Notes by Ioan Steffan Thomas

Modal Logic

Modal logic extends propositional logic by introducing new connectives called modalities. Modal logic is more expressive than propositional logic, but less expressive than first-order logic.

There are many different types of modal logics, depending on the choice of modalities provided and the interpretation of the modalities. The most common form of modal logic provides modalities to capture possibility and necessity.

Modality – a classification of logical propositions according to their asserting or denying the possibility, impossibility, contingency, or necessity of their content.

Syntax for Classic Modal Logic:

- Assuming there is an infinite alphabet, $A ::= P, Q, R, \ldots$, of propositional variables/letters.
- The standard grammar for all well-formed modal formulas ϕ is:

$$
\begin{array}{ll}
\phi ::= A & \\
\quad \mid \neg\phi & \text{(negation)} \\
\quad \mid (\phi \wedge \phi) & \text{(conjunction)} \\
\quad \mid (\phi \vee \phi) & \text{(disjunction)} \\
\quad \mid (\phi \rightarrow \phi) & \text{(implication)} \\
\quad \mid (\phi \leftrightarrow \phi) & \text{(equivalence)} \\
\quad \mid \Diamond\phi & \text{(possibility)} \\
\quad \mid \Box\phi & \text{(necessity)}
\end{array}
$$

- **Precedence (highest to lowest):** modalities and negations, conjunctions, disjunctions, implications, equivalences. This is a common precedence convention.

Logical Connectives:

- All logical connectives and laws found in propositional logic are valid here. I.e. all propositional logical equivalences are also logical equivalences in modal logic.

- **Possibility:**
  - It is possible that φ. This states that there is a possible world (see semantics) in which φ holds.
  - This can be written '$\Diamond\varphi$', this expression is true if φ is true for some world.

- **Necessity:**
  - It is a necessity that φ (or necessarily φ). This states that all accessible worlds (see semantics) are such that φ holds.
  - This can be written '$\Box\varphi$', this expression is true if φ is true for all accessible worlds.

- By chaining together n modalities, it can be expressed that accessible worlds can be reached in n 'steps' (traversals of edges) such that some proposition holds.

- **Modality Laws and Logical Equivalences:**
  - **Duality of Modalities:**

Easily shown by semantic equivalences and using De Morgan's for FOL. Clearly the box and diamond connectives are duals of each other.
$\neg\Diamond\varphi \equiv \Box\neg\varphi$
$\neg\Box\varphi \equiv \Diamond\neg\varphi$

- **Distributivity of Modalities:**
  Also easily shown by semantic equivalences and distributivities for FOL and equivalences for implications.
  $\Box(\varphi \wedge \psi) \equiv \Box\varphi \wedge \Box\psi$
  $\Diamond(\varphi \vee \psi) \equiv \Diamond\varphi \vee \Diamond\psi$

Relational Semantics:

- **Frames:**
  - Sometimes called a Kripke Frame.
  - A **frame** M is a pair (W, R), i.e. $M = (W, R)$.
  - W is a nonempty set (of '**worlds**').
  - R is a binary relation on W, i.e. $R \subseteq W \times W$, also known as the '**accessibility**' relation on worlds.
  - M can be visualised as a directed graph, with vertices contained in W and edges represented by R.
  - **Accessibility Relation:**
    - The relation R represents which worlds are accessible from each other.
    - For some $w \in W$ and $w' \in W$, Rww' is shorthand for $(w, w') \in R$, and represents that the world w' is accessible from the world w.
    - An alternative for Rww' is wRw'.
    - When visualised as a graph, Rww' represents a directed edge from w to w'.
- **Valuations:**
  - Given a fixed frame $M = (W, R)$, a valuation for M is a function ρ from the propositional letters to sets of worlds, i.e. $\rho: A \to \mathcal{P}(W)$, where $\mathcal{P}(S)$ represents the powerset of a set S.
  - The valuation ρ of some propositional letter P, $\rho(P)$, can be thought of as the set of worlds in which the propositional letter/proposition it true.
- Valuations of well-formed formulas under the valuation ρ for a frame $M = (W, R)$ for some world $w \in W$:
  $M, w \vDash_\rho P \equiv w \in \rho(P)$, i.e. P is true at world w (in frame M and valuation ρ)
  $M, w \nvDash_\rho P \equiv w \notin \rho(P)$
  $M, w \vDash_\rho \neg\varphi \equiv M, w \nvDash_\rho \varphi \equiv not\ M, w \vDash_\rho \neg\varphi$
  $M, w \vDash_\rho (\varphi \wedge \psi) \equiv \left(M, w \vDash_\rho \varphi\right)\ and\ \left(M, w \vDash_\rho \psi\right)$
  $M, w \vDash_\rho (\varphi \vee \psi) \equiv \left(M, w \vDash_\rho \varphi\right)\ or\ \left(M, w \vDash_\rho \psi\right)$
  $M, w \vDash_\rho (\varphi \Rightarrow \psi) \equiv \left(M, w \nvDash_\rho \varphi\right)\ or\ \left(M, w \vDash_\rho \psi\right)$
  $M, w \vDash_\rho (\varphi \Leftrightarrow \psi) \equiv$
  $\left(\left(M, w \vDash_\rho \varphi\right)\ and\ \left(M, w \vDash_\rho \psi\right)\right)\ or\ \left(\left(M, w \nvDash_\rho \varphi\right)\ and\ \left(M, w \nvDash_\rho \psi\right)\right)$
  $M, w \vDash_\rho \Diamond\varphi \equiv \exists w' \in W. Rww'\ and\ M, w' \vDash_\rho \varphi$, i.e. there is at least one possible world which is accessible from the world w which satisfies φ
  $M, w \vDash_\rho \Box\varphi \equiv \forall w' \in W. if\ Rww'\ then\ M, w' \vDash_\rho \varphi$, i.e. every world accessible from the world w satisfies φ.

