \(\text{take}\) and \(\text{take'}\) is a useful template for writing stream functions: one function specializes to streams, the other to fronts, and they hand data back and forth between each other.

### 3.3 Example: Reals as Infinite Streams of Integers

A stream of integers could be used to represent a **real number**. We could use the head of the stream to represent the **integer part** of the real number, and then the rest of the stream to represent the decimal digits, with the most significant digit (the tenths part) first, then the hundredths part, and so on. According to this representation, \(\text{ones}\) represents the real number whose decimal expansion is \(1.11111\ldots\).

\[
\begin{align*}
\text{val} \ X & = \text{lazyappend} \ ([5, 8], \ \text{repeat} \ [1,4,4]) \\
(* \ X \ \text{represents the real number equal to 3227/555} (*) \\
(* \ This \ \text{number has decimal expansion 5.8144144144.} \ldots (*)
\end{align*}
\]

\[
\begin{align*}
\text{val} \ Y & = \text{repeat} \ [0,1,2,3,4,5,6,7,9] \\
(* \ Y \ \text{represents 10/81} (*)
\end{align*}
\]

4
(* The decimal expansion for 10/81 is 0.12345679012345679... *)

val Z = repeat [0,9]
(* Z represents 10/11 *)
(* The decimal expansion for 10/11 is 0.909090... *)

fun realDiv (x, y) : int stream =
  delay (fn () =>
    if x = 0 then Empty
    else
      let val d = x div y
      in
        if d > 9 then raise Fail "invalid input"
        else
          Cons (d, realDiv ((x - y * d) * 10, y))
      end)

3.4 Higher-Order Functions on Streams

The following functions appear within the Stream structure appearing in the code associated with today's lecture, as do the types and key functions presented so far. We omit the "Stream." prefix except when needed to distinguish from similarly named list functions.

- **map** : ('a -> 'b) -> 'a stream -> 'b stream
  map' : ('a -> 'b) -> 'a front -> 'b front

  fun map f s = delay (fn () => map' f (expose s))
  and map' f (Empty) = Empty
  | map' f (Cons(x, s)) = Cons(f(x), map f s)

If \( f: t_1 \rightarrow t_2 \) is total, and \( S: t_1 \) stream has elements \( d_0, d_1, \ldots, d_n, \ldots \), then \( \text{map} f S \) has type \( t_2 \) stream, and elements \( f(d_0), f(d_1), \ldots, f(d_n), \ldots \).

Again, notice the mutual recursion between a function that operates on streams and one that operates on fronts. Also, observe that \( \text{map} f s \) returns a delayed computation. Not until someone, if ever, evaluates \( \text{expose}(\text{map} f s) \) does any actual mapping occur, and then only to the first stream element, with further mapping of the rest of the stream itself delayed until requested via an \( \text{expose} \) of the rest of the mapped stream.

- **zip** : 'a stream * 'b stream -> ('a * 'b) stream
  zip' : 'a front * 'b front -> ('a * 'b) front

  fun zip (s1, s2) = delay (fn () => zip'(expose s1, expose s2))
  and zip' (_ , Empty) = Empty
  | zip' (Empty , _) = Empty
  | zip' (Cons(x, s1), Cons(y, s2)) = Cons((x, y), zip(s1, s2))
If $A$: $t_1$ stream has elements $a_0, a_1, ..., a_n, ...$
and $B$: $t_2$ stream has elements $b_0, b_1, ..., b_n, ...$,
then $\text{zip}(A, B)$ evaluates to a value $Z$: $(t_1 \times t_2)$ stream,
with elements $(a_0, b_0), (a_1, b_1), ..., (a_n, b_n), ...$.

- $\text{filter} : (\text{'a} \to \text{bool}) \to \text{'a stream} \to \text{'a stream}$

\[
\text{filter}' : (\text{'a} \to \text{bool}) \to \text{'a front} \to \text{'a front}
\]

\[
\begin{align*}
\text{fun filter } p \ s &= \text{delay (fn () => filter}' p \ (\text{expose } s)) \\
\text{and filter}' p \ (\text{Empty}) &= \text{Empty} \\
\mid \text{filter}' p \ (\text{Cons}(x, s)) &= \\
\quad \text{if } p(x) \text{ then } \text{Cons}(x, \text{filter } p \ s) \\
\quad\text{else filter}' p \ (\text{expose } s)
\end{align*}
\]

If $S$: $t$ filter and $p$: $t$->bool is total, $\text{filter } p \ S$ computes the stream of elements of $S$
that satisfy $p$.

If $S$ has an element satisfying $p$, and $p$ is total, then $\text{expose}(\text{filter } p \ S)$ returns a front
$\text{Cons}(x, s)$, where $x$ is the first element of $S$ that satisfies $p$ and $s$ is a stream representing
the remaining elements of $S$ satisfying $p$.

What happens if we evaluate $\text{expose}(\text{filter } (fn _ => false) \ \text{ones})$?

