Chapter 7: Computations lifted to a functor context II. Monads

Part 2: Laws and structure of semimonads

Sergei Winitzki

Academy by the Bay

2018-03-11

Semimonad laws I: The intuitions

What properties of functor block programs do we expect to have?

- In $x \leftarrow c$, the value of x will go over items held in container c
- Manipulating items in container is followed by a generator:

```
x \leftarrow cont1
                                                                v \leftarrow cont1
      y = f(x)
                                                                         .map(x \Rightarrow f(x))
      z \leftarrow cont2(y)
                                                                z \leftarrow cont2(y)
cont1.flatMap(x \Rightarrow cont2(f(x))) = cont1.map(f).flatMap(y \Rightarrow cont2(y))
```

Manipulating items in container is preceded by a generator:

```
x \leftarrow cont1
                                                       x \leftarrow cont1
      y \leftarrow cont2(x)
                                                       z \leftarrow cont2(x)
      z = f(v)
                                                                  .map(f)
cont1.flatMap(cont2).map(f) = cont1.flatMap(x \Rightarrow cont2(x).map(f))
```

• Within a generator, for {...} yield can be inlined:

```
x \leftarrow cont
                                                              vv \leftarrow for \{ x \leftarrow cont \}
y \leftarrow p(x)
                                                                                   y \leftarrow p(x) } yield y
z \leftarrow cont2(y)
                                                              z \leftarrow cont2(yy)
```

```
cont.flatMap(x \Rightarrow p(x).flatMap(cont2)) = cont.flatMap(p).flatMap(cont2)
```

Semimonad laws II: The laws for flatMap

For brevity, write flm instead of flatMap

A semimonad S^A has $flm^{[A,B]}: (A \Rightarrow S^B) \Rightarrow S^A \Rightarrow S^B$ with 3 laws:

fmap
$$f^{A\Rightarrow B}$$
 S^B flm $g^{B\Rightarrow S^C}$

$$S^A \Longrightarrow S^C$$
flm $(f^{A\Rightarrow B} \circ g^{B\Rightarrow S^C})$

2 flm $(f^{A \Rightarrow S^B} \circ \operatorname{fmap} g^{B \Rightarrow C}) = \operatorname{flm} f \circ \operatorname{fmap} g$ (naturality in B)

$$S^{A} \xrightarrow{\text{flm } f^{A \Rightarrow S^{B}}} S^{B} \xrightarrow{\text{fmap } g^{B \Rightarrow C}} S^{C}$$

$$flm (f^{A \Rightarrow S^{B}} \circ fmap g^{B \Rightarrow C})$$

3 $\operatorname{flm}(f^{A\Rightarrow S^B} \circ \operatorname{flm} g^{B\Rightarrow S^C}) = \operatorname{flm} f \circ \operatorname{flm} g$ (associativity)

$$S^{A} \xrightarrow{\text{flm } f^{A \Rightarrow S^{B}}} S^{B} \xrightarrow{\text{flm } g^{B \Rightarrow S^{C}}} S^{C}$$

Is there a shorter and clearer formulation of these laws?

Semimonad laws III: The laws for flatten

The methods flatten (denoted by ftn) and flatMap are equivalent:

$$\operatorname{flm}(f^{A]}: S^{S^{A}} \Rightarrow S^{A} \equiv \operatorname{flm}^{\left[S^{A}, A\right]}(m^{S^{A}} \Rightarrow m)$$

$$\operatorname{flm}(f^{A \Rightarrow S^{B}}) \equiv \operatorname{fmap} f \circ \operatorname{ftn}$$

$$S^{A} \xrightarrow{\operatorname{flm}(f^{A \Rightarrow S^{B}})} S^{S^{D}}$$

It turns out that flatten has only 2 laws:



2 fmap $(ftn^{[A]}) \circ ftn^{[A]} = ftn^{[S^A]} \circ ftn^{[A]}$ (associativity)



Equivalence of a natural transformation and a "lifting"

- Equivalence of flm and ftn: ftn = flm (id); flm $f = \text{fmap } f \circ \text{ftn}$
- We saw this before: deflate = fmapOpt(id); $fmapOpt f = fmap f \circ deflate$
 - ▶ Is there a general pattern where two such functions are equivalent?
- Let $tr: F^{G^A} \Rightarrow F^A$ be a natural transformation (F and G are functors)
- Define ftr: $(A \Rightarrow G^B) \Rightarrow F^A \Rightarrow F^B$ by ftr $f = \operatorname{fmap} f \circ \operatorname{tr}$
- It follows that tr = ftr(id), and we have equivalence between tr and ftr:

$$\operatorname{tr}: F^{G^A} \Rightarrow F^A = \operatorname{ftr}(m^{G^A} \Rightarrow m)$$

$$\operatorname{ftr}(f^{A \Rightarrow G^B}) = \operatorname{fmap} f \circ \operatorname{tr}$$

$$f^A \xrightarrow{\operatorname{ftr}(f^{A \Rightarrow G^B})} F^B$$

- An automatic law for ftr ("naturality in A") follows from the definition: fmap $g \circ \text{ftr } f = \text{fmap } g \circ \text{fmap } f \circ \text{tr} = \text{fmap } (g \circ f) \circ \text{tr} = \text{ftr } (g \circ f)$
 - ► This is why tr always has one law fewer than ftr
- To demonstrate equivalence in the direction ftr → tr: Start with an arbitrary ftr satisfying "naturality in A", then obtain tr = ftr (id) from it, then verify ftr f = fmap f ∘ tr with that tr; fmap f ∘ ftr (id) = ftr (f ∘ id) = ftr f

Semimonad laws IV: Deriving the laws for flatten

Denote for brevity $q^{\uparrow} \equiv \text{fmap } q$ for any function q ("lifting" $q^{A \Rightarrow B}$ to S) Express flm $f = f^{\uparrow} \circ$ ftn and substitute that into flm's 3 laws:

- flm $(f \circ g) = f^{\uparrow} \circ \text{flm } g \text{ gives } (f \circ g)^{\uparrow} \circ \text{ftn} = f^{\uparrow} \circ g^{\uparrow} \circ \text{ftn}$ - this law holds automatically due to functor composition law
- ② flm $(f \circ g^{\uparrow}) = \text{flm } f \circ g^{\uparrow} \text{ gives } (f \circ g^{\uparrow})^{\uparrow} \circ \text{ftn} = f^{\uparrow} \circ \text{ftn} \circ g^{\uparrow};$ using the functor composition law, we reduce this to $g^{\uparrow\uparrow} \circ \text{ftn} = \text{ftn} \circ g^{\uparrow} - \text{this is the naturality law}$
- § $flm(f \circ flm g) = flm f \circ flm g$ with functor composition law gives $f^{\uparrow} \circ g^{\uparrow\uparrow} \circ \operatorname{ftn}^{\uparrow} \circ \operatorname{ftn} = f^{\uparrow} \circ \operatorname{ftn} \circ g^{\uparrow} \circ \operatorname{ftn}$; using ftn's naturality and omitting the common factor $f^{\uparrow} \circ g^{\uparrow\uparrow}$, we get $ftn^{\uparrow} \circ ftn = ftn \circ ftn - associativity law$
 - flatten has the simplest type signature and the fewest laws
 - It is usually easy to check naturality!
 - ▶ Parametricity theorem: Any pure, fully parametric code for a function of type $F^A \Rightarrow G^A$ will implement a natural transformation
- Checking flatten's associativity needs a lot more work!

The cats library has a FlatMap type class, defining flatten via flatMap

Checking the associativity law for standard monads

- Implement flatten for these functors and check the laws (see code):
 - ▶ Option monad: $F^A \equiv 1 + A$; ftn : $1 + (1 + A) \Rightarrow 1 + A$
 - ▶ Either monad: $F^A \equiv Z + A$; ftn : $Z + (Z + A) \Rightarrow Z + A$
 - ▶ List monad: $F^A \equiv \text{List}^A$; ftn : List List $\Rightarrow \text{List}^A$
 - ▶ Writer monad: $F^A \equiv A \times W$; ftn : $(A \times W) \times W \Rightarrow A \times W$
 - ▶ Reader monad: $F^A \equiv R \Rightarrow A$; ftn : $(R \Rightarrow (R \Rightarrow A)) \Rightarrow R \Rightarrow A$
 - ▶ State: $F^A \equiv S \Rightarrow A \times S$; ftn : $(S \Rightarrow (S \Rightarrow A \times S) \times S) \Rightarrow S \Rightarrow A \times S$
 - ► Continuation monad: $F^A \equiv (A \Rightarrow R) \Rightarrow R$; ftn : $((((A \Rightarrow R) \Rightarrow R) \Rightarrow R) \Rightarrow (A \Rightarrow R) \Rightarrow R$
- Code implementing these flatten functions is fully parametric in A
 - ▶ Naturality of these functions follows from parametricity theorem
 - Associativity needs to be checked for each monad!
- Example of a useful semimonad that is *not* a full monad:
 - $F^A \equiv A \times V \times W; \text{ ftn } ((a \times v_1 \times w_1) \times v_2 \times w_2) = a \times v_1 \times w_2$
- Examples of *non-associative* (i.e. wrong) implementations of flatten:
 - $F^A \equiv A \times W \times W; \text{ ftn} ((a \times v_1 \times v_2) \times w_1 \times w_2) = a \times w_2 \times w_1$
 - $ightharpoonup F^A \equiv \text{List}^A$, but flatten concatenates the nested lists in reverse order

Motivation for monads

- Monads represent values with a "special computational context"
- Specific monads will have methods to create various contexts
- Monadic composition will "combine" the contexts associatively
- It is generally useful to have an "empty context" available:

pure :
$$A \Rightarrow M^A$$

Adding the empty context to another context should be a no-op

Empty context is followed by a generator:

```
y \leftarrow pure(x)
                                                           v = x
                                                           z \leftarrow cont(y)
      z \leftarrow cont(y)
                                                          pure \circ flm f = f – left identity
pure(x).flatMap(y \Rightarrow cont(y)) = cont(x)
```

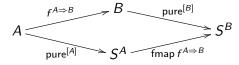
Empty context is preceded by a generator:

```
x \leftarrow cont
                                                               x \leftarrow cont
y \leftarrow pure(x)
                                                               v = x
```

flm(pure) = id - right identity $cont.flatMap(x \Rightarrow pure(x)) = cont$

The monad laws formulated in terms of pure and flatten

Naturality law for pure: f ∘ pure = pure ∘ fmap f



Left identity: pure ∘ flm f = pure ∘ fmap f ∘ ftn = f ∘ pure ∘ ftn = f
 requires that pure ∘ ftn = id (both sides applied to S^A)

$$S^{A} \xrightarrow{\text{pure}^{[S^{A}]}} S^{S^{A}} \xrightarrow{\text{ftn}^{[A]}} S^{A}$$

• Right identity: $\mathsf{flm}(\mathsf{pure}) = \mathsf{fmap}(\mathsf{pure}) \circ \mathsf{ftn} = \mathsf{id}^{S^A \Rightarrow S^A}$



Formulating laws via Kleisli functions

- Recall: we formulated the laws of filterables via fmapOpt
 - type signature of fmapOpt : $(A \Rightarrow 1 + B) \Rightarrow S^A \Rightarrow S^B$
 - ▶ and then we had to compose functions of types $A \Rightarrow 1 + B$ via \diamond_{Ont}
- Here we have flm : $(A \Rightarrow S^B) \Rightarrow S^A \Rightarrow S^B$ instead of fmapOpt
- Can we compose **Kleisli functions** with "twisted" types $A \Rightarrow S^B$?
- Use flm to define Kleisli composition: $f^{A\Rightarrow S^B} \diamond g^{B\Rightarrow S^C} \equiv f \circ \text{flm } g$
- Define **Kleisli identity** id_{\diamond} of type $A \Rightarrow S^A$ as $id_{\diamond} \equiv pure$
- Composition law: $flm(f \diamond g) = flm f \circ flm g$ (same as for fmapOpt)
 - ▶ Shows that flatMap is a "lifting" of $A \Rightarrow S^B$ to $S^A \Rightarrow S^B$
- These laws are similar to functor "lifting" laws...
 - ▶ except that ⋄ is used for composing Kleisli functions
- What are the properties of <?</p>
 - \triangleright Exactly similar to the properties of function composition $f \circ g$

Reformulate flm's laws in terms of the \diamond operation:

- flm's left and right identity laws: pure $\diamond f = f$ and $f \diamond$ pure = f
- Associativity law: $(f \diamond g) \diamond h = f \diamond (g \diamond h)$
 - ▶ Follows from the flm law: $f \circ \text{flm}(g \circ \text{flm}h) = f \circ \text{flm} g \circ \text{flm} h$

* Motivation for categories and functors

Compare different "liftings" seen so far, and generalize

Category	Type $A \rightsquigarrow B$	Identity	Composition
plain functions	$A \Rightarrow B$	$id: A \Rightarrow A$	$f^{A\Rightarrow B}\circ g^{B\Rightarrow C}$
lifted to F	$F^A \Rightarrow F^B$	$id: F^A \Rightarrow F^A$	$f^{F^A \Rightarrow F^B} \circ g^{F^B \Rightarrow F^C}$
Kleisli over F	$A \Rightarrow F^B$	pure : $A \Rightarrow F^A$	$f^{A\Rightarrow F^B} \diamond g^{B\Rightarrow F^C}$

Category theory generalizes this situation

Category: a certain class of "twisted functions" $A \rightsquigarrow B$ called **morphisms**

- For any two morphisms $f^{A \leadsto B}$ and $g^{B \leadsto C}$ the **composition** morphism $f \diamond g$ of type $A \leadsto C$ must exist
- For each type A, the **identity** morphism id_{\diamond} of type $A \rightsquigarrow A$ must exist
- Composition respects identity: $id_{\diamond} \diamond f = f$ and $f \diamond id_{\diamond} = f$
- Composition is associative: $(f \diamond g) \diamond h = f \diamond (g \diamond h)$

General functor: a map from one category to another

- A functor must fmap each morphism from one category to the other
- Functor laws: fmap must preserve identity and composition
 - ▶ What we call "functor" is called **endofunctor** in category theory
- ► An endofunctor's fmap goes from plain functions to F-lifted functions

 Sergei Winitzki (ABTB) Chapter 7: Functor-lifted computations || 2018-03-11 || 11

* From Kleisli back to flatMap

The Kleisli functions, $A \rightsquigarrow B \equiv A \Rightarrow S^B$, form a category iff S is a monad

- map and flatMap are computationally equivalent to Kleisli composition:
 - ▶ Define flatMap through Kleisli: flm $f^{A\Rightarrow S^B} \equiv id^{S^A\Rightarrow S^A} \diamond f$
 - ► Require two additional laws that connect ⋄, fmap, and ⋄:
 - ★ Left naturality: $f^{A\Rightarrow B} \circ g^{B\Rightarrow S^c} = (f \circ pure) \diamond g$
 - ★ Right naturality: $f^{A\Rightarrow S^B} \circ \operatorname{fmap} g^{B\Rightarrow C} = f \diamond (g \circ \operatorname{pure})$
 - ▶ So, can define fmap through Kleisli: fmap $g^{A\Rightarrow B} \equiv \operatorname{id}^{S^A\Rightarrow S^A} \diamond (g \circ \operatorname{pure})$

The laws for pure and flatMap then follow from category axioms for Kleisli:

- Left and right identity laws follow from id \diamond pure = id and pure \diamond f = f
 - Associativity for flatMap follows from $(id \diamond f) \diamond g = id \diamond (f \diamond g)$
 - Use "left naturality", get: $(f \circ g) \diamond h = (f \circ pure) \diamond g \diamond h = f \circ (g \diamond h)$
 - Naturality for pure: pure \circ fmap $f = \text{pure} \diamond (f \circ \text{pure}) = f \circ \text{pure}$
 - Define flatten: $\mathsf{ftn} = \mathsf{id}^{S^S} \Rightarrow S^S$ $\diamond \mathsf{id}^{S^A} \Rightarrow S^A$
 - Naturality for flatten: $ftn \circ fmap f = id \diamond id \diamond (f \circ pure) = id \diamond fmap f$ and $fmap (fmap f) \circ ftn = id \diamond ((fmap f) \circ pure) \circ id \diamond id = id \diamond fmap f$

Structure of semigroups and monoids

- Semimonad contexts are combined associatively, as in a semigroup
 - ▶ A full monad includes an "empty" context, i.e. the identity element
 - Semigroup with an identity element is a monoid

Some constructions of semigroups and monoids (see code):

- **1** Any type Z is a semigroup with operation $z_1 \circledast z_2 = z_1$ (or z_2)
- ② 1+S is a monoid if S is (at least) a semigroup (or $S\equiv 0$)
- **3** List^A is a monoid (for any type A), also Seq^A etc.
- The function type $A \Rightarrow A$ is a monoid (for any type A)
 - ▶ The operation $f \circledast g$ can be either $f \circ g$ or $g \circ f$
- ullet Any totally ordered type is a monoid, with \circledast defined as max or min
- $\circ S_1 + S_2$ is a semigroup (monoid) if S_1 , S_2 are semigroups (monoids)
- M[S] is a monoid if M[_] is a monad and S is a monoid.
- - ▶ The "action" is $\alpha: S \Rightarrow P \Rightarrow P$ such that $\alpha(s_2) \circ \alpha(s_1) = \alpha(s_1 \circledast s_2)$.
 - ▶ $S \times P$ is a "twisted product." Examples: $(A \Rightarrow A) \times A$; Bool \times (1 + A).
 - Other examples of monoids: Int (many), String, Set^A, Akka's Route

Structure of (semi)monads

How to recognize a (semi)monad by its type? Open question!

Intuition from flatten: reshuffle data in F^{FA} to fit into F^{A} Some constructions of exponential-polynomial (semi)monads:

- $F^A \equiv Z$ (constant functor) for a fixed type Z
 - For a full monad, need to choose Z=1
- $F^A \equiv A \times G^A$ for any functor G^A (a full monad only if $G^A \equiv 1$)
- $F^A \equiv G^A \times H^A$ for any (semi)monads G^A and H^A
 - but $G^A + H^A$ is generally *not* a semimonad
- \bullet $F^A \equiv R \Rightarrow G^A$ is a (semi)monad for any (semi)monad G^A
- $F^A \equiv A + G^A$ is a monad for a monad G^A (free pointed over G)
- **6** $F^A \equiv G^{Z+A\times W}$ is a monad if G is a monad and W a monoid
- $F^A \equiv A + G^{F^A}$ (recursive) for any functor G^A (free monad over G) Semimonad-only constructions:
- § $F^A \equiv G^A + G^{F^A}$ (recursive) for any functor G^A
- - ▶ Obtain a full monad only when $G^A \equiv 1$, i.e. $F^A \equiv H^A \Rightarrow A$

Exercises II

- Show that M[S] is a monoid if M[_] is a monad and S is a monoid.
- ② A framework implements a "route" type R as $R \equiv Q \Rightarrow (E+S)$, where Q is a query, E is an error response, and S is a success response. A server is defined as a "sum" of several routes. For a given query Q, the response is the first route (if it exists) that yields a success. Implement the route "summation" operation and show that it makes R into a semigroup. What would be necessary to make R into a monoid?
- **3** Verify the associativity law for the semimonad $F^A \equiv Z + \text{Bool} \times A$.
- **③** Show that the functor $F^A \equiv \text{Boolean} \times M^A$ (where M^A is an arbitrary monad) can be made into a semimonad but not into a monad.
- **9** If W and R are arbitrary fixed types, which of the functors can be made into a semimonad: $F^A \equiv W \times (R \Rightarrow A)$, $G^A = R \Rightarrow (W \times A)$?
- Show that $F^A \equiv (P \Rightarrow A) + (Q \Rightarrow A)$ is not a semimonad (cannot define flatMap) when P and Q are arbitrary, different types.
- Implement the flatten and pure methods for $D^A \equiv 1 + A \times A$ (type D[A] = Option[(A, A)]) in at least two significantly different ways, and show that the monad laws always fail to hold. (D^A is not a monad!)

Exercises II (continued)

8 A programmer implemented the fmap method for $F^A \equiv A \times (A \Rightarrow Z)$ as

```
def fmap[A,B](f: A\RightarrowB): ((A, A\RightarrowZ)) \Rightarrow (B, B\RightarrowZ) =
   { case (a, az) \Rightarrow (f(a), (_: B) \Rightarrow az(a)) }
```

Show that this implementation fails to satisfy the functor laws.

- **9** Show that $P^A \equiv Z + W \times A$ is a (full) monad if W is a monoid.
- Verify that the full monad laws hold for construction 4.
- Implement flatten and pure for $F^A \equiv A + (R \Rightarrow A)$, where R is a fixed type, and show that all the monad laws hold.
- For construction 5, show that an identity law would fail if pure were defined as $a \Rightarrow Right(Monad[G].pure(a))$ instead of as Left(a).
- Implement the monad methods for $F^A \equiv (Z \Rightarrow 1 + A) \times \text{List}^A$ using the known monad constructions (no need to check the laws).
- Implement the semimonad construction 2 by discarding the first effect (not the second), and show that the associativity law is still satisfied.
- For semimonad construction 8, show that the associativity law holds.
- Verify the identity laws for the State and Continuation monads.