- **Satisfaction** – a formula φ is satisfied in a frame $M = (W, R)$ for a world $w \in W$, by a valuation ρ for M if and only if $M, w \vDash_\rho \varphi$. If the formula is not satisfied by ρ with the world w, it can be denoted $M, w \nvDash_\rho \varphi$.
- **Satisfiability** ((in a specified frame $M = (W, R)$) – a formula φ is satisfiable in M if and only if there exists some valuation ρ for M and some world $w \in W$ such that $M, w \vDash_\rho \varphi$.
- **Validity** (in a specified frame $M = (W, R)$) – a formula φ is valid in M if and only is it $M, w \vDash_\rho \varphi$ for all valuations ρ for M and for all $w \in W$, which can be denoted $M \vDash \varphi$, and if unsatisfiable in M is denoted $M \nvDash \varphi$.
- **Satisfiability** (in general) – a formula φ is satisfiable if and only if it is satisfiable in some frame M.
- **Validity** (in all frames) – a formula φ is valid if and only if it is valid in all frames. This is denoted $\vDash \varphi$.
- Two modal formulas φ and ψ are **logically equivalent** (and hence interchangeable), $\varphi \equiv \psi$, if and only if for any frame $M = (W, R)$, any valuation ρ for M, and any world $w \in W$, $M, w \vDash_\rho \varphi \Leftrightarrow M, w \vDash_\rho \psi$.
- The same relationships and consequences of satisfiability and validity which occur for first-order logic applies here, obviously except relating to bound and free variables.

Example proving validity in modal logic:

- Show $(\Box A \to \Box(A \to B)) \to \Box B$ is valid.
- Proof: for any frame $M = (W, R)$, any valuation ρ for M and any $w \in W$:
    - RTP: $M, w \vDash_\rho (\Box A \wedge \Box(A \to B)) \to \Box B$.
    - Assume $M, w \vDash_\rho \Box A$ (1) and $M, w \vDash_\rho \Box(A \to B)$ (2).
    - RTP: $M, w \vDash_\rho \Box B$:
        - $M, w \vDash_\rho \Box B \equiv \forall w' \in W. Rww' \to M, w' \vDash_\rho B$
        - Assume: Rww' for a $w' \in W$.
        - RTP $M, w' \vDash_\rho B$:
            - From (1) and Rww', it follows that $M, w' \vDash_\rho A$. And from (2) and Rww', it follows that $M, w' \vDash_\rho A \to B$.

$$M, w' \vDash_\rho A \text{ and } M, w' \vDash_\rho A \to B$$
$$\equiv M, w' \vDash_\rho A \wedge (A \to B)$$
$$\text{by modus ponens } M, w' \vDash_\rho B \qquad Q.E.D$$

Different Types of Modal Logics

What has been described so far is the weakest form of modal logic, named K after Saul Kripke. Stronger logics can be defined, in which more formulas are valid, by requiring the frames satisfy various additional mathematical properties.

The typical system K is formed by the two axioms:

- **Necessitation Rule, N:** $(\vDash \varphi) \to (\vDash \Box\varphi)$, i.e. if φ is a tautology, then so is $\Box\varphi$.
- **Distribution Axiom, K:** $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$.

The other systems of 'normal' modal logic are built with additional axioms on top of those for K.

Reflexive Frames:

- A binary relation R on W is reflexive if and only if $\forall w \in W. Rww$.
- A frame (W, R) is reflexive if and only if R is reflexive.

