Introduction to Modalities

August 26, 2022

1 Modalities

Motivation:

- The modal logic **S4** extends classical propositional logic by *modal operators* $\Diamond$ ('necessity') and $\Box$ ('possibility'), which come with axioms and rules that say that $\Diamond$ behaves like a category theoretic *monad*, and its De Morgan dual $\Box$ behaves like a *comonad*, i.e. the formulas

  $$\Box A \to A \qquad \Box A \to \Box\Box A \qquad A \to \Diamond A \qquad \Diamond\Diamond A \to \Diamond A$$

  are theorems of **S4** for all formulas A.

- Moggi's *monadic λ-calculus* [Mog91] extends the simply typed λ-calculus with a unary type constructor T and term formation rules

  $$\frac{\Gamma \vdash t : A}{\Gamma \vdash \eta(t) : TA} \qquad \frac{\Gamma \vdash s : TA \qquad \Gamma, x{:}A \vdash t : TB}{\Gamma \vdash \mathsf{let}\ x := s\ \mathsf{in}\ t : TB}$$

  saying that T behaves like a monad.

-

Type theoretic *modalities* can be viewed type theoretic version of the monadic possibility modality $\Diamond$, or as idempotent version of Moggi's monads[1]. In type theory, it is easiest to define modalities as specific *reflective subuniverses.* Following [RSS20, Uni13], we use the symbol $\bigcirc$ instead of $\Diamond$ for type theoretic modalities.

1.1 Subuniverses. Let U be a type theoretic universe. Given a predicate $P_\bigcirc : \mathsf{U} \to \mathsf{Prop}$ on U, we may form the *subtype*

$$\mathsf{U}_\bigcirc = \{A : \mathsf{U} \mid P_\bigcirc(A)\} \hookrightarrow \mathsf{U}$$

of U. We will refer to either of the predicate $P_\bigcirc$ and the subtype $\mathsf{U}_\bigcirc$ as a *subuniverse* of U. $\Diamond$

[1] Strictly speaking the modalities corresponding to possibility are the *monadic* modalities. There are also *comonadic* modalities in type theory corresponding to necessity in modal logic, but if we just say 'modality' we normally refer to the monadic type, and the first section is exclusively devoted to those. Comonadic modalities will come up in the second section.

1.2 Definition A subuniverse is called *reflective*, if for every type A in U there exists a type $\bigcirc A$ and a function $\eta_A : A \to \bigcirc A$ such that

$$\forall (B : \mathsf{U}_\bigcirc)(f : A \to B)\; \exists ! (\bar{f} : \bigcirc A \to B)\, .\, \bar{f} \circ \eta_A = f. \tag{1.1}$$

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "\eta_A"'] & B \\ \bigcirc A \arrow[ur, dashed, "\bar{f}"'] & \end{tikzcd}$$

$\diamondsuit$

1.3 Lemma *If $\mathsf{U}_\bigcirc \hookrightarrow \mathsf{U}$ is a reflective subuniverse and A is a type in U, then the type $\bigcirc A$ together with the map η_A is determined uniquely up to equivalence – and therefore by univalence up to identification – by property (1.1).*

1.4 Terminology We refer to $\bigcirc : \mathsf{U} \to \mathsf{U}$ as a *modal operator*, and to η as the *modal unit.*

A type A is called $\bigcirc$-connected if $\bigcirc A = 1$, and $\bigcirc$-modal if η_A is an equivalence. Thus, $\mathsf{U}_\bigcirc$ is the subuniverse of modal types. $\diamondsuit$

1.5 Lemma *Let $\mathsf{U}_\bigcirc \hookrightarrow \mathsf{U}$ be a reflective subuniverse.*

1. *Given types $A : U$ and $B : \mathsf{U}_\bigcirc$ and functions $f, g : \bigcirc A \to B$, we have $f = g$ whenever $f \circ \eta = g \circ \eta$ (and more precisely we have an equivalence between identity types $(f = g) \simeq (f \circ \eta = g \circ \eta)$).*

2. *A type B is modal iff $\forall (A : \mathsf{U})(f : A \to B)\; \exists ! (\bar{f} : \bigcirc A \to B)\, .\, \bar{f} \circ \eta = f$.*

