Dyads, a generalisation of monads

Maarten Fokkinga

University of Twente, Dept INF, PO Box 217, NL 7500 AE Enschede e-mail: fokkinga@cs.utwente.nl

Version of December 6, 1993

The concept of dyad is defined as the least common generalisation of monads and co-monads. So, taking some of the ingredients to be the identity, the concept specialises to the concept of monad, and taking other ingredients to be the identity it specialises to co-monads. Except for one axiom, all have a nice ("natural") form.

Introduction: monads

Let us first describe one way in which monads can be motivated. We shall motivate dyads by a similar argument later. (Other descriptions have been given by Barr and Wells [1] and Wadler [2].) We use a fairly standard notation: $\mathcal{A}, \mathcal{B}, \ldots$ denote categories, a, b, \ldots objects, f, g, \ldots arrows, F, G, \ldots functors, greek letters denote natural transformations, and x, y, \ldots various things. Composition is denoted in diagrammatic order: f : g is normally written $g \circ f$.

Example. Let L be the list functor: La is the set of lists over a, and for $f: a \to b$ we have $Lf: La \to Lb$ as the well-known f-map. Given list producing functions $f: a \to Lb$ and $g: b \to Lc$, we often see the list producing "composition":

$$f, Lq, \#/: a \to Lc$$
.

Here $\#/: LL \to L$ is the flattening, or concatenation, of lists of list into lists. Functions of type $a \to Lb$, for varying a, b, turn up frequently in actual programming. For example, unconditional list comprehensions can be described in that way:

$$x \mapsto [z \mid y \leftarrow f x; \ z \leftarrow g y] = f, Lg, \#/$$
.

Notice, moreover, the existence of a particularly nice list producing function of type $a \to Lb$ with a = b, namely the singleton former $\eta = \lambda x :: [x]$. It has the property that $f : L\eta : \#/= f = \eta : Lf : \#/= f$.

Generalisation. The situation above can be described more elegantly in categorical terms as follows. Let F be an endofunctor. Under what conditions can we consider arrows of type $f: a \to Fb$, for varying a, b, to be arrows of type $a \to b$ in another category? This question means, amongst others, that there must be a way to "compose" arrows $f: a \to Fb$ and $g: b \to Fc$ into an arrow of type $a \to Fc$, and that this composition is associative, and that there exists a function $\eta_a: a \to Fa$ for each a that is the identity for the new "composition".

The new category is known as the Kleisli category, and we shall indicate it by $\mathcal{K}(F)$, or simply \mathcal{K} if the functor F is understood. The construction of \mathcal{K} is straightforward. Define:

```
a object in \mathcal{K} \equiv a object in the given category f arrow in \mathcal{K} \equiv f arrow in the given category of type a \to Fb for some a, b f \colon a \to_{\mathcal{K}} b \equiv f \colon a \to Fa f \colon_{\mathcal{K}} g \equiv f \colon_{\mathcal{F}} g \colon_{\mathcal{H}} \text{ whenever } f \colon_{\mathcal{A}} a \to Fb, \quad g \colon_{\mathcal{B}} b \to_{\mathcal{F}} c id_{\mathcal{K},a} \equiv \eta_a
```

where natural transformations μ and η are assumed to exist:

$$\begin{array}{cccc} \mu & : & FF \to F \\ \eta & : & I \to F \end{array}.$$

The associativity of composition $_{,\kappa}$, and the neutrality of id_{κ} for this composition, are equivalent to the following three properties of μ and η :

The first two of these say that $F\mu$ equals μF , and $F\eta$ equals ηF , when followed by μ . Indeed, a proof of associativity reads:

$$(f_{;\kappa} g)_{;\kappa} h = f_{;\kappa} (g_{;\kappa} h)$$

$$\equiv \text{ definition }_{;\kappa}$$

$$(f_{;\kappa} F g_{;\mu})_{;\kappa} F h_{;\mu} = f_{;\kappa} F (g_{;\kappa} F h_{;\mu})_{;\mu}$$

$$\equiv \text{ category, functor}$$

$$f_{;\kappa} F g_{;\mu}_{;\kappa} F h_{;\mu} = f_{;\kappa} F g_{;\kappa} F F h_{;\kappa} F \mu_{;\mu}$$

$$\equiv \text{ in lhs: naturality } \mu$$

$$f_{;\kappa} F g_{;\kappa} F F h_{;\mu} F_{;\mu} = f_{;\kappa} F g_{;\kappa} F F h_{;\kappa} F \mu_{;\mu}$$

$$\equiv \text{ for } \Leftarrow : \text{ Leibniz; for } \Rightarrow : \text{ instantiate with } f, g, h \text{ all equal to } id$$

$$\mu F_{;\kappa} \mu = F \mu_{;\kappa} \mu_{;\kappa}.$$

A proof of neutrality of $id_{\mathcal{K}}$ for $_{\mathcal{K}}$ reads:

The three remaining category axioms for K read:

The first two of these are evidently true; the latter one isn't true in general. So the construction yields a pre-category, and by a standard construction one obtains a category, the Kleisli category \mathcal{K} .

In summary, the ingredients that enable the construction of a Kleisli category are:

$$\begin{array}{lll} (F, & \mu \hbox{:}\; FF \to F, & \eta \hbox{:}\; Id \to F) \\ \text{satisfying} \\ F\mu \ ; \mu & = & \mu F \ ; \mu \\ F\eta \ ; \mu & = & \eta F \ ; \mu & = & id F \end{array} \ .$$

Such a triple is known as a **monad**.

A generalisation wanted

Loosely formulated the above construction of the Kleisli category enables us to "compose" arrows of type $a \to Fb$ (considering them to be of type $a \to' b$) for fixed F and varying a,b. Doaitse Swierstra posed the problem of "composing" arrows of type $a \times b \to Fc$, for fixed F and varying a,b,c. The Kleisli construction doesn't help here.

Abstracting a little bit, the new problem is to construct a Kleisli-like category for arrows of type $Fa \to Gb$ (considering them to be of type $a \to' b$), for fixed F, G and varying a, b. Taking F = Id we get the original problem that is solved by Kleisli's construction and the existence of a monad for G. Taking G = Id we get the dual problem; solved by the dual of Kleisli's construction and the existence of a co-monad for F. Thus our current problem formulation is the least common generalisation of the "Kleisli problem" and its dual; and our solution will be the least common generalisation of the Kleisli construction and its dual. We will term the ingredients of the solution a **dyad**: the least common generalisation of a monad and a co-monad.

Dyads

Let F, G be functors with a common source and a common target. We will construct a category $\mathcal{D}(F, G)$, or simply \mathcal{D} , whose arrows (of type $a \to_{\mathcal{D}} b$) are the given arrows of type $Fa \to Gb$. To this end we define:

```
a 	ext{ object in } \mathcal{D} \equiv a 	ext{ object in source category of } F, G
f 	ext{ arrow in } \mathcal{D} \equiv f 	ext{ arrow in target category of } F, G
f 	ext{ of type } Fa \to Gb 	ext{ for some } a, b
f 	ext{: } a \to_{\mathcal{D}} b \equiv f 	ext{: } Fa \to Gb
f 	ext{:}_{\mathcal{D}} g \equiv \mu 	ext{: } Ff 	ext{: } \gamma 	ext{: } Gg 	ext{: } \nu 	ext{ (explained below)}
id_{\mathcal{D}} \equiv \eta 	ext{.}
```

where natural transformations γ, η, μ, ν are assumed to exist:

 $\begin{array}{lll} \gamma & : & FG \to GF & F,G\text{-commuting transformation} \\ \eta & : & F \to G & F\text{- to }G\text{-unit transformation} \\ \mu & : & F \to FF & F\text{-generating transformation} \\ \nu & : & GG \to G & G\text{-reducing transformation} \end{array}.$

Notice that G and ν play the role of F and μ in the usual nomenclature for monads and the Kleisli construction. The above definition of $f_{:\mathcal{D}} g$ is the simplest general way to combine f and g in that order into an arrow of type $Fa \to Gc$. Indeed, in order that a transformed f and transformed f both have the same ingredients in their target and source type, respectively, f has to be applied to f and f to f this gives f this gives f and f and f and f are f this gives f and f are f the f and f are f and f are f and f are f and f and f are f and f are f are f are f are f and f are f and f are f are f are f and f are f are f and f are f are f are f and f are f are f are f are f and f are f are f are f and f are f are f are f and f are f are f are f are f and f are f are f are f are f and f are f are f and f are f are f are f are f are f are f and f are f ar

$$Fa \stackrel{?}{\longrightarrow} FFa \stackrel{Ff}{\longrightarrow} FGb \stackrel{?}{\longrightarrow} GFb \stackrel{Gg}{\longrightarrow} GGb \stackrel{?}{\longrightarrow} Gb$$