- Example proof for reflexive frames: prove $M \vDash \Box\varphi \to \varphi$ (i.e. is valid) for any reflexive frame $M = (W, R)$:

  RTP for any $w \in W$ and any valuation ρ for M, $if\ M, w \vDash_\rho \Box\varphi\ then\ M, w \vDash_\rho \varphi$.
  Assume the antecedent $M, w \vDash_\rho \Box\varphi$, i.e. $\forall w' \in W.\ if\ Rww'\ then\ M, w' \vDash_\rho \varphi$.
  So, when $w' = w$, Rww is true as M is reflexive, and hence $M, w \vDash_\rho \varphi$ is true.
  Q.E.D

- Example proof of the converse: prove $M \vDash \Box\varphi \to \varphi$ only in reflexive frames:

  RTP for any non-reflexive frame $M = (W, R)$, for some $w \in W$ and some valuation ρ for M, $M, w \vDash_\rho \Box\varphi\ and\ M, w \nvDash_\rho \varphi$.
  Since M is not reflexive, $\exists w \in W.\ \neg Rww$.
  Using such a w, let $\varphi \equiv P$ for some letter P, and $\rho(P) = W - \{w\}$.
  Then $M, w \vDash_\rho \Box\varphi\ and\ M, w \nvDash_\rho \varphi$ becomes $\forall w' \in W\big(Rww' \to w' \in \rho(P)\big)\ and\ w \notin \rho(P)$, which is also $\forall w' \in W(Rww' \to w' \in W - \{w\})\ and\ w \notin W - \{w\}$.
  Clearly $w \notin W - \{w\}$ holds, and further $\forall w' \in W(Rww' \to w' \in W - \{w\})$ holds, as by assuming Rww', $w \neq w'$ as the frame is not reflexive, and since $w' \in W$, then $w' \in W - \{w\}$.
  So clearly the negation is satisfiable in non-reflexive frames, so the original is invalid in non-reflexive frames.
  Q.E.D
- Clearly then, $M \vDash \Box\varphi \to \varphi \equiv M\ is\ reflexive$. So, restricting all frames to be reflexive is the same as taking $\Box\varphi \to \varphi$ to be an **axiom** of modal logic defined with reflexive frames. This is the reflexivity axiom (**T**), if P is necessary, then it is that P.

Transitive Frames:

- A binary relation R on W is transitive if and only if $\forall x, y, z \in W.\ Rxy \wedge Ryz \to Rxz$.
- A frame (W, R) is transitive if and only if R is transitive.

- Example proof for transitive frames: prove $M \vDash \Box\varphi \to \Box\Box\varphi$ (i.e. is valid) for any transitive frame $M = (W, R)$:

  RTP for any $w \in W$ and any valuation ρ for M, $if\ M, w \vDash_\rho \Box\varphi\ then\ M, w \vDash_\rho \Box\Box\varphi$.
  Assume the antecedent $M, w \vDash_\rho \Box\varphi$ (1), i.e. $\forall w' \in W.\ if\ Rww'\ then\ M, w' \vDash_\rho \varphi$.
  Showing $M, w \vDash_\rho \Box\Box\varphi$:

  $\equiv \forall w' \in W.\ if\ Rww'\ then\ M, w' \vDash_\rho \Box\varphi$
  $\equiv \forall w' \in W.\ if\ Rww'\ then\ \forall w'' \in W\big(if\ Rw'w''\ then\ M, w'' \vDash_\rho \varphi\big)$

  To show this step, assume Rww' (2) and then assume $Rw'w''$ (3).
  RTP $M, w'' \vDash_\rho \varphi$.
  By transitivity of the frame, (2) and (3) implies Rww''.
  By (1), Rww'' implies $M, w'' \vDash_\rho \varphi$.
  Q.E.D

- Example proof of the converse: prove $M \vDash \Box\varphi \to \Box\Box\varphi$ only in transitive frames:

  RTP for any non-reflexive frame $M = (W, R)$, for some $w \in W$ and some valuation ρ for M, $M, w \vDash_\rho \Box\varphi\ and\ M, w \nvDash_\rho \Box\Box\varphi$.

