lisp-lazy-seq

February 25, 2016

1 Lazy sequences in (a few lines of) Common Lisp

Copyright (C) 2016 F. Peschanski, CC-BY-SA 3.0

The Clojure programming language comes with a versatile <u>lazy sequence</u> abstraction, which also exists under other names such as <u>generators</u> in e.g. Python, (implicitly) <u>lazy lists</u> in e.g. Haskell, <u>streams</u> in e.g. Ocaml, etc.

There is nothing like in the Common Lisp standard, in particular the CL sequences are IMHO less interesting (providing a too restricted kind of polymorphism).

In this document I address the following question:

• is it easy to implement such an abstraction in Common Lisp?

And as it is almost always the case in Common Lisp, the answer is:

• yes it is!

And the remaining of this document is the way I would do it.

Disclaimer: the Lisp code below is not of production quality, and it is by no mean a port of the Clojure code, although I guess there aren't a million ways to implement the lazy sequence abstraction...

1.1 What is a lazy sequence?

Let's start with the basics.

1.1.1 A mathematical detour

According to https://en.wikipedia.org/wiki/Sequence,

a (mathematical) sequence is an ordered collection of objects in which repetitions are allowed.

Put in other terms, this is an ordered way to encode a <u>multiset</u>, and the simplest encoding is to associate to each element of the multiset a natural number giving its "position" in the ordering relation. The first element is at position 0, the second at position 1, etc. This gives a functional interpretation of a sequence. In this interpretation a sequence can have a <u>finite</u> or <u>infinite</u> domain, in which case we say that the sequence is finite or infinite.

There is also an inductive definition of a finite sequence, being the smallest set composed of either:

- the empty sequence nil
- or a non-empty sequence, constructed from a first element and the rest of the sequence.

You then obtained the definition of a list, our ubiquitous data structure (well, its proper, immutable variant)! The guarantee that the sequence is finite here is that we take the <u>smallest</u> set satisfying the rules. But what about infinite sequences?

There is indeed a related mathematical definition of an infinite sequence, being the <u>largest</u> set composed of

• a head elements followed by an infinite tail sequence

Note that there is no base case and moreover we consider this time the <u>largest</u> set satisfying the unique rule, touching infinity <u>for real</u>! Such definition is said <u>coinductive</u> and we will not go any deeper in this very interesting theory (if you're interested, cf. http://www.cs.unibo.it/~sangio/IntroBook.html).

Note that it is possible to mix the two definitions, by keeping the empty list and identifying first with head, and rest with tail.

So our short mathematical detour leads to the following Lisp code :

1.1.2 Strict vs. lazy sequences

Let's go back to computers and programming... In a computer everything's, in a way, is <u>finite</u> so the finite/infinite dichotomy is at best a view of the mind. We can perhaps more interestingly talk about <u>stict</u> vs. <u>lazy</u> sequences. Both are in fact finite, even though it is sometimes useful to consider lazy sequences as "infinite" (please <u>do</u> mind the double quotes).

A **strict sequence** follows the inductive definition above: it is either empty or non-empty providing a head value and a strict tail sequence. A consequence is that all the elements of strict sequences are computed values.

As long as you abolish setf-ing things, we can safely identify strict sequences with <u>lists</u>, hence the following Lisp definitions.

```
Out[7]: 1
In [8]: (tail '(1 2 3 4 5))
Out[8]: (2 3 4 5)
In [9]: (print-cell '(1 2 3 4 5) *standard-output*)
1 2 3 4 5
Out[9]: NIL
```

Now, a **lazy sequence** is designed after the coinductive case: we know that it has a head and a tail and since we want to be lazy, we do not want to say <u>too early</u> what exactly are the head and tail. In particular, the head is <u>not yet</u> considered a computed value until we need it, and the only thing we know about the tail is that is we need it, then it will also be a lazy sequence.

1.2 Strict sequences in Lisp

It is now time to get some actual code supporting the notions discussed above. First, we consider the case of strict sequences. Of course, implementing strict sequences is neither difficult nor increddibly usefull since, as we've seen previously, proper lists serve as quite a good representation of strict sequences. However, I think it is interesting to show the slight modifications involved when going from the strict to the lazy case.