It is natural to see the first two functions as “extensions” of the corresponding functions on ordinary
lists, because one can prove the following results:

For all non-negative $n$, and all suitably typed values $f, S$ and $T$, with $f$ total:

\[
\begin{align*}
\text{take}(\text{Stream.map } f \ S, n) &= \text{List.map } f \ (\text{take}(S, n)) \\
\text{take}(\text{Stream.zip } (S, T), n) &= \text{List.zip } (\text{take}(S, n), \text{take}(T, n))
\end{align*}
\]

### 3.5 Equivalence of Streams

When should we say that two stream expressions are equivalent? In what sense are

\[
\begin{align*}
\text{Stream.map } (f \circ g) \ S \\
\text{Stream.map } f \ (\text{Stream.map } g \ S)
\end{align*}
\]
equivalent, assuming as usual that $f$ and $g$ are total functions whose types allow them to be composed?

A notion of equivalence that makes sense for streams is that stream values $X$ and $Y$ are extensionally equivalent iff, for all $n \geq 0$, $\text{take}(X, n)$ and $\text{take}(Y, n)$ are equivalent. This definition
also allows for the possibility that an attempt to extract an element may loop forever (as in the $\text{filter}$ example when there is no element satisfying the predicate). So two streams whose first 42 elements are equivalent and whose 43rd elements fail to terminate are also regarded as equivalent.

It is impossible to implement an equality-check for infinite streams, even for int streams. Unlike ordinary lists of integers, we cannot use the built-in $=$ operator of SML to check for equality. Instead, one generally must establish stream equivalence by proof.
3.6 Lazy Fusion Equivalences

Given the notion of stream equivalence above, the following lazy fusion results are provable:

For all types \( t_1, t_2, t_3 \), all total functions \( f : t_2 \rightarrow t_3 \) and \( g : t_1 \rightarrow t_2 \), and all values \( S : t_1 \) stream,

\[
\text{Stream.map } (f \circ g) \ S = \text{Stream.map } f \ (\text{Stream.map } g \ S).
\]

The proof proceeds by showing that (under the given assumptions) for all \( n \geq 0 \),

\[
\text{take}(\text{Stream.map } (f \circ g) \ S, n) = \text{take}(\text{Stream.map } f \ (\text{Stream.map } g \ S), n).
\]

In fact, we can appeal to the earlier result linking \( \text{Stream.map} \) and \( \text{List.map} \), and the familiar fusion property of \( \text{List.map} \), i.e., that

\[
\text{List.map } (f \circ g) (\text{take}(S, n)) = \text{List.map } f (\text{List.map } g (\text{take}(S, n)))
\]

to derive the lazy fusion equivalence stated above.

Note that we can paraphrase this lazy fusion equivalence, taking advantage of extensionality.

The following is an equivalent statement to the one given above:

For all types \( t_1, t_2, t_3 \), and all total functions \( f : t_2 \rightarrow t_3 \) and \( g : t_1 \rightarrow t_2 \),

\[
\text{Stream.map } (f \circ g) = (\text{Stream.map } f) \circ (\text{Stream.map } g).
\]

4 Examples

This section presents two examples illustrating how to program with streams. The art is in the design of thunks (often using recursion), and making the control flow sufficiently lazy that we never prematurely reveal the head-tail (i.e., front) structure of part of a stream until we need to.

4.1 Streams of Factorials

Useful facts about factorials:

- \( 0! = 1 \)
- \( k! \ast (k + 1) = (k + 1)! \) when \( k \geq 0 \).

The following lazy way to generate factorials is based on these useful facts. The first fact tells us that the first element of the stream should be 1. The second fact tells us that we can obtain the later factorials by zipping each integer \( k + 1 \) with the factorial of its predecessor and multiplying. So we embed a recipe for generating factorials starting from 0!, by zipping and multiplying, in the tail thunk.

\[
(*) \text{ factorials : unit } \rightarrow \text{ int front } *
\]

\[
\text{fun factorials } ( ) = \text{Cons}(1, \text{map } (\text{op } \ast) (\text{zip}(\text{delay factorials}, \text{delay}(\text{natsFrom 1}))))
\]

\[
(*) \text{ Factorials : int stream } *
\]

\[
\text{val Factorials = delay factorials}
\]
Factorials is a stream consisting of all factorials $n!$, for $n \geq 0$.
In other words, for all non-negative $n$, take(Factorials, $n+1$) evaluates to the list

$$[0!, 1!, \ldots, n!].$$

$\text{take(Factorials, 6) } \Rightarrow [1,1,2,6,24,120].$

### 4.2 Stream of Prime Numbers

This code is based on the Sieve of Eratosthenes, an ancient algorithm for enumerating the primes.

(* notDivides : int -> int -> bool
  REQUIRES: true
  ENSURES: notDivides p q ==> true iff p does not divide q, i.e.,
  q is not a multiple of p.
*)

fun notDivides p q = (q mod p <> 0)

(*
  sieve : int stream -> int stream
  sieve' : int front -> int front
*)

fun sieve s = delay (fn () => sieve' (expose s))
and sieve' (Empty) = Empty
  | sieve' (Cons(p, s)) = Cons(p, sieve (filter (notDivides p) s))

val primes = sieve(natsFrom 2)

primes is a stream representing all the prime numbers, in ascending order. Here are the first ten:

$\text{take(primes, 10) } \Rightarrow [2,3,5,7,11,13,17,19,23,29]$