Since M is not transitive, $\exists x, y, z \in W. Rxy \land Ryz \land \neg Rxz$.
Using such an x, y, z, let $\varphi \equiv P$ for some letter P, and $\rho(P) = W - \{z\}$, and take w as x.
Then $M, x \vDash_\rho \Box\varphi\ and\ M, x \nvDash_\rho \Box\Box\varphi$ becomes $\forall w' \in W(Rxw' \to w' \in \rho(P))\ and\ not\ \forall w', w'' \in W\left(Rxw' \to \left(Rw'w'' \to w'' \in \rho(P)\right)\right)$.
Equivalently $\forall w' \in W(Rxw' \to w' \in \rho(P))\ and\ \exists w', w'' \in W(Rxw' \land Rw'w'' \land w'' \notin \rho(P))$ by applying De Morgan's.
Clearly $\forall w' \in W(Rxw' \to w' \in \rho(P))$ holds as Rxw' is only true when $w' \neq z$, and since $w' \in W$, then $w' \in \rho(P)$. And further $\exists w', w'' \in W(Rxw' \land Rw'w'' \land w'' \notin \rho(P))$ holds, as Rxw' is true only when $w' \neq z$ and $Rw'w''$ is true for $w' = y$ and $w'' = z$, and since $z \notin \rho(P)$, this holds.
So clearly the negation is satisfiable in non-transitive frames, so the original is invalid in non-transitive frames.
Q.E.D

- Clearly then, $M \vDash \Box\varphi \to \Box\Box\varphi \equiv M\ is\ transitive$. So, restricting all frames to be transitive is the same as taking $\Box\varphi \to \Box\Box\varphi$ to be an **axiom** of modal logic defined with transitive frames. This is the transitivity axiom (**4**).

The $S4$ modal logic is formed by taking **T** and **4** as it's axioms, i.e. it considers only reflexive-transitive frames. This is one of the most popular modal logics.

Symmetric Frames:

- A binary relation R on W is symmetric if and only if $\forall x, y \in W. Rxy \to Ryx$ (so clearly the converse $Ryx \to Rxy$ is also true).
- A frame (W, R) is symmetric if and only if R is symmetric.

- Example proof for symmetric frames: prove $M \vDash \varphi \to \Box\Diamond\varphi$ (i.e. is valid) for any symmetric frame $M = (W, R)$:

  RTP for any $w \in W$ and any valuation ρ for $M,\ if\ M, w \vDash_\rho \varphi\ then\ M, w \vDash_\rho \Box\Diamond\varphi$.
  Assume the antecedent $M, w \vDash_\rho \varphi$ (1).
  Showing $M, w \vDash_\rho \Box\Diamond\varphi$:

  $\equiv \forall w' \in W. if\ Rww'\ then\ M, w' \vDash_\rho \Diamond\varphi$

  $\equiv \forall w' \in W. if\ Rww'\ then\ \exists w'' \in W(Rw'w''\ and\ M, w'' \vDash_\rho \varphi)$

  To show this step, assume Rww' (2).
  RTP $Rw'w''\ and\ M, w'' \vDash_\rho \varphi$.
  Take $w'' = w$, then by (1), $M, w'' \vDash_\rho \varphi$ is clearly true.
  By symmetricity of the frame, and by (2), $Rw'w$ is true.
  Hence, the case for $\exists w'' \in W. Rw'w''\ and\ M, w'' \vDash_\rho \varphi$ when $w'' = w$ holds whenever Rww' holds for any $w, w' \in W$.
  Q.E.D

- Example proof of the converse: prove $M \vDash \varphi \to \Box\Diamond\varphi$ only in symmetric frames:

  RTP for any non-symmetric frame $M = (W, R)$, for some $w \in W$ and some valuation ρ for M, $M, w \vDash_\rho \varphi\ and\ M, w \nvDash_\rho \Box\Diamond\varphi$.
  Since M is not symmetric, $\exists x, y \in W. Rxy \land \neg Ryx$.
  Using such an x, y, let $\varphi \equiv P$ for some letter P and $\rho(P) = \{x\}$, and take w as x.

Then $M, w \vDash_\rho \varphi \text{ and } M, w \nvDash_\rho \Box\Diamond\varphi$ becomes $x \in \rho(P) \text{ and not } \forall w' \in W\left(Rxw' \rightarrow \exists w'' \in W\big(Rw'w'' \wedge w'' \in \rho(P)\big)\right)$.

Equivalently $x \in \rho(P) \text{ and } \exists w' \in W \forall w'' \in W\left(Rxw' \wedge \big(Rw'w'' \rightarrow w'' \notin \rho(P)\big)\right)$.

Clearly $x \in \rho(P)$ as $\rho(P) = \{x\}$. Further, when $w' = y$, then clearly Rxw' holds. Then by assuming Ryw'', clearly w'' cannot be x, by the earlier choice, and whenever $w'' \neq x, w'' \notin \{x\}$, so $\forall w'' \in W. Rw'w'' \rightarrow w'' \notin \rho(P)$ holds.

Q.E.D

- Clearly then, $M \vDash \varphi \rightarrow \Box\Diamond\varphi \equiv M \text{ is symmetric}$. So, restricting all frames to be symmetric is the same as taking $\varphi \rightarrow \Box\Diamond\varphi$ to be an **axiom** of modal logic defined with symmetric frames. This is the symmetricity axiom (**B**).