Now an arrow of type $Fa \to Gc$ may be obtained by an F-generating transformation first, followed by Ff and Gg with an F,G-commuting transformation in between, and a G-reducing transformation at the end:

$$Fa \xrightarrow{\mu} FFa \xrightarrow{Ff} FGb \xrightarrow{\gamma} GFb \xrightarrow{Gg} GGb \xrightarrow{\nu} Gb$$

Clearly, the typing axioms, except for uniqueness of typing, are satisfied:

$$f: a \to_{\mathcal{D}} b, \quad g: b \to_{\mathcal{D}} c \quad \Rightarrow \quad f:_{\mathcal{D}} g: a \to_{\mathcal{D}} c$$

$$id_{\mathcal{D},a}: a \to_{\mathcal{D}} a \quad .$$

So in order to prove that \mathcal{D} is a pre-category, it remains to show that $_{i\mathcal{D}}$ is associative and $id_{\mathcal{D}}$ is neutral for $_{i\mathcal{D}}$. We shall now give those proofs, assuming suitable properties on

 γ, η, μ, ν along the way. First, for associativity, let $f: a \to_{\mathcal{D}} b$, $g: b \to_{\mathcal{D}} c$, and $h: c \to_{\mathcal{D}} d$ be arbitrary; then:

```
f_{:\mathcal{D}}(g_{:\mathcal{D}}h) = (f_{:\mathcal{D}}g)_{:\mathcal{D}}h
                                                                                               definition \mathcal{D}
                                                                       commutes
                                                                       Fa \overset{F}{\rightarrow} FFa \overset{FF}{\rightarrow} FFFa \overset{FFf}{\rightarrow} FFGb \overset{F\gamma}{\rightarrow} FGFb \overset{FGg}{\rightarrow} FGGc \overset{F\nu}{\rightarrow} FGc \overset{\gamma}{\rightarrow} GFc \overset{Gh}{\rightarrow} GGd \overset{\nu}{\rightarrow} GGd \overset
                                                                                                   { in order to split the terms into several parts at isomorphic objects: }
(*)
                                                                                                 assume \gamma has inverse \check{\gamma}, so that \gamma F \colon FGF \to GFF has inverse \check{\gamma}F
                                                                       \gamma F \| \check{\gamma} F - \check{\gamma} G \| \gamma G
                                                                                                                                                                                                                                                                                                                                                                                                                                                 commutes
                                                                       Fa \overset{\mu}{\rightarrow} FFa \overset{F\mu}{\rightarrow} FFFa \overset{FFf}{\rightarrow} FFGb \overset{F\gamma}{\rightarrow} FGFb \overset{FGg}{\rightarrow} FGGc \overset{F\nu}{\rightarrow} FGc \overset{\gamma}{\rightarrow} GFc \overset{Gh}{\rightarrow} GGd \overset{\nu}{\rightarrow} Gd
                                                                                                diagram notation
                                                                       \mu, Ff, \gamma, G\mu = \mu, F\mu, FFf, F\gamma, \gamma F \land 
                                                                       GFq, \check{\gamma}G = \check{\gamma}F, FGq \wedge
                                                                       \gamma G, G\gamma, GGh, G\nu, \nu = F\nu, \gamma, Gh, \nu
                                                                                                   { for readability: }
                                                                                                 define \gamma_{2,1} = F\gamma, \gamma F: FFG \rightarrow GFF and
                                                                                                 define \gamma_{1,2} = \gamma G, G\gamma: FGG \rightarrow GGF
                                                                       \mu, Ff, \gamma, G\mu = \mu, F\mu, FFf, \gamma_{2.1} \wedge
                                                                       GFq, \check{\gamma}G = \check{\gamma}F, FGq \wedge
                                                                       \gamma_{1,2}, GGh, G\nu, \nu = F\nu, \gamma, Gh, \nu
                                                                                                 2nd conjunct: naturality \check{\gamma}: GF \to FG;
                                                                                                 1st, 3rd conjunct: { aiming at the next two steps, }
                                                                                                 assume \gamma, G\mu = \mu G, \gamma_{2,1} and F\nu, \gamma = \gamma_{1,2}, \nu F
                                                                       \mu, Ff, \mu G, \gamma_{2,1} = \mu, F\mu, FFf, \gamma_{2,1} \wedge
                                                                       \gamma_{1,2}, GGh, G\nu, \nu = \gamma_{1,2}, \nu F, Gh, \nu
                                                                                                naturality \mu: F \to FF and \nu: GG \to G
                                                                       \mu, \mu F, FFf, \gamma_{2,1} = \mu, F\mu, FFf, \gamma_{2,1} \wedge
                                                                       \gamma_{1,2} , GGh , G\nu , \nu=\gamma_{1,2} , GGh , \nu G , \nu
                                                                                               Leibniz
                                                                       \mu, \mu F = \mu, F\mu \wedge
                                                                       G\nu; \nu = \nu G; \nu
                                                                                                 assume \mu, \mu F = \mu, F\mu and G\nu, \nu = \nu G, \nu
                                                                       true .
```

The $\gamma_{2,1}$ and $\gamma_{1,2}$ defined above are instances of a more general $\gamma_{m,n}\colon F^mG^n\to G^nF^m$ that one may define easily by induction on m and n in several distinct but semantically equal ways. In step (*) there is no other nontrivial way to split the giant terms into several parts than the way indicated: in general two intermediate objects can be isomorphic only if their denotations contain the same ingredients. More precisely, in step (*) we have assumed that FGFb and GFFb are isomorphic via γ_{F} ; an alternative is to assume that FFGb and GFFb are isomorphic, via $\gamma_{2,1}$. But since $\gamma_{2,1} = F\gamma$, γ_{F} and since the term $F\gamma$ is present at just the right place, this alternative assumption is equivalent to ours.

Here are the five assumptions made along the way (the equalities are assumed, the typing is provable):

$$\gamma \text{ has inverse } \breve{\gamma}$$

$$\gamma \text{ ; } G\mu = \mu G \text{ ; } \gamma_{2,1} : FG \rightarrow GFF$$

$$F\nu \text{ ; } \gamma = \gamma_{1,2} \text{ ; } \nu F : FGG \rightarrow GF$$
and
$$\mu \text{ ; } \mu F = \mu \text{ ; } F\mu : F \rightarrow FF$$

$$G\nu \text{ ; } \nu = \nu G \text{ ; } \nu : GG \rightarrow G$$
where
$$\gamma_{2,1} = F\gamma \text{ ; } \gamma F : FFG \rightarrow GFF$$

$$\gamma_{1,2} = \gamma G \text{ ; } G\gamma : FGG \rightarrow GGF$$

Taking F = Id and also $\mu, \gamma = id, id$, these assumptions specialise to those that make the Kleisli composition for G, ν associative.

The derivation of *nice* assumptions on η in order that $id_{\mathcal{D}}$ is neutral for $_{:\mathcal{D}}$ is problematic. Let us first consider only the first equality in " $f_{:\mathcal{D}}$ $id = id_{:\mathcal{D}}$ f = f". So, let $f: a \to_{\mathcal{D}} b$ be arbitrary; then:

$$f_{:_{\mathcal{D}}} id_{\mathcal{D}} = id_{\mathcal{D}}_{:_{\mathcal{D}}} f$$

$$\equiv \operatorname{definition} \mathcal{D}$$

$$Fa \xrightarrow{\mu} FFa \xrightarrow{Ff} FGb \xrightarrow{\gamma} GFb \xrightarrow{G\eta} GGb \xrightarrow{\nu} Gb =$$

$$Fa \xrightarrow{\mu} FFa \xrightarrow{F\eta} FGa \xrightarrow{\gamma} GFa \xrightarrow{Gf} GGb \xrightarrow{\nu} Gb$$

$$\Leftarrow \operatorname{Leibniz} \left\{ \text{no nontrivial intermediate isomorphism seems plausible} \right\}$$

$$(*) Ff : \gamma : G\eta = F\eta : \gamma : Gf$$

$$\Leftarrow \operatorname{naturality} \eta : F \to G \text{, so } Ff : \eta G = \eta F : Gf$$

$$(\star) \gamma : G\eta = \eta G \quad \land \quad F\eta : \gamma = \eta F \quad .$$

Equation (*), for all f, is acceptably nice: it asserts a sort of naturality $F \to G$. Since we have already assumed natural transformation $\eta \colon F \to G$, it seems reasonable to require that $\gamma \colon G\eta$ is ηG , as in line (*). However, when instantiating with $F, \gamma := Id, id$, both line (*) and line (*) give requirements that are stronger than those for a monad: line (*) becomes $f \colon G\eta = \eta \colon Gf$ (which is not just naturality $\eta \colon Id \to G$), and line (*) becomes

 $G\eta = \eta G$. So, we look for another sufficient condition that implies f, $id_{\mathcal{D}} = id_{\mathcal{D}}$, f for all f: $a \to_{\mathcal{D}} b$. The following line of reasoning has been suggested by Lambert Meertens.

$$f_{:_{\mathcal{D}}} id_{\mathcal{D}} = id_{\mathcal{D}:_{\mathcal{D}}} f$$

$$\equiv \operatorname{definition} \mathcal{D}$$

$$Fa \xrightarrow{\mu} FFa \xrightarrow{Ff} FGb \xrightarrow{\gamma} GFb \xrightarrow{G\eta} GGb \xrightarrow{\nu} Gb = Fa \xrightarrow{\mu} FFa \xrightarrow{F\eta} FGa \xrightarrow{\gamma} GFa \xrightarrow{Gf} GGb \xrightarrow{\nu} Gb$$

$$\equiv \{ \text{in order to shift } Ff \text{ and } Gf \text{ towards each other, } \}$$

$$\mathbf{assume} \ \gamma : G\eta : \nu = \eta G : \nu \text{ and } \mu : F\eta : \gamma = \mu : \eta F$$

$$\mu : Ff : \eta G : \nu = \mu : \eta F : Gf : \nu$$

$$\Leftarrow \text{Leibniz}$$

$$Ff : \eta G = \eta F : Gf$$

$$\equiv \text{naturality } \eta : F \to G$$

$$true .$$

The two assumptions are acceptably nice.

As regards to the second equality in " $f_{;D}$ $id = id_{;D}$ f = f" we argue as follows. Let $f: a \to_D b$ be arbitrary, then:

$$f_{:_{\mathcal{D}}} id_{\mathcal{D}} = id_{\mathcal{D}}$$

$$\equiv \text{ definition } \mathcal{D}$$

$$\mu_{:} Ff_{:}, \gamma_{:} G\eta_{:}, \nu = f$$

$$\equiv \text{ assumption in previous calculation}$$

$$(*) \qquad \mu_{:} Ff_{:}, \eta G_{:}, \nu = f$$

$$\equiv \text{ naturality } \eta: F \to G$$

$$(\star) \qquad \mu_{:}, \eta F_{:}, Gf_{:}, \nu = f_{:}.$$

Thus we are led to assume the equation of line (*) (for all $f: a \to_{\mathcal{D}} b$), or equivalently, of line (\star) . However, both equations are too complicated to be called nice; in particular, the "f" occurs in the *middle* of the term and not at one end, as in naturality assertions. Fortunately, the instantiation with $F, \gamma, \mu := Id, id, id$ does give a monad law:

$$\mu : Ff : \eta G : \nu = f \quad \text{for all } f : a \to_{\mathcal{D}} b$$

$$\equiv \quad \text{substitution } F, \gamma, \mu := Id, id, id$$

$$f : \eta G : \nu = f \quad \text{for all } f : a \to Gb$$

$$\equiv \quad \text{for } \Leftarrow : \text{ Leibniz};$$

$$\text{for } \Rightarrow : \text{ take } a, f := Gb, id_{Gb}$$

$$\eta G : \nu = idG \quad .$$

Summary. Let F, G be functors. Then we call $(F, G, \gamma, \eta, \mu, \nu)$ a **dyad** if:

$$\begin{array}{cccc} \gamma & : & FG \to GF \\ \eta & : & F \to G \\ \mu & : & F \to FF \\ \nu & : & GG \to G \end{array},$$

satisfy the following conditions:

After substituting $F, \gamma, \mu := Id, id, id$, these requirements are equivalent to the statement that (G, η, ν) is a monad.

References

- [1] M. Barr and C. Wells. Category Theory for Computing Science. Prentice Hall, 1990.
- [2] P. Wadler. Comprehending monads. In ACM Conference on Lisp and Functional Programming, June 1990. To appear in Mathematical Structures in Computer Science.