1.6 Theorem *TFAE for a reflective subuniverse $\mathsf{U}_\bigcirc \hookrightarrow \mathsf{U}$.*

1. *$\mathsf{U}_\bigcirc$ is Σ-closed, i.e. if $A : \mathsf{U}_\bigcirc$ and $B : A \to \mathsf{U}_\bigcirc$ then $\sum_{(a:A)} B(a)$ is in $\mathsf{U}_\bigcirc$.*

2. *$\mathsf{U}_\bigcirc$ admits* unique dependent elimination, *i.e. given $A : \mathsf{U}$ and $B : \bigcirc A \to \mathsf{U}_\bigcirc$ and $f : \prod_{(a:A)} B(\eta(a))$ there exists a unique $\tilde{f} : \prod_{(a:\bigcirc A)} B(a)$ such that $\tilde{f} \circ \eta = f$.*

$$\begin{tikzcd} & \sum_{(a:\bigcirc A)} B(a) \arrow[d, bend left, "\mathsf{pr}_1"] \\ A \arrow[ur, "\langle \eta, f \rangle"] \arrow[r, "\eta"'] & \bigcirc A \arrow[u, dashed, bend left, "\langle 1, \tilde{f} \rangle"] \end{tikzcd}$$

Proof. First assume 1, let A, B, and f as in 2, and define $g := \langle \eta, f \rangle : A \to \sum_{(a:\bigcirc A)} B(a)$, i.e. $g(a) = (\eta(a), f(a))$. The type $\sum_{(a:\bigcirc A)} B(a)$ is in $\mathsf{U}_\bigcirc$ and therefore there exists a unique $\bar{g} : \bigcirc A \to \sum_{(a:\bigcirc A)} B(a)$ with $\bar{g} \circ \eta = g$. We can argue

$\mathsf{pr}_1 \circ \bar{g} \circ \eta = \mathsf{pr}_1 \circ g = \eta$ and therefore $\mathsf{pr}_1 \circ \bar{g} = 1$ by the universal property of η.

$$\begin{tikzcd}
 & \sum_{(a:\bigcirc A)} B(a) \arrow[d, bend left, "\mathsf{pr}_1"] \\
A \arrow[ur, "g=\langle \eta, f\rangle"] \arrow[r, "\eta"'] & \bigcirc A \arrow[u, dashed, bend left, "\bar{g}"]
\end{tikzcd}$$

This means that $\bar{g} = \langle 1, \tilde{f} \rangle$ for $\tilde{f} : \prod_{(a:\bigcirc A)} B(a)$. We have $\tilde{f} \circ \eta = \mathsf{pr}_2 \circ \bar{g} \circ \eta = \mathsf{pr}_2 \circ g = f$. Uniqueness of $\tilde{f}$ follows from uniqueness of $\bar{g}$.

Conversely, assume 2 and let $A : \mathsf{U}_\bigcirc$ and $B : A \to \mathsf{U}_\bigcirc$. We have to show that for every $C : \mathsf{U}_\bigcirc$ and $f : C \to \sum_{(a:A)} B(a)$ there exists a unique $\bar{f} : \bigcirc C \to \sum_{(a:A)} B(a)$ with $\bar{f} \circ \eta = f$. Writing $f = \langle g, h \rangle$ we get $\bar{f} = \langle \bar{g}, \tilde{h} \rangle$. ■

1.7 Definition A *modality* is a reflective subuniverse satisfying the equivalent conditions of the theorem. ♢

The 'official' definition is a bit different:

1.8 Definition [Uni13, Def. 7.7.5] A *modality* is an operation $\bigcirc : \mathsf{U} \to \mathsf{U}$ for which there are

1. functions $\eta_A : A \to \bigcirc(A)$ for every type $A : \mathsf{U}$,
2. for every $A : \mathsf{U}$ and every type family $B : \bigcirc(A) \to \mathsf{U}$, a function
$$\mathsf{ind}_\bigcirc : \left(\prod_{a:A} \bigcirc(B(\eta_A(a))) \right) \to \prod_{z:\bigcirc(A)} \bigcirc(B(z)),$$
and
3. a path $\mathsf{ind}_\bigcirc(f)(\eta_A(a)) = f(a)$ for each $f : \prod_{(a:A)} \bigcirc(B(\eta_A(a)))$, such that
4. for all $z, z' : \bigcirc(A)$, the map $\eta_{z=z'} : (z = z') \to \bigcirc(z = z')$ is an equivalence. ♢