According to the c2 wiki (cf. http://c2.com/cgi/wiki?ConsCell):

ConsCells are the building blocks of lists in LispLanguages.

Our version of a cons cell structure for strict sequences is as follows:

So a strict cell has two slots : a computed head ${\tt hd}$ and a strict tail ${\tt tl}$.

And much of what is already done for lisp can be done based on this simple structure.

We will stop <u>right now</u> to reimplement a list-like datastructure in Lisp: there's something disturbing about the whole <u>exercise!</u>

1.3 Lazy sequences in Lisp

We are now in a position of updating our strict cells to make them lazier.

A lazy cell is a specific kind of cell with too possible states:

- either it is <u>computed</u>, in which case the genfn slot is nil and its hd and tl slots (included from the cell structure) are directly accessible
- or it is <u>not yet computed</u> in which case both the hd and tl slots are nil, and the genfn slot contains a generator function.

Initially a lazy cell must be in the not yet computed state, that's the whole point of laziness!

Before asking for a head or tail, it is necessary to switch to the <u>computed</u> state, which happens thanks to the following function.

An important requirement is that the function above may only be called in the <u>not yet computed</u> state. The generator function is then called, which produce a value resulting from the cell computation. The state change is indicated by setting the **genfn** slot to nil.

The computation itself results in ... a sequence composed of a head and a tail. In many case we will have a cons pair (hence a kind of sequence) but it can be any kind of sequence distinct (of course distinct from the cell we are working on). For truely lazy sequences, the <u>head</u> should be a computed value, and the tail should be either empty or a (distinct) lazy cell. We will see how to fulfill these requirements below. For now, let's just remember that the extracted informations is stored in the two slots hd and t1.

Now we can define the head and tail methods in a straightforward way.

In both cases the first step is to check wether we are in the <u>computed state</u> (i.e. genfn is nil) or not. If not, the <u>compute-lazy-cell</u> function is called, and then the <u>value</u> of the hd (resp. tl) slot is returned.

And that is about all we need for manipulating lazy sequences!

1.4 Constructing lazy sequences

In order to construct lazy sequences, we can define a lazy variant of cons.

Note that we <u>have to</u> define it as a macro, otherwise we would compute both the hd and t1 parameters! The (lambda () ...) used as a generator function is used to delay the computation of the head and tail. Hence, of course, strict sequences are required to implement the lazy ones (we could have used cells also but let's use lists everywhere they work).

1.4.1 Constructing a finie sequence

Any finite sequence should be constructible in a lazy way, so let's try.

```
In [27]: (defparameter ls1 (lazy-cons (+ 1 0) (lazy-cons (+ 1 1) (lazy-cons (+ 1 2) nil))))
Out[27]: LS1
In [28]: LS1
Out[28]: #<lazy:?>
```

The output above confirms that nothing has been computed yet. The question mark marks (pun intended!) the <u>not yet computed</u> part of the sequence (everything, for now). Of course, taking either the head or the tail of the lazy sequence will trigger some computation.

```
In [29]: (head ls1)
Out[29]: 1
```

The computed value of the head is 1 as expected, and we can see the effect on the sequence itself.

```
In [30]: ls1
Out[30]: #<lazy:1 ?>
```

The computed part of the sequence now comprises the first element. Of course, we can dig a little deeper.

```
In [31]: (head (tail ls1))
Out[31]: 2
In [32]: ls1
Out[32]: #<lazy:1 2 ?>
```

To make the sequence fully computed, we need to dig a little deeper.

```
In [33]: (tail (tail (tail ls1)))
Out[33]: NIL
In [34]: ls1
Out[34]: #<lazy:1 2 3>
```

Note that the list is still considered $\underline{\text{lazy}}$ but it is in computed state now so there's no computation to delay anymore.

One important take away of the previous example is that laziness can only obtained through side effects, however very controlled (and overally "safe") ones.

1.4.2 Constructing an "infinite" sequence

Once computed a sequence must of course be finite, but the lazy sequence abstraction allows to define <u>intentionally inifinite</u> sequences. A good example is the sequence of natural numbers. Let's try a definition using the lazy-cons macro.

