Modal Logic Mechanisation in HOL4

Yiming Xu

15/6/2018

1 Getting Start

In our formalisation, we only consider the most standard modal language, called the basic modal language. The basic modal language is defined using a set Φ of propositional formulas, and only one modal operator $\Diamond$. A formula in the basic modal language is either a propositional symbol p where $p \in \Phi$, or a disjunction $\psi \vee \phi$ of prmodal formulas ψ and ϕ, or the negation $\neg\phi$ of a modal formula ϕ, or else of the form $\Diamond\phi$ obtained by putting a diamond before a modal formula. In the HOL, we create a datatype called 'form' of the formulas of this modal language.

$$\begin{aligned}
&\alpha\ \texttt{chap1\$form} = \\
&\quad \mathsf{VAR}\ \alpha \\
&\quad \mid \mathsf{DISJ}\ (\alpha\ \texttt{chap1\$form})\ (\alpha\ \texttt{chap1\$form}) \\
&\quad \mid \bot \\
&\quad \mid (\neg)\ (\alpha\ \texttt{chap1\$form}) \\
&\quad \mid \Diamond\ (\alpha\ \texttt{chap1\$form})
\end{aligned}$$

When we say an 'α form', we mean a formula in the basic modal language defined on a set Φ with elements of type α, such a set is called an α set, and of type $\alpha \to \mathit{bool}$. Note that we have the notion of the conjunction '$\wedge$', the implication $\to$, and the truth '$\top$', but they are not primitive, and defined as:

$$\vdash \mathsf{AND}\ f_1\ f_2 = \neg\mathsf{DISJ}\ (\neg f_1)\ (\neg f_2) \quad \vdash (f_1 \to f_2) = \mathsf{DISJ}\ (\neg f_1)\ f_2 \quad \vdash \mathsf{TRUE} = \neg\bot$$

As an analogue of the duality of the universal quantifier is the dual of the existential quantifier, in the sense that $\exists$ is defined to be $\neg\forall\neg$, we have a modal operator that is dual to the diamond, defined by $\Box\phi := \neg\Diamond\neg\phi$.

$$\vdash \Box\ \phi = \neg\Diamond\ (\neg\phi)$$

If a modal formula does not use any diamond, then it is also a propositional formula:

$$\begin{aligned}
\vdash\ &(\forall p.\ \mathsf{propform}\ (\mathsf{VAR}\ p) \iff \mathsf{T}) \wedge \\
&(\forall \phi_1\ \phi_2.\ \mathsf{propform}\ (\mathsf{DISJ}\ \phi_1\ \phi_2) \iff \mathsf{propform}\ \phi_1 \wedge \mathsf{propform}\ \phi_2) \wedge \\
&(\forall f.\ \mathsf{propform}\ (\neg f) \iff \mathsf{propform}\ f) \wedge (\mathsf{propform}\ \bot \iff \mathsf{T}) \wedge \\
&\forall f.\ \mathsf{propform}\ (\Diamond\ f) \iff \mathsf{F}
\end{aligned}$$

Knowing how does the modal formulas we are interested in look like, now we want to consider what does they mean. If the modal formula defined using propositional symbols in an α-set Φ is a propositional formula, then to talk about its truth value, what we need is just an assignment of truth value of the propositional letters in Φ, that is, a function $\Phi \to \mathit{bool}$. In the HOL, to define a function from an α-set, where α is any type, we are not allowed to only assign values to the elements of the set, instead, we must define the function on for all the terms that has type α. Hence we can define the function of evaluating propositional formulas as takes a function $\sigma : \alpha \to \mathit{bool}$ and an α propositional formula, and let it spit out the truth value of the propositional formula under the assignment σ.

$$\begin{aligned}
\vdash\ &(\forall \sigma\ p.\ \mathsf{peval}\ \sigma\ (\mathsf{VAR}\ p) \iff \sigma\ p) \wedge \\
&(\forall \sigma\ f_1\ f_2. \\
&\quad \mathsf{peval}\ \sigma\ (\mathsf{DISJ}\ f_1\ f_2) \iff \mathsf{peval}\ \sigma\ f_1 \vee \mathsf{peval}\ \sigma\ f_2) \wedge \\
&(\forall \sigma.\ \mathsf{peval}\ \sigma\ \bot \iff \mathsf{F}) \wedge \\
&(\forall \sigma\ f.\ \mathsf{peval}\ \sigma\ (\neg f) \iff \neg\mathsf{peval}\ \sigma\ f) \wedge \\
&\forall \sigma\ f.\ \mathsf{peval}\ \sigma\ (\Diamond\ f) \iff \mathsf{F}
\end{aligned}$$

However, to interpret a modal formula that involves diamonds, an assignment of truth value is not enough. We need a model of our language. A model of the basic modal language consists of two pieces of informations. The first thing we need is a frame. A frame consists of two things, a set, and a relation defined on elements in this set. If the underlying set of a frame is an β-set, then the frame is called an β-frame.

$$\alpha\ \texttt{frame} = \texttt{<|}\ \mathsf{world}\ :\ \alpha\ \rightarrow\ \texttt{bool};\ \mathsf{rel}\ :\ \alpha\ \rightarrow\ \alpha\ \rightarrow\ \texttt{bool}\ \texttt{|>}$$

The second thing we need is a valuation. If we are taking about interpret an α-form on a β-frame, a valuation is a function of type $\beta \to \alpha \to bool$. The information that a valuation gives is that, for each point in the frame, an assignment of truth value of each propositional symbol. We get a model by putting a frame and a valuation together.

$$\begin{array}{l}(\alpha,\ \beta)\ \texttt{chap1\$model} = \texttt{<|}\\ \quad \mathsf{frame}\ :\ \beta\ \texttt{frame};\\ \quad \mathsf{valt}\ :\ \alpha\ \rightarrow\ \beta\ \rightarrow\ \texttt{bool}\\ \texttt{|>}\end{array}$$

When we say an (α, β)-model, we mean a model for α-forms with a β-set as its underlying set.

Now we can interpret modal formulas in the basic modal language on a model by defining satisfication.

$$\begin{array}{l}\vdash (\forall\,\mathfrak{M}\ w\ p.\\ \qquad \mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{VAR}\ p) \iff\\ \qquad\quad w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ w\ \in\ \mathfrak{M}.\mathsf{valt}\ p)\ \wedge\\ \quad (\forall\,\mathfrak{M}\ w.\ \mathsf{satis}\ \mathfrak{M}\ w\ \bot\ \iff\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ \mathsf{F})\ \wedge\\ \quad (\forall\,\mathfrak{M}\ w\ \phi.\\ \qquad \mathsf{satis}\ \mathfrak{M}\ w\ (\neg\phi)\ \iff\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ \neg\mathsf{satis}\ \mathfrak{M}\ w\ \phi)\ \wedge\\ \quad (\forall\,\mathfrak{M}\ w\ \phi_1\ \phi_2.\\ \qquad \mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{DISJ}\ \phi_1\ \phi_2)\ \iff\ \mathsf{satis}\ \mathfrak{M}\ w\ \phi_1\ \vee\ \mathsf{satis}\ \mathfrak{M}\ w\ \phi_2)\ \wedge\\ \quad \forall\,\mathfrak{M}\ w\ \phi.\\ \qquad \mathsf{satis}\ \mathfrak{M}\ w\ (\Diamond\ \phi)\ \iff\\ \qquad\quad w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\\ \qquad\quad\ \exists\,v.\\ \qquad\qquad \mathfrak{M}.\mathsf{frame.rel}\ w\ v\ \wedge\ v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\\ \qquad\qquad \mathsf{satis}\ \mathfrak{M}\ v\ \phi\end{array}$$

(insert an example of model and satisfication here)

Just as in first order logic and propositional logic, we can define the notion of logical equivalence of modal formulas which is an equivalence relation between formulas which use propositional symbols of the same type:

$$\vdash \mathsf{equiv0}\ tyi\ f_1\ f_2\ \iff\ \forall\,\mathfrak{M}\ w.\ \mathsf{satis}\ \mathfrak{M}\ w\ f_1\ \iff\ \mathsf{satis}\ \mathfrak{M}\ w\ f_2\ \ \vdash \mathsf{equiv0}\ \mu\ \mathsf{equiv\_on}\ s$$

A notable thing is that $\mathsf{equiv0}$ need to take a type itself, denoted as μ, and if μ denotes the type α itself and f_1, f_2 are β formulas, $\mathsf{equiv0}\ \mu\ f_1\ f_2$ means for any (α, β)-model $\mathfrak{M}$ and any world w in it, we have $\mathsf{satis}\ \mathfrak{M}\ w\ f_1 \iff \mathsf{satis}\ \mathfrak{M}\ w\ f_2$. We are not allowed to omit the μ in the definition, since then we will introduce a type, namely the type of models $\mathfrak{M}$, that only appear on the right hand side but not on the left hand side of the definition, which is not allowed. Also we are not allowed to quantify over the μ, and something like 'equiv f1 f2 ⇔ !μ. equiv0 μ f1 f2' is also not allowed. This is actually a inconvenience of the HOL, since when we mention equivalence of formulas in usual mathematical language, we are implicitly taking about the class of all models, but the constrain here bans us from talking about all models at once.

We can immediately prove that for μ denotes any type, if $\mathsf{equiv0}\ \mu\ f\ g$ then $\mathsf{equiv0}\ \mu\ (\Diamond\ f)\ (\Diamond\ g)$, but the converse does not hold under our definition. In the usual mathematical notion of equivalence, if two diamond formulas $\Diamond\ f$ and $\Diamond\ g$ are equivalent, then f and g must be equivalent. It is because if f and g are not equivalent, then there eixsts a model $\mathfrak{M}$ and a world w such that w satisfies f but not g, then we can attach a world that is only linked to w, and the world we add is a witness of the fact that $\Diamond\ f$ and $\Diamond\ g$ are not equivalent. But under our definition in the HOL, if the μ denotes a finite type, then it is possible that we have $\mathsf{equiv0}\ \mu\ (\Diamond\ f)\ (\Diamond\ g)$ but $\neg\mathsf{equiv0}\ \mu\ f\ g$. For example, consider the type that has only two inhabitants a, b, then in the model below, we have $\mathsf{equiv0}\ \mu\ (\Diamond\ (\mathsf{VAR}\ p))\ (\Diamond\ (\mathsf{VAR}\ q))$ but $\neg\mathsf{equiv0}\ \mu\ (\mathsf{VAR}\ p)\ (\mathsf{VAR}\ q)$.

(add the picture of example, how to draw it?)

As the sitiution above is only because of our special definition, such case is uninteresting. For any model, regardless its world set is of a finite type or not, we can always create a copy of the model in an infinite type. So it is harmless to just consider equivalence of formulas for models whose underlying set is of an infinite type. When μ denotes an infinite type, as we can always come up with a fresh world that allows the proof of equivalence between f and g from the equivalence of $\Diamond f$ and $\Diamond g$ as in the usual sense to go through, we do have a double implication:

$$\vdash \mathsf{INFINITE}\,\mathcal{U}(:\beta) \Rightarrow \\ (\mathsf{equiv0}\ \mu\ (\Diamond\ f)\ (\Diamond\ g) \iff \mathsf{equiv0}\ \mu\ f\ g)$$

Given a modal formula, we can extract the set of propositional symbols appear in it, as:

$$\vdash (\forall p.\ \mathsf{prop\_letters}\ (\mathsf{VAR}\ p) = \{\ p\ \}) \land \mathsf{prop\_letters}\ \bot = \emptyset \land \\ (\forall f_1\ f_2. \\ \mathsf{prop\_letters}\ (\mathsf{DISJ}\ f_1\ f_2) = \\ \mathsf{prop\_letters}\ f_1 \cup \mathsf{prop\_letters}\ f_2) \land \\ (\forall f.\ \mathsf{prop\_letters}\ (\neg f) = \mathsf{prop\_letters}\ f) \land \\ \forall f.\ \mathsf{prop\_letters}\ (\Diamond\ f) = \mathsf{prop\_letters}\ f$$

Then we can show when evaluating a formula ϕ on a model, the only relevant information in the valuation is the assignment it makes to the propositional letters actually occurring in ϕ:

$$\vdash \mathfrak{M}_1.\mathsf{frame} = \mathfrak{M}_2.\mathsf{frame} \Rightarrow \\ (\forall p.\ p \in \mathsf{prop\_letters}\ phi \Rightarrow \mathfrak{M}_1.\mathsf{valt}\ p = \mathfrak{M}_2.\mathsf{valt}\ p) \Rightarrow \\ \forall w.\ w \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow (\mathsf{satis}\ \mathfrak{M}_1\ w\ phi \iff \mathsf{satis}\ \mathfrak{M}_2\ w\ phi)$$

In particular, if we are interpreting a propositional formula, then we do not need the relation on the model, as we can prove from definition:

$$\vdash \mathsf{propform}\ f \land w \in \mathfrak{M}.\mathsf{frame.world} \Rightarrow \\ (\mathsf{satis}\ \mathfrak{M}\ w\ f \iff \mathsf{peval}\ (\lambda\,a.\ w \in \mathfrak{M}.\mathsf{valt}\ a)\ f)$$

Moreover, we only need the truth values of the propositional symbols that appears in the formula:

$$\vdash \mathsf{propform}\ f \land (\forall a.\ \mathsf{VAR}\ a \in \mathsf{subforms}\ f \Rightarrow a \in s) \land \\ w \in \mathfrak{M}.\mathsf{frame.world} \Rightarrow \\ (\mathsf{satis}\ \mathfrak{M}\ w\ f \iff \mathsf{peval}\ ((\lambda\,a.\ w \in \mathfrak{M}.\mathsf{valt}\ a) \cap s)\ f)$$

One may confused why are we allowed to write $\lambda\ a.w \in M.valt\ a) \cap s$ here, for $\lambda\ a.(w \in M.valt\ a)$ is a function but s is a set. This is because both the α-set s and the function $\lambda\ a.(w \in M.valt)$ has type $\alpha \to bool$, so they can be feeded to the function $\cap$, which has type $(\alpha \to bool) \to (\alpha \to bool) \to (\alpha \to bool)$. Hence for the set s, we can either view it as a collection of elements of type α, or a function that assigns each term of type α a truth value, where the truth value assigned to $a : \alpha$ is T if and only if $a \in s$.

(change subforms to propsyms later?)

2 Invariant results and bisimulations

The key concept we are interested in this chapter is called 'modal equivalent', by 'two worlds $w \in M.frame.world, w' \in M'.frame.world$ are modal equivalent', we mean they satisfies exactly the same modal formulas.

$$\vdash \mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w' \iff \forall \phi.\ \mathsf{satis}\ \mathfrak{M}\ w\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ w'\ \phi$$

In this chapter, we investigate when can we get modal equivalence. In the first section, we talk about how can we get new models from old models without affect modal satisfication, and then we introduce morphisms between models, and find out under which condition we can map a point of a model to another model and preserve satisfication. In the second section, we investigate a kind of relation between models, called bisimulation, that gives arise to modal equivalence, in the sense of two worlds that are related by a bisimulation satisfies the same modal formulas.

2.1 Operations on models

Here we introduce two operations on models, called disjoint union and generated submodels respectively.

The operation on models that leads to invarience of formulas in the most straightforward way is the disjoint union:

```
DU $(f, dom)$ $\overset{\text{def}}{=}$
  <|frame :=
    <|world :=
       $\{\ w$ | FST $w\ \in\ dom\ \wedge$ SND $w\ \in\ (f$ (FST $w$)).frame.world $\}$ ;
     rel :=
       $(\lambda\, w_1\ w_2.$
          FST $w_1\ =$ FST $w_2\ \wedge$ FST $w_1\ \in\ dom\ \wedge$
          $(f$ (FST $w_1$)).frame.rel (SND $w_1$) (SND $w_2$))|>;
  valt := $(\lambda\, v\ w.\ (f$ (FST $w$)).valt $v$ (SND $w$))|>
```

Disjoint union is defined as a function that takes a function that takes a function $f : \beta \to (\alpha, \gamma)$-model and a β-set *dom*, which plays the role of the index set that we take the disjoint union over. The function f are defined for all terms of type β, but actually we only need its value on elements in *dom*. Given a family of models indexed by the set *dom*, the world set of the disjoint union of this family has elements of form (i, w) where $i \in dom$ and $w \in (f\ i).frame.world$. For a propositional letter p and a world (i, w), we have $DU\ (f, dom).valt\ p\ (i, w)$ iff $(f\ i).valt\ p\ w$, and $DU\ (f, dom).rel\ (i, w)\ (i', w')$ iff $i = i'$ and $(f\ i).frame.rel\ w\ w'$, means that (i, w) and (i', w') come from the same model and are related by the relation in that model.

The disjoint union does nothing except for collecting a set of models together, so it is unsurprising that we can prove by induction that:

```
⊢ FST $w\ \in\ dom\ \Rightarrow$
    (satis $(f$ (FST $w$)) (SND $w$) $phi\ \iff$
      satis (DU $(f, dom)$) $w\ phi$)
```

Disjoint union is the operation we use when we want to expand out scope from a smaller model to a large model. Accordingly, we have an operation that allows us to restrict our scope to a smaller model, this is called 'generated submodel' construction.

When we say 'M_1 is a submodel of M_2, we mean M_1 is a part of M_2.

```
SUBMODEL $\mathfrak{M}_1\ \mathfrak{M}_2$ $\overset{\text{def}}{=}$
  $\mathfrak{M}_1$.frame.world $\subseteq$ $\mathfrak{M}_2$.frame.world $\wedge$
  $\forall\, w_1.$
     $w_1\ \in\ \mathfrak{M}_1$.frame.world $\Rightarrow$
     $(\forall\, v.\ \mathfrak{M}_1$.valt $v\ w_1\ \iff\ \mathfrak{M}_2$.valt $v\ w_1)\ \wedge$
     $\forall\, w_2.$
        $w_2\ \in\ \mathfrak{M}_1$.frame.world $\Rightarrow$
        $(\mathfrak{M}_1$.frame.rel $w_1\ w_2\ \iff\ \mathfrak{M}_2$.frame.rel $w_1\ w_2)$
```

It is not necessary that submodel construction preserves modal satisfication, since we can discard relations between worlds and hence destroy the satisfication of formulas with diamonds.

But after adding some constrains to the relations that we need to preserve, we can have a special kind of submodels, called generated submodels, that preserves modal satisfication.

```
⊢ GENSUBMODEL $\mathfrak{M}_1\ \mathfrak{M}_2\ \iff$
    SUBMODEL $\mathfrak{M}_1\ \mathfrak{M}_2\ \wedge$
     $\forall\, w_1.$
        $w_1\ \in\ \mathfrak{M}_1$.frame.world $\Rightarrow$
        $\forall\, w_2.$
           $w_2\ \in\ \mathfrak{M}_2$.frame.world $\wedge\ \mathfrak{M}_2$.frame.rel $w_1\ w_2\ \Rightarrow$
           $w_2\ \in\ \mathfrak{M}_1$.frame.world
```

Note that for a model M_1 to be a generated submodel of M_2, we only require all the worlds w' which have a link from a world w in M_1 to be also included in the world set of M_1, and we are allowed to completely ignore everthing that is linked to a world in M_1. This is because modal formulas cannot 'look back', in the sense that linking or discard links to a world does not change the set of satisfied formulas at the world.

Also by induction, we can prove the invariance of modal satisfication under taking generated submodels:

```
⊢ GENSUBMODEL $\mathfrak{M}_1\ \mathfrak{M}_2\ \wedge\ n\ \in\ \mathfrak{M}_1$.frame.world $\Rightarrow$
   (satis $\mathfrak{M}_1\ n\ phi\ \iff$ satis $\mathfrak{M}_2\ n\ phi$)
```

2.2 Morphisms between models

In this section, we talk about various kind of 'morphisms' between models. Similar as in mathematics, the notion of 'morphism' here is used to describe maps that preserves structures. For instance, a homomorphism is a rather notion of 'structure-preserving':

$$
\begin{array}{l}
\vdash \mathsf{hom}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \iff \\
\quad \forall w. \\
\qquad w \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow \\
\qquad f\ w \in \mathfrak{M}_2.\mathsf{frame.world} \land \\
\qquad (\forall p.\ w \in \mathfrak{M}_1.\mathsf{valt}\ p \Rightarrow f\ w \in \mathfrak{M}_2.\mathsf{valt}\ p) \land \\
\qquad \forall u. \\
\qquad\quad u \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow \\
\qquad\quad \mathfrak{M}_1.\mathsf{frame.rel}\ w\ u \Rightarrow \\
\qquad\quad \mathfrak{M}_2.\mathsf{frame.rel}\ (f\ w)\ (f\ u)
\end{array}
$$

Although a homomorphism preserves link between worlds in the source model, we allow the existence of extra link in the target model which has no counterpart in the source. Because of this, we cannot guarentee any world in the source is mapped to a world that satisfies exactly the same set of modal formulas under a homomorphism. As an attempt to obtain an equivlence, we strengthen the condition of being a homomorphism, and what we get after strengthening is called a strong homomorphism:

$$
\begin{array}{l}
\mathsf{strong\_hom}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
\quad \forall w. \\
\qquad w \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow \\
\qquad f\ w \in \mathfrak{M}_2.\mathsf{frame.world} \land \\
\qquad (\forall p.\ w \in \mathfrak{M}_1.\mathsf{valt}\ p \iff f\ w \in \mathfrak{M}_2.\mathsf{valt}\ p) \land \\
\qquad \forall u. \\
\qquad\quad u \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow \\
\qquad\quad (\mathfrak{M}_1.\mathsf{frame.rel}\ w\ u \iff \mathfrak{M}_2.\mathsf{frame.rel}\ (f\ w)\ (f\ u))
\end{array}
$$

A strong homomorphism that is a bijection on the world set is called an isomorphism, and an isomorphism will certainly preserves everthing. Just for the sake of modal equivalence, we do not really require an isomorphism, instead, a surjective strong morphism is enough:

$$
\begin{array}{l}
\vdash \mathsf{strong\_hom}\ f\ \mathfrak{M}\ \mathfrak{M}' \land f\ w = w' \land w \in \mathfrak{M}.\mathsf{frame.world} \land \\
\quad \mathsf{SURJ}\ f\ \mathfrak{M}.\mathsf{frame.world}\ \mathfrak{M}'.\mathsf{frame.world} \Rightarrow \\
\quad \mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'
\end{array}
$$

Now we have a desired modal equivalence from strong homomorphism. The problem is that the condition of being a strong homomorphism is too strong. We want a suitably weakened condition on a morphism that gives arise of such equivalence, our answer is bounded morphism:

$$
\begin{array}{l}
\mathsf{bounded\_mor}\ f\ \mathfrak{M}\ \mathfrak{M}' \overset{\mathsf{def}}{=} \\
\quad \forall w. \\
\qquad w \in \mathfrak{M}.\mathsf{frame.world} \Rightarrow \\
\qquad f\ w \in \mathfrak{M}'.\mathsf{frame.world} \land \\
\qquad (\forall a.\ \mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{VAR}\ a) \iff \mathsf{satis}\ \mathfrak{M}'\ (f\ w)\ (\mathsf{VAR}\ a)) \land \\
\qquad (\forall v. \\
\qquad\quad v \in \mathfrak{M}.\mathsf{frame.world} \land \mathfrak{M}.\mathsf{frame.rel}\ w\ v \Rightarrow \\
\qquad\quad \mathfrak{M}'.\mathsf{frame.rel}\ (f\ w)\ (f\ v)) \land \\
\qquad \forall v'. \\
\qquad\quad v' \in \mathfrak{M}'.\mathsf{frame.world} \land \mathfrak{M}'.\mathsf{frame.rel}\ (f\ w)\ v' \Rightarrow \\
\qquad\quad \exists v.\ v \in \mathfrak{M}.\mathsf{frame.world} \land \mathfrak{M}.\mathsf{frame.rel}\ w\ v \land f\ v = v'
\end{array}
$$

The invariance result that bounded morphism gives is stated as:

$$
\begin{array}{l}
\vdash \mathsf{bounded\_mor}\ f\ \mathfrak{M}\ \mathfrak{M}' \land w \in \mathfrak{M}.\mathsf{frame.world} \Rightarrow \\
\quad (\mathsf{satis}\ \mathfrak{M}\ w\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ (f\ w)\ \phi)
\end{array}
$$

As an application, we will use it to prove the tree-like property of modal formulas. The tree-like property of modal formula says that for any formula ϕ satisfied on any point in any model, there exists a tree-like model such that ϕ is satisfied at the root of the tree. As the name indicates, a tree-like model is a model such that its underlying frame is a tree. Formally, a tree is defined as:

$$
\begin{array}{l}
\mathsf{tree}\ S\ r \stackrel{\mathsf{def}}{=} \\
\quad r \in S.\mathsf{world} \land \\
\quad (\forall t.\ t \in S.\mathsf{world} \Rightarrow (\mathsf{RESTRICT}\ S.\mathsf{rel}\ S.\mathsf{world})^{*}\ r\ t) \land \\
\quad (\forall r_0.\ r_0 \in S.\mathsf{world} \Rightarrow \neg S.\mathsf{rel}\ r_0\ r) \land \\
\quad \forall t.\ t \in S.\mathsf{world} \land t \neq r \Rightarrow \exists! t_0.\ t_0 \in S.\mathsf{world} \land S.\mathsf{rel}\ t_0\ t
\end{array}
$$

Here RTC denotes the reflexive and transitive closure, and RESTRICT is a function that takes a relation and a set and return a relation that can only possibly hold on the set we give:

$$\mathsf{RESTRICT}\ R\ s\ x\ y \stackrel{\mathsf{def}}{=} R\ x\ y \land x \in s \land y \in s$$

We read '*tree S r*' as 'S is a tree with root r'. By definition, any tree like model is rooted. An induction on reflextive and transitive closure proves a tree has no loop:

$$\vdash \mathsf{tree}\ s\ r \Rightarrow \forall t_0\ t.\ (\mathsf{RESTRICT}\ s.\mathsf{rel}\ s.\mathsf{world})^{+}\ t_0\ t \Rightarrow t_0 \neq t$$

We now prove the tree-like property of modal formulas:

$$
\begin{array}{l}
\vdash \mathsf{satis}\ \mathfrak{M}\ w\ \phi \Rightarrow \\
\quad \exists\ MODEL\ s.\ \mathsf{tree}\ MODEL.\mathsf{frame}\ s \land \mathsf{satis}\ MODEL\ s\ \phi
\end{array}
$$

Proof: Suppose satis $\mathfrak{M}$ w *phi* By the invariance result of rooted model, we have satis $\mathfrak{M}'$ w *phi* where $\mathfrak{M}$ is the rooted model generated by w. By the invariance result for bounded morphisms, it suffices to prove $\mathfrak{M}'$ is the image of some bounded morphism from some tree-like model where the root of the tree is mapped to the root of $\mathfrak{M}'$. We construct such a model M'' as follows: Take the set of worlds to be the finite sequences $[w = u_0; u_1; \cdots ; u_n]$ such that $n > 0$ and $M.frame.rel\ u_i\ u_{i+1}$ for all i. Define $M''.frame.rel\ [w; u_1; \cdots ; u_n]\ [w; v_1; \cdots ; v_m]$ iff $m = n+1$, $m_i = v_i$ for $i \leq n$ and $M.frame.rel\ u_n\ v_m$. The valuation is given by $[w; u_1; \cdots ; u_n] \in M''.valt\ p$ iff $u_n \in M.valt\ p$. In the HOL, such a model looks like:

```
bounded_preimage_rooted $\mathfrak{M}$ $x$ $\stackrel{\mathsf{def}}{=}$
 <|frame :=
  <|world :=
     { $l$ |
      HD $l$ = $x$ ∧ LENGTH $l$ > 0 ∧
      ∀ $m$.
        $m$ < LENGTH $l$ − 1 ⇒
        RESTRICT $\mathfrak{M}$.frame.rel $\mathfrak{M}$.frame.world (EL $m$ $l$)
          (EL ($m$ + 1) $l$) };
   rel :=
     (λ $l_1$ $l_2$.
        LENGTH $l_1$ + 1 = LENGTH $l_2$ ∧
        RESTRICT $\mathfrak{M}$.frame.rel $\mathfrak{M}$.frame.world (LAST $l_1$)
          (LAST $l_2$) ∧
        ∀ $m$. $m$ < LENGTH $l_1$ ⇒ EL $m$ $l_1$ = EL $m$ $l_2$)|>;
 valt := (λ $v$ $n$. $\mathfrak{M}$.valt $v$ (LAST $n$))|>
```

It is straightforward to check the map that sends a world in bounded_preimage_rooted $\mathfrak{M}'$ w to its last member is a bounded morphism, and $[w]$ in M'' is sent to w in $\mathfrak{M}'$, as desired.

2.3 Bisimulation

The three approachs to obtain modal equivalence has a common feature: all of them leads to a relation between models such that related states satisfies exactly the same set of propositional letters, and once we are able to make a transition in one model, we can make a corresponding transition in the other. This observation leads us to the concept of bisimulation:

$$\begin{array}{l}
\mathsf{bisim}\ Z\ \mathfrak{M}\ \mathfrak{M}' \stackrel{\mathsf{def}}{=} \\
\quad \forall\, w\ w'. \\
\qquad w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ w'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ Z\ w\ w'\ \Rightarrow \\
\qquad (\forall\, a.\ \mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{VAR}\ a)\ \iff\ \mathsf{satis}\ \mathfrak{M}'\ w'\ (\mathsf{VAR}\ a))\ \wedge \\
\qquad (\forall\, v. \\
\qquad\quad v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ \mathfrak{M}.\mathsf{frame.rel}\ w\ v\ \Rightarrow \\
\qquad\quad \exists\, v'.\ v'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ Z\ v\ v'\ \wedge\ \mathfrak{M}'.\mathsf{frame.rel}\ w'\ v')\ \wedge \\
\qquad \forall\, v'. \\
\qquad\quad v'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ \mathfrak{M}'.\mathsf{frame.rel}\ w'\ v'\ \Rightarrow \\
\qquad\quad \exists\, v.\ v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ Z\ v\ v'\ \wedge\ \mathfrak{M}.\mathsf{frame.rel}\ w\ v
\end{array}$$

If we have $\mathsf{bisim}\ Z\ \mathfrak{M}\ \mathfrak{M}'$, it means that Z is a bisimulation relation between worlds in $\mathfrak{M}$ and $\mathfrak{M}'$. We require any two worlds related by a bisimulation to have same atomic information and matching transition possibilities. We can see all of the three constructions we introduced before give arise to bisimulations:

$$\begin{array}{l}
\vdash i\ \in\ \mathit{dom}\ \wedge\ w\ \in\ (f\ i).\mathsf{frame.world}\ \Rightarrow \\
\quad \mathsf{bisim\_world}\ (f\ i)\ (\mathsf{DU}\ (f, \mathit{dom}))\ w\ (i, w)\ \ \vdash \mathsf{GENSUBMODEL}\ \mathfrak{M}\ \mathfrak{M}'\ \Rightarrow \\
\quad \forall\, w.\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow\ \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w\ \ \vdash \mathsf{bounded\_mor\_image}\ f\ \mathfrak{M}\ \mathfrak{M}'\ \Rightarrow \\
\quad \forall\, w.\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow\ \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ (f\ w)
\end{array}$$

Proof: For (i), the bisimulation relation is $\lambda\, a\ b.\ b\ =\ (i, a)$, which is linking a world to its corresponding copy in the disjoint union. For (ii), the relation is $\lambda\, n_1\ n_2.\ n_1\ =\ n_2$. And for (iii), the relation is $\lambda\, n_1\ n_2.\ n_2\ =\ f\ n_1$

The clauses on forth and back condition for a bisimulation relation provide precisely the information to push the induction on formula for proving the following invariance theorem about bisimulations through:

$$\vdash \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'\ \Rightarrow\ \mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'$$

The theorem above provides alternative proofs to the invariance theorems for disjoint union, generated submodels, and bounded morphisms. Hence we can say the three former constructions is actually 'the same thing'. But it is not the end of the story. A natural question to ask is : is bisimulation and modal equivalence the 'same thing'? More precisely, a bisimulation will always give a modal equivalence, conversely, is that the fact that a modal equivalence always give a bisimulation?

The answer is no. Nonetheless, we can prove the converse of the theorem above with an extra condition on the models. A model $\mathfrak{M}$ is called image finite if for any world $w\ \in\ \mathfrak{M}.\mathsf{frame.world}$, there are only finitely many worlds in $\mathfrak{M}$ that is related to w.

$$\begin{array}{l}
\mathsf{image\_finite}\ \mathfrak{M} \stackrel{\mathsf{def}}{=} \\
\quad \forall\, x. \\
\qquad x\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow \\
\qquad \mathsf{FINITE}\ \{\ y\ \mid\ y\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ \mathfrak{M}.\mathsf{frame.rel}\ x\ y\ \}
\end{array}$$

Our main theorem is called Hennessy-Milner theorem, it says that for image finite models, modal equivalence and bisimulation coincides:

$$\begin{array}{l}
\vdash \mathsf{image\_finite}\ \mathfrak{M}\ \wedge\ \mathsf{image\_finite}\ \mathfrak{M}'\ \wedge\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge \\
\quad w'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \Rightarrow \\
\quad (\mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'\ \iff\ \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w')
\end{array}$$

Proof: The implication from right to left is just the invariance theorem for bisimulation. We prove the implication from left to right. Given w and w' are worlds in $\mathfrak{M}$ and $\mathfrak{M}'$ which are modal equivalent, we prove the relation $\lambda\, n_1\ n_2.\ \forall\, \phi.\ \mathsf{satis}\ \mathfrak{M}\ n_1\ \phi\ \iff\ \mathsf{satis}\ \mathfrak{M}'\ n_2\ \phi$ gives a bisimulation. The first clause is immediate to check. For the second one, assume $\mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ n_1\ n_2$ and $\mathfrak{M}.\mathsf{frame.rel}\ n_1\ n_1'$ for some $n_1'\ \in\ \mathfrak{M}.\mathsf{frame.world}$, we prove the existence of the world $n_2'\ \in\ \mathfrak{M}'.\mathsf{frame.world}$ such that $\mathfrak{M}'.\mathsf{frame.rel}\ n_2\ n_2'$ and $\mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ n_1'\ n_2'$. Suppose such a n_2' does not exist, we derive a contradiction. Consider the set $S := u'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ \mathfrak{M}'.\mathsf{frame.rel}\ n_2\ u'$, the first claim is that this set is finite and nonempty. Finiteness comes from the fact that $\mathfrak{M}'$ is image finite, and if the set is empty, then $\Box\ \bot$ will be a formula which holds for n_2 but not for n_1, contradicts the modal equivalence between n_1 and n_2. By assumption, for each world in S, there is a formula phi such that $\mathsf{satis}\ \mathfrak{M}\ n_1'\ \mathit{phi}$ but $\neg\mathsf{satis}\ \mathfrak{M}'\ n_2'\ \mathit{phi}$. As the set S is finite, the set of such phis is finite. Then we can take the conjunction of such phis to obtain a formula psi. Then we will have $\mathsf{satis}\ \mathfrak{M}\ n_1\ (\Diamond\ \mathit{psi})$ but $\neg\mathsf{satis}\ \mathfrak{M}\ n_2\ (\Diamond\ \mathit{psi})$.

The trick is what to do to capture the big conjunction. Certainly we can define a big conjunction inductively as a function that takes a finite set and give us the formula that conjuncts them together, but here is a much more direct approach such that allows us to directly obtain the *psi*. We can prove:

$$\begin{aligned}
&\vdash \mathsf{FINITE}\ s \Rightarrow \\
&\quad s \neq \emptyset \Rightarrow \\
&\quad v \in \mathfrak{M}.\mathsf{frame.world} \land \\
&\quad (\forall v'.\ v' \in s \Rightarrow \exists\, phi.\ \mathsf{satis}\ \mathfrak{M}\ v\ phi \land \neg\mathsf{satis}\ \mathfrak{M}'\ v'\ phi) \Rightarrow \\
&\quad \exists\, psi.\ \mathsf{satis}\ \mathfrak{M}\ v\ psi \land \forall v'.\ v' \in s \Rightarrow \neg\mathsf{satis}\ \mathfrak{M}'\ v'\ psi
\end{aligned}$$

Using this lemma, we obtain the *psi* we want by plugging in S to be the s.

3 Finite model property

In this chapter, we prove the finite model property of our modal language, which says if a modal formula is satisfied on an arbitary mode, then it can be satisfied on a finite model, where finite model means the finitness of the set of worlds. We will discuss two methods for building finite models for satisfiable modal formulas, namely via filteration and selection.

3.1 Finite model property via filteration

One way to get a finite model is by taking the quotient of the world set. The quotient model we will get is called filtration of a model. The equivalence relation that we will use for taking the quotient is defined using the concept of 'closed under subformulas'. Firstly, subformulas is a function that takes a formula and gives a set of the subformulas of this formula, defined by:

$$\begin{aligned}
&\mathsf{subforms}\ (\mathsf{VAR}\ a) \overset{\mathsf{def}}{=} \{\ \mathsf{VAR}\ a\ \} \\
&\mathsf{subforms}\ \bot \overset{\mathsf{def}}{=} \{\ \bot\ \} \\
&\mathsf{subforms}\ (\neg f) \overset{\mathsf{def}}{=} \neg f\ \mathsf{INSERT}\ \mathsf{subforms}\ f \\
&\mathsf{subforms}\ (\mathsf{DISJ}\ f_1\ f_2) \overset{\mathsf{def}}{=} \\
&\quad \mathsf{DISJ}\ f_1\ f_2\ \mathsf{INSERT}\ \mathsf{subforms}\ f_1 \cup \mathsf{subforms}\ f_2 \\
&\mathsf{subforms}\ (\Diamond\ f) \overset{\mathsf{def}}{=} \Diamond\ f\ \mathsf{INSERT}\ \mathsf{subforms}\ f
\end{aligned}$$

The definition says a subformula of a formula is a part of a formula which is itself a formula. Some properties of subformulas can be proved immediately, for instance, we have any formula is a subformula of itself. Also subformulas are transitive, and the set of subformulas for any formula is finite.

$$\vdash phi \in \mathsf{subforms}\ phi \quad \vdash f \in \mathsf{subforms}\ phi \land phi \in \mathsf{subforms}\ psi \Rightarrow f \in \mathsf{subforms}\ psi \quad \vdash \mathsf{FINITE}\ (\mathsf{subforms}\ phi)$$

We say a set Σ is closed under formula if for any formula *phi* in the set, any subformula of *phi* is also in Σ.

$$\begin{aligned}
&\mathsf{CUS}\ \Sigma \overset{\mathsf{def}}{=} \\
&\quad \forall f\ f'. \\
&\qquad (\mathsf{DISJ}\ f\ f' \in \Sigma \Rightarrow f \in \Sigma \land f' \in \Sigma) \land \\
&\qquad (\neg f \in \Sigma \Rightarrow f \in \Sigma) \land (\Diamond\ f \in \Sigma \Rightarrow f \in \Sigma)
\end{aligned}$$

We can easily check the set of subformulas of any formulas is closed under subformulas:

$$\vdash \mathsf{CUS}\ (\mathsf{subforms}\ phi)$$

And for a model $\mathfrak{M}$, the equivalence relation we will use to filtrate its world set is:

$$\begin{aligned}
&\mathsf{REL\_CUS}\ \Sigma\ \mathfrak{M} \overset{\mathsf{def}}{=} \\
&\quad (\lambda\, w\ v. \\
&\qquad w \in \mathfrak{M}.\mathsf{frame.world} \land v \in \mathfrak{M}.\mathsf{frame.world} \land \\
&\qquad \forall\, phi.\ phi \in \Sigma \Rightarrow (\mathsf{satis}\ \mathfrak{M}\ w\ phi \iff \mathsf{satis}\ \mathfrak{M}\ v\ phi))
\end{aligned}$$

Under this equivalence relation, for any world $w \in \mathfrak{M}.\mathsf{frame.world}$, the equivalence class it belongs to looks like:

$$\mathsf{EC\_CUS}\ \Sigma\ \mathfrak{M}\ w \overset{\text{def}}{=} \{\ v \mid \mathsf{REL\_CUS}\ \Sigma\ \mathfrak{M}\ w\ v\ \}$$

We can take these equivalence class to be the set of worlds for the filtrated model, but then the model we get will have different type from the original model. To avoid changing the type, we define $\mathsf{EC\_REP}\ \Sigma\ \mathfrak{M}\ w$ to be the representative from the equivalence class where w lives in, and we will use $\mathsf{EC\_REP\_SET}\ \Sigma\ \mathfrak{M}$ to be the set of worlds for the filtrated model.

$$\mathsf{EC\_REP}\ \Sigma\ \mathfrak{M}\ w \overset{\text{def}}{=} \mathsf{CHOICE}\ (\mathsf{EC\_CUS}\ \Sigma\ \mathfrak{M}\ w)\mathsf{EC\_REP\_SET}\ \Sigma\ \mathfrak{M} \overset{\text{def}}{=}$$
$$\{\ n \mid \exists\, w.\ w \in \mathfrak{M}.\mathsf{frame.world} \land n = \mathsf{EC\_REP}\ \Sigma\ \mathfrak{M}\ w\ \}$$

The definition of filtration of a model $\mathfrak{M}$ via a subformula-closed set is given by a relation, we read as $\mathsf{filtration}\ \mathfrak{M}\ \Sigma\ FLT$ as 'FLT is a filteration of $\mathfrak{M}$ under Σ.

```
filtration $\mathfrak{M}\ \Sigma\ FLT \overset{\text{def}}{=}$
  CUS $\Sigma \land FLT$.frame.world $=$ EC_REP_SET $\Sigma\ \mathfrak{M} \land$
  $(\forall\, w\ v.$
     $w \in \mathfrak{M}$.frame.world $\land\ v \in \mathfrak{M}$.frame.world $\land\ \mathfrak{M}$.frame.rel $w\ v \Rightarrow$
     $FLT$.frame.rel (EC_REP $\Sigma\ \mathfrak{M}\ w$) (EC_REP $\Sigma\ \mathfrak{M}\ v$)) $\land$
  $(\forall\, w\ v.$
     $w \in \mathfrak{M}$.frame.world $\land\ v \in \mathfrak{M}$.frame.world $\land$
     $FLT$.frame.rel (EC_REP $\Sigma\ \mathfrak{M}\ w$) (EC_REP $\Sigma\ \mathfrak{M}\ v$) $\Rightarrow$
     $\forall\, phi\ psi.$
       $phi \in \Sigma \land phi = \Diamond\, psi \Rightarrow$
       satis $\mathfrak{M}\ v\ psi \Rightarrow$
       satis $\mathfrak{M}\ w\ phi) \land$
  $\forall\, p\ s.$
    $FLT$.valt $p\ s \iff$
     $\exists\, w.\ s =$ EC_REP $\Sigma\ \mathfrak{M}\ w \land$ satis $\mathfrak{M}\ w$ (VAR $p$)
```

The definition above forces filtration to only make sense for subformula closed sets. For any model $\mathfrak{M}$ and set Σ which is closed under subformulas, we can find some model such that it is a filtration of $\mathfrak{M}$ under Σ. One construction of filtration is given by:

```
FLT $\mathfrak{M}\ \Sigma \overset{\text{def}}{=}$
  <|frame :=
   <|world := EC_REP_SET $\Sigma\ \mathfrak{M}$;
    rel :=
      $(\lambda\, n_1\ n_2.$
         $\exists\, w_1\ w_2.$
            $w_1 \in \mathfrak{M}$.frame.world $\land\ w_2 \in \mathfrak{M}$.frame.world $\land$
            $n_1 =$ EC_REP $\Sigma\ \mathfrak{M}\ w_1 \land n_2 =$ EC_REP $\Sigma\ \mathfrak{M}\ w_2 \land$
            $\exists\, w'\ v'.$
               $w' \in \mathfrak{M}$.frame.world $\land\ v' \in \mathfrak{M}$.frame.world $\land$
               $w' \in$ EC_CUS $\Sigma\ \mathfrak{M}\ w_1 \land$
               $v' \in$ EC_CUS $\Sigma\ \mathfrak{M}\ w_2 \land\ \mathfrak{M}$.frame.rel $w'\ v')$|>;
  valt := $(\lambda\, p\ s.\ \exists\, w.\ s =$ EC_REP $\Sigma\ \mathfrak{M}\ w \land$ satis $\mathfrak{M}\ w$ (VAR $p$))|>
```

In fact, it is not the fact that we always have a unique filtration of a model under a subformula closed set, but we will just use the one defined above. It is routine to check the above definition does give a filtration:

$$\vdash \mathsf{CUS}\ \Sigma \Rightarrow \mathsf{filtration}\ \mathfrak{M}\ \Sigma\ (\mathsf{FLT}\ \mathfrak{M}\ \Sigma)$$

We are interested in two properties of filtration of models that directly give the proof of finite model property. Firstly, filtrating a model using a finite set gives a finite model:

$$\vdash \mathsf{FINITE}\ \Sigma \land \mathsf{filtration}\ \mathfrak{M}\ \Sigma\ FLT \Rightarrow$$
$$\mathsf{CARD}\ FLT.\mathsf{frame.world} \leq 2 ** \mathsf{CARD}\ \Sigma$$

Proof: As Σ is finite, it suffices to give an injection from the world set of FLT to $\mathsf{POW}\ \Sigma$. Such an injection is given by sending a world w to the set of formulas in Σ that is satisfies at w.

Secondly, we need the satisification of modal formula to be preserved under filtration, in the following sense:

$$\begin{array}{l}\vdash \mathit{phi} \in \Sigma \Rightarrow \\ \quad \forall w. \\ \qquad w \in \mathfrak{M}.\mathsf{frame.world} \wedge \mathsf{filtration}\ \mathfrak{M}\ \Sigma\ \mathit{FLT} \Rightarrow \\ \qquad (\mathsf{satis}\ \mathfrak{M}\ w\ \mathit{phi} \iff \mathsf{satis}\ \mathit{FLT}\ (\mathsf{EC\_REP}\ \Sigma\ \mathfrak{M}\ w)\ \mathit{phi})\end{array}$$

Putting the last two theorems together and we get the finite model property via filtration:

$$\begin{array}{l}\vdash \mathsf{satis}\ \mathfrak{M}\ w\ \mathit{phi} \Rightarrow \\ \quad \exists \mathfrak{M}'\ w'. \\ \qquad w' \in \mathfrak{M}'.\mathsf{frame.world} \wedge \mathsf{satis}\ \mathfrak{M}'\ w'\ \mathit{phi} \wedge \\ \qquad \mathsf{FINITE}\ \mathfrak{M}'.\mathsf{frame.world}\end{array}$$

3.2 Finite model property via filtration

Another method to build finite models for an arbitary model is by selection. The intuition of this method is that only finitely many diamonds can occur in a modal formula, so any formula can only capture the information of finitely depth. To make the notion of 'depth' precise, we define the degree of a modal formula, which count the diamonds appear in a model formula and hence measure how much of the model can a modal formula see from the current state:

$$\begin{array}{l}\mathsf{DEG}\ (\mathsf{VAR}\ p) \overset{\text{def}}{=} 0 \\ \mathsf{DEG}\ \bot \overset{\text{def}}{=} 0 \\ \mathsf{DEG}\ (\neg\phi) \overset{\text{def}}{=} \mathsf{DEG}\ \phi \\ \mathsf{DEG}\ (\mathsf{DISJ}\ \phi_1\ \phi_2) \overset{\text{def}}{=} \mathsf{MAX}\ (\mathsf{DEG}\ \phi_1)\ (\mathsf{DEG}\ \phi_2) \\ \mathsf{DEG}\ (\Diamond\ \phi) \overset{\text{def}}{=} \mathsf{DEG}\ \phi + 1\end{array}$$

A modal formula of degree zero is exactly a propositional formula:

$$\vdash \mathsf{DEG}\ f = 0 \iff \mathsf{propform}\ f$$

The crucial fact that we will use about the degree of formulas is the following lemma:

$$\begin{array}{l}\vdash \mathsf{FINITE}\ s \wedge \mathsf{INFINITE}\ \mathcal{U}(:\beta) \Rightarrow \\ \quad \forall n. \\ \qquad \mathsf{FINITE} \\ \qquad (\{\ f \mid \mathsf{DEG}\ f \le n \wedge \forall a.\ \mathsf{VAR}\ a \in \mathsf{subforms}\ f \Rightarrow a \in s\ \}//E \\ \qquad \mu)\end{array}$$

The proof of this lemma is by induction on the degree n. We prove it in the following interlude:

4 Interlude I: Finiteness of non-equivalent modal formulas in each degree

4.1 Base case

For the base case, we need to prove if we only use propositional letters in s where s is a finite set, then up to equivalence, we can only onbtain finitely many propositional formulas. To prove this, it suffices to find out an injection from the set of equivalence class of propositional formulas that only uses propositional letters in s to a finite set. We will prove the function that sends an equivalence class of formulas to its image under the function $\lambda f.\{\sigma \mid \mathit{peval}\ \sigma\ f\} \cap (\mathit{POW}\ s)$, which sends a formula f to the set $\{\sigma \mid \mathit{peval}\ \sigma\ f\} \cap (\mathit{POW}\ s)$ of assignments of truth value on propositional letters in s that makes f holds, is an injection to the finite set $\mathsf{POW}\ (\mathsf{POW}\ (\mathsf{POW}\ s))$. We can see the function we define has correct codomain: A formula f is sent to a set of subsets of s, which is an element of $\mathsf{POW}\ (\mathsf{POW}\ s)$, so a equivalence class has image as a set of elements in $\mathsf{POW}\ (\mathsf{POW}\ s)$, which lies in $\mathsf{POW}\ (\mathsf{POW}\ (\mathsf{POW}\ s))$.

Let us firstly investigate what the function $\lambda f.\{\sigma \mid \mathit{peval}\ \sigma\ f\} \cap (\mathit{POW}\ s)$ does. We claim the image of each equivalence class of this function is a singleton:

$$\begin{array}{l} \vdash x \in \\ \quad \{\, f \mid \mathsf{propform}\ f \ \wedge\ \forall\, a.\ \mathsf{VAR}\ a \ \in\ \mathsf{subforms}\ f \ \Rightarrow\ a \ \in\ s \,\} // E\ \mu \ \wedge \\ \quad \phi \ \in\ x \ \Rightarrow \\ \quad \mathsf{IMAGE}\ (\lambda f.\ \{\, \sigma \mid \mathsf{peval}\ \sigma\ f \,\} \ \cap\ \mathsf{POW}\ s)\ x \ = \\ \quad\quad \{\ \{\, \sigma \mid \mathsf{peval}\ \sigma\ \phi \,\} \ \cap\ \mathsf{POW}\ s \,\} \end{array}$$

Indeed, if two $f1$ and $f2$ formulas are equivalent, then for any assignment of truth value of propositional symbols, it makes $f1$ true iff it makes $f2$ true:

$$\begin{array}{l} \vdash \mathsf{propform}\ f_1 \ \wedge\ \mathsf{propform}\ f_2 \ \wedge \\ \quad (\forall\, \mathfrak{M}\ w.\ \mathsf{satis}\ \mathfrak{M}\ w\ f_1 \iff \mathsf{satis}\ \mathfrak{M}\ w\ f_2) \ \Rightarrow \\ \quad \forall\, \sigma.\ \mathsf{peval}\ \sigma\ f_1 \iff \mathsf{peval}\ \sigma\ f_2 \end{array}$$

It is because if we have a $\mathsf{peval}\ \sigma\ f_1 \ \wedge\ \neg\mathsf{peval}\ \sigma\ f_2$, then we can construct a model with only one world w, no relation, and define the valuation at w for propositional symbols to be σ. Then by `peval_satis`, we get $\mathsf{satis}\ \mathfrak{M}\ w\ f_1 \ \wedge\ \neg\mathsf{satis}\ \mathfrak{M}\ w\ f_2$, a contradiction. As a consequence, we have:

$$\begin{array}{l} \vdash \mathsf{propform}\ f_1 \ \wedge\ \mathsf{propform}\ f_2 \ \wedge\ \mathsf{equiv0}\ \mu\ f_1\ f_2 \ \Rightarrow \\ \quad (\lambda f\ s.\ \mathsf{peval}\ s\ f)\ f_1 \ =\ (\lambda f\ s.\ \mathsf{peval}\ s\ f)\ f_2 \end{array}$$

which says if two formulas are equivalent, then the sets of valuations that makes them hold are identical. As a consequence, these two sets will still be equal after taking inter section with $POW\ s$. This proves `IMAGE_peval_singlton_strengthen`.

Recall we proved the condition for a propositional formula to hold as the lemma `peval_satis_strengthen`, it gives arise to this useful lemma:

$$\begin{array}{l} \vdash \mathsf{propform}\ f_1 \ \wedge\ \mathsf{propform}\ f_2 \ \wedge \\ \quad (\forall\, a.\ \mathsf{VAR}\ a \ \in\ \mathsf{subforms}\ f_1 \ \Rightarrow\ a \ \in\ s) \ \wedge \\ \quad (\forall\, a.\ \mathsf{VAR}\ a \ \in\ \mathsf{subforms}\ f_2 \ \Rightarrow\ a \ \in\ s) \ \Rightarrow \\ \quad (\forall\, \sigma.\ \sigma \ \in\ \mathsf{POW}\ s \ \Rightarrow\ (\mathsf{peval}\ \sigma\ f_1 \iff \mathsf{peval}\ \sigma\ f_2)) \ \Rightarrow \\ \quad \forall\, \mathfrak{M}\ w.\ \mathsf{satis}\ \mathfrak{M}\ w\ f_1 \iff \mathsf{satis}\ \mathfrak{M}\ w\ f_2 \end{array}$$

Proof: Suppose $\mathsf{satis}\ \mathfrak{M}\ w\ f_1$ for some model $\mathfrak{M}$, then by `peval_satis_strengthen`, we have $\mathsf{peval}\ ((\lambda\, a.\ w \ \in\ \mathfrak{M}.\mathsf{valt}\ a) \ \cap\ s)\ f_1$. As $(\lambda\, a.\ w \ \in\ \mathfrak{M}.\mathsf{valt}\ a) \ \cap\ s$ gives a subset of s, we conclude $\mathsf{peval}\ ((\lambda\, a.\ w \ \in\ \mathfrak{M}.\mathsf{valt}\ a) \ \cap\ s)\ f_2$ by assumption, and then $\mathsf{satis}\ \mathfrak{M}\ w\ f_2$ by `peval_satis_strengthen` again. The other direction is the same.

We can now prove the injection:

$$\begin{array}{l} \vdash \mathsf{INJ}\ (\lambda\, eqc.\ \mathsf{IMAGE}\ (\lambda f.\ \{\, s \mid \mathsf{peval}\ s\ f \,\} \ \cap\ \mathsf{POW}\ s)\ eqc) \\ \quad (\{\, f \mid \mathsf{propform}\ f \ \wedge\ \forall\, a.\ \mathsf{VAR}\ a \ \in\ \mathsf{subforms}\ f \ \Rightarrow\ a \ \in\ s \,\} // E\ \mu) \\ \quad (\mathsf{POW}\ (\mathsf{POW}\ (\mathsf{POW}\ s))) \end{array}$$

Proof: If two equivalence classes represented by x and x' respectively has identical image under the function $\lambda f.\{s \mid peval\ sf\} \cap POW\ s$, then by `IMAGE_peval_singlton_strengthen`, it gives $\{\{\sigma \mid peval\ \sigma x\} \cap POW\ s\}$. This means $\forall\, \sigma.\ \sigma \ \in\ \mathsf{POW}\ s \ \Rightarrow\ \mathsf{peval}\ \sigma\ x \iff \mathsf{peval}\ \sigma\ x'$, which means x and x' are equivalent by `equiv0_peval_strengthen`.

4.2 Step case

The key observation for proving the step case is that any formulas which only uses propositional letters in a fixed finite set ss and of degree no more than $n + 1$ is obtained by boolean combination on the set s formed by formulas of form $\Diamond\ phi$ which also only uses propositional letters in ss where $\mathsf{DEG}\ phi \ \leq\ n$ and propositional letters in ss. Saying a formula f is a boolean combination of a set s of formulas is to say f is either itself an element in s, or can be obtained by combining elements in s using the connectives $\neg$ and $\vee$. We define boolean combination as an inductive relation, where $\mathsf{IBC}\ f\ s$ reads 'f is a boolean combination of elements in the set s':

$$\begin{array}{l}
\vdash (\forall f_1\, f_2\, s.\, \mathsf{IBC}\, f_1\ s\ \wedge\ \mathsf{IBC}\, f_2\ s\ \Rightarrow\ \mathsf{IBC}\ (\mathsf{DISJ}\, f_1\, f_2)\ s)\ \wedge \\
\quad (\forall s.\, \mathsf{IBC}\ \bot\ s)\ \wedge\ (\forall f\ s.\, \mathsf{IBC}\ f\ s\ \Rightarrow\ \mathsf{IBC}\ (\neg f)\ s)\ \wedge \\
\quad \forall f\ s.\, f\ \in\ s\ \Rightarrow\ \mathsf{IBC}\, f\ s\ \vdash (\forall f_1\, f_2\, s.\, \mathit{IBC}'\, f_1\ s\ \wedge\ \mathit{IBC}'\, f_2\ s\ \Rightarrow\ \mathit{IBC}'\ (\mathsf{DISJ}\, f_1\, f_2)\ s)\ \wedge \\
\quad (\forall s.\, \mathit{IBC}'\ \bot\ s)\ \wedge\ (\forall f\ s.\, \mathit{IBC}'\ f\ s\ \Rightarrow\ \mathit{IBC}'\ (\neg f)\ s)\ \wedge \\
\quad (\forall f\ s.\, f\ \in\ s\ \Rightarrow\ \mathit{IBC}'\ f\ s)\ \Rightarrow \\
\quad \forall a_0\ a_1.\, \mathsf{IBC}\ a_0\ a_1\ \Rightarrow\ \mathit{IBC}'\ a_0\ a_1\ \vdash \mathsf{IBC}\ a_0\ a_1\ \iff \\
\quad\ (\exists f_1\, f_2.\ a_0\ =\ \mathsf{DISJ}\, f_1\, f_2\ \wedge\ \mathsf{IBC}\, f_1\ a_1\ \wedge\ \mathsf{IBC}\, f_2\ a_1)\ \vee \\
\quad\quad a_0\ =\ \bot\ \vee\ (\exists f.\ a_0\ =\ \neg f\ \wedge\ \mathsf{IBC}\, f\ a_1)\ \vee\ a_0\ \in\ a_1
\end{array}$$

We can prove our claim in the last paragraph with induction:

$$\begin{array}{l}
\vdash \mathsf{DEG}\, x\ \le\ n\ +\ 1\ \wedge\ (\forall a.\, \mathsf{VAR}\ a\ \in\ \mathsf{subforms}\, x\ \Rightarrow\ a\ \in\ s)\ \iff \\
\quad \mathsf{IBC}\ x \\
\quad\quad (\{\ \mathsf{VAR}\ v\ \mid\ v\ \in\ s\ \}\ \cup \\
\quad\quad\ \{\ \Diamond\ \mathit{psi}\ \mid \\
\quad\quad\quad \mathsf{DEG}\ \mathit{psi}\ \le\ n\ \wedge\ \forall a.\, \mathsf{VAR}\ a\ \in\ \mathsf{subforms}\ \mathit{psi}\ \Rightarrow\ a\ \in\ s\ \})
\end{array}$$

Observe a formula which is a boolean combinations on an empty set is either equivalent to $\bot$ or $\top$:

$$\vdash \mathsf{IBC}\, f\ s\ \Rightarrow\ s\ =\ \emptyset\ \Rightarrow\ \mathsf{equiv0}\ \mu\, f\ \mathsf{TRUE}\ \vee\ \mathsf{equiv0}\ \mu\, f\ \bot$$

so we are interested in boolean combination of non-empty sets along our proof, and leave the empty case to be treat separately at the very end. Our aim is to prove the set of boolean combinations on a set which contains only finitely many non-equivalent formulas only contain finitely many non-equivalent formulas:

$$\vdash \mathit{fs}\ \neq\ \emptyset\ \Rightarrow\