Given a modality in this presentation, the reflective subuniverse is recovered by setting

$$P_\bigcirc(A) = \mathsf{isequiv}(\eta_A) \quad \text{and thus} \quad \mathsf{U}_\bigcirc = \{A : \mathsf{U} \mid \mathsf{isequiv}(\eta_A : A \to \bigcirc A)\}.$$

2 Examples of modalities

2.1 Truncation

Some of the most important examples of modalities are given by truncation. Recall from Emily's class the definition of the truncation levels.

$$\begin{aligned}
\mathsf{istrunc}_n &: \mathsf{U} \to \mathsf{Prop} && \text{for } n \geq -2 \\
\mathsf{istrunc}_{-2}(A) &:= \mathsf{isContr}(A) \\
\mathsf{istrunc}_{n+1}(A) &:= \forall (x\, y : A) \,.\, \mathsf{istrunc}_n(x = y)
\end{aligned}$$

2.1 Theorem *For all $n \geq -2$, the subuniverse*

$$n\text{-}\mathsf{Type} := \{A : \mathsf{U} \mid \mathsf{istrunc}_n(A)\}$$

is reflective and $\sum$-closed, i.e. a modality.

Proof. The closure under sums is easily shown by induction n. The case -2 is trivial. Assume that n-Type is closed under sums, and let $A \in_{-} \mathsf{Type} n+1$ and $B : A \to_{-} \mathsf{Type} n+1$, and $x, y : \sum_{()} \sum_{(a:A)} B(a)$. We have to show that the type $(x = y)$ is n-truncated. By Σ-induction we may assume that $x \equiv (a_1, b_1)$ and $y \equiv (a_2, b_2)$, and we have

$$(x = y) \simeq \sum_{p:(a_1 = a_2)} p_*(b_1) = b_2.$$

The right type is n-truncated by assumption and the claim follows since n-types are closed under sums.

To show that the subuniverses $n\text{-}\mathsf{Type} \hookrightarrow \mathsf{U}$ are reflective we have to construct *n-truncations* of types A in U, i.e. types $\|A\|_n$ together with functions $|-|_n : A \to \|A\|_n$ such that for all n-truncated types B and functions $f : A \to B$ there exists a unique $\bar{f} : \|A\|_n \to B$ with $\bar{f} \circ |-|_n = f$. There exist several constructions for n-truncation. The HoTT book gives a construction involving higher inductive types, specifically the "hub and spokes" construction [Uni13, Section 7.3]. A more elegant proof is given in Egbert Rijke's book: the case -2 is trivial. For $n = -1$ we speak of *propositional truncation*, which can either be constructed 'impredicatively' (using *propositional resizing*) as

$$\|A\|_{-1} = \forall (Q : \mathsf{Prop}) \,.\, (A \to Q) \to Q,$$

or 'predicatively' – but relying on pushouts in U – using Rijke's *join construction* [Rij17].

For the inductive step assume that U admits n-truncations for $n \geq -1$. To construct an $(n+1)$-truncation of $A : \mathsf{U}$ define

$$R_k : A \to n\text{-}\mathsf{Type}^A, \qquad R_k(a) = \lambda b \,.\, \|a = b\|_n.$$

An $(n+1)$-truncation of A is given by the *image* of R_n, which is the subtype

$$\|A\|_{n+1} := \mathsf{im}(R_n) := \{P : A \to n\text{-}\mathsf{Type} \mid \exists (a : A) \,.\, R_n(a) = P\} \hookrightarrow n\text{-}\mathsf{Type}. \quad \blacksquare$$

2.2 Open and closed modalities

For every $Q : \mathsf{Prop}$ we can define *open* and *closed* modalities

$$\mathsf{Op}_Q, \mathsf{Cl}_Q : \mathsf{U} \to \mathsf{U} \qquad \mathsf{Op}(A) = A^Q \qquad \mathsf{Cl}_Q(A) = A * Q$$

where in the $A * Q$ is the *join* of A and Q which is defined by the pushout

$$\begin{tikzcd} A \times Q \arrow[r, "\mathsf{pr}_2"'] \arrow[d, "\mathsf{pr}_1"] & Q \arrow[d] \\ A \arrow[r] \arrow[ur, phantom, "\ulcorner", very near end] & A * Q \end{tikzcd} .$$

These modalities are higher analogues of the open and closed *Lawvere Tierney topologies* known from 1-topos theory.

2.3 Localization and nullification

Given a function $f : Y \to X$, we call an object $A : \mathsf{U}$ f*-local* if the precomposition function

$$A^f : A^X \to A^Y \qquad A^f(g) = g \circ f$$

is an equivalence. Using higher inductive types, one can show that the subuniverse of f-local objects is reflective [RSS20, Theorem 2.16], and a *modality* whenever the codomain X of f is contractible [RSS20, Theorem 2.17]. In the latter case one also speaks of *nullification at* X. For example, n-truncation can be defined as nullification at S^{n+1}.

3 Factorization

3.1 Definition An *orthogonal factorization system* is a pair of predicates

$$\mathcal{L}, \mathcal{R} : \prod_{A,B:\mathsf{U}} (A \to B) \to \mathsf{Prop}$$

on functions, such that

1. both $\mathcal{L}$ and $\mathcal{R}$ are closed under composition and contain all identities, and
2. every function $f : A \to B$ factors *uniquely* as

   $$f = (A \xrightarrow{f_{\mathcal{L}}} E(f) \xrightarrow{f_{\mathcal{R}}} B)$$

   i.e. the space of factorizations is contractible.

The factorization system is called *stable*, if the left class is closed under pullbacks.◇

It turns out that stable factorizations systems are equivalent to modalities!

Specifically, given a modal reflective subuniverse $\mathsf{U}_{\bigcirc} \hookrightarrow \mathsf{U}$, we obtain a stable factorization system by setting

$$\mathcal{L} = \{\text{functions with modal fibers}\}$$
$$\mathcal{R} = \{\text{functions with connected fibers}\}.$$

A factorization of $f : B \to A$ is obtained by 'applying the modality fiberwise':

$$\begin{tikzcd} B \arrow[r, "\simeq"] \arrow[dr, "f"'] & \sum_{(a:A)} \mathsf{fib}_f(a) \arrow[r, "\sum_{(a:A)} \eta"] & \sum_{(a:A)} \bigcirc(\mathsf{fib}_f(a)) \arrow[dl, "\mathsf{pr}_1"] \\ & A & \end{tikzcd}$$

Conversely, we get a modality from a stable factorization system $(\mathcal{L}, \mathcal{R})$ by setting $\mathsf{U}_\bigcirc = \{A : \mathsf{U} \mid (A \to 1) \in \mathcal{R}\}$.

Formally étale maps

Surprisingly there is *another* construction producing OFSs from modalities. Here, the right maps are those $f : A \to B$ for which the square

$$\begin{tikzcd} A \arrow[r, "f"'] \arrow[d, "\eta"] & B \arrow[d, "\eta"] \\ \bigcirc A \arrow[r, "\bigcirc f :\equiv \overline{\eta \circ f}"] & \bigcirc B \end{tikzcd}$$

is a pullback. These maps are called $\bigcirc$-étale in [CR21].

The two OFS constructions coincide whenever the modality is *lex*, which intuitively means that the modality viewed as a functor on U preserves finite limits, and can be characterized in a variety of different ways [RSS20, Theorem 3.1].

References

[CR21] F. Cherubini and E. Rijke. Modal descent. *MSCS. Mathematical Structures in Computer Science*, 31(4):363–391, 2021.

[Mog91] E. Moggi. Notions of computation and monads. volume 93, pages 55–92. 1991. Selections from the 1989 IEEE Symposium on Logic in Computer Science.

[Rij17] E. Rijke. The join construction. January 2017.

[RSS20] E. Rijke, M. Shulman, and B. Spitters. Modalities in homotopy type theory. *Logical Methods in Computer Science*, 16(1):Paper No. 2, 79, 2020.

[Uni13] The Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics*. `https://homotopytypetheory.org/book`, Institute for Advanced Study, 2013.