In this state, only the finite prefix 0 1 2 of the lazy sequence has been computed. To dig deeper, we can start to implement some utility functions.

The first one consists in taking a finite prefix of the sequence, materializing it (of course!) as a list.

Note that since we keep a reference to the first cell (with head 0), the more we consume informations in the LS2 sequence the more we increase the amount of memory required to store the sequence cells. There is a Clojure proverb saying:

don't hold to your head!

In [52]: (defun iterate (f x)

In [53]: (take 10 (iterate #'1+ 1))

Out[53]: (1 2 3 4 5 6 7 8 9 10)

Out [52]: ITERATE

The example above is a good illustration of what happens in this case. One way to skip unwanted cells is to use the **drop** function defined below:

```
In [45]: (defun drop (n s)
              "Drops the first 'n' elements of the sequence 's'."
              (loop repeat n
                    for cell = s then (tail cell)
                    when (not cell) do (return cell)
                    finally (return cell)))
Out [45]: DROP
  Let's try to drop a few natural numbers.
In [46]: (drop 10 (nats 0))
Out[46]: #<lazy:?>
   The result is a lazy sequence without any computed cell... Of course, it's more useful if we take some
information from the sequence:
In [47]: (head (drop 10 (nats 0)))
Out [47]: 9
In [48]: (take 10 (drop 10000 (nats 1)))
Out [48]: (10000 10001 10002 10003 10004 10005 10006 10007 10008 10009)
   Of course, be careful not holding to your head as in the take case.
In [49]: (drop 100 ls2)
Out[49]: #<lazy:?>
In [50]: 1s2
Out [50]: #<lazy:0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 3
1.4.3 The lazy-seq macro
Clojure does not provide (anymore) a lazy-cons macro, but rather allows to mix strict and lazy sequences
using a simple albeit useful lazy-seq macro.
   The way I understand it, is simply that the cons part remains implicit, which gives:
In [51]: (defmacro lazy-seq (expr)
              '(make-lazy-cell :hd nil :tl nil :genfn (lambda () ,expr)))
Out [51]: LAZY-SEQ
  The consing part is now the responsability of the expr argument of lazy-seq.
  Let's port the iterate function of the Clojure standard library using this macro.
```

(lazy-seq (cons x (iterate f (funcall f x)))))

1.5 The three stooges: map, filter and reduce

Lazy sequences, as you can see, are not difficult to introduce in Common Lisp. And thanks to the lazy-seq macro it is rather easy to write generator functions such as iterate above. We will also see that sequences are an adequate tool to perform traversals of structured data.

But once the sequences have been created, what we want to do is to manipulate them. The three most important manipulations are: mappings, filterings and reductions. The corresponding functions are only a few lines aways.

Because there is already a map in Common Lisp (although not $\underline{\text{that}}$ useful), let call our version maps for mapping sequence.

```
In [54]: (defun maps (f s)
             (when s
                 (lazy-seq (cons (funcall f (head s)) (maps f (tail s))))))
Out [54]: MAPS
In [55]: (take 10 (maps (lambda (x) (* x 10)) (nats 1)))
Out [55]: (10 20 30 40 50 60 70 80 90 100)
  Filtering is not much more difficult.
In [56]: (defun filters (pred s)
             (when s
                 (if (funcall pred (head s))
                      (lazy-seq (cons (head s) (filters pred (tail s))))
                      (filters pred (tail s)))))
Out [56]: FILTERS
In [57]: (take 10 (filters #'evenp (nats 1)))
Out[57]: (2 4 6 8 10 12 14 16 18 20)
In [58]: (take 10 (filters #'oddp (nats 1)))
Out[58]: (1 3 5 7 9 11 13 15 17 19)
```

The reduction function is very powerfull, although unlike the map and filter cases, it requires the whole sequence to be computed. Efficiency is prominent here, and we have to be careful not feeding the reduction with an infinite sequence.

1.6 The final word ...

In a few lines of Lisp we were able to implement (a prototypal version of) the sequence abstraction (through the head and tail generic functions) and their lazy representation. The <u>lisp-lazy-seq</u> library contains much more facilities for your enjoyment but you now know how all this works.

```
That's all folks!
```

```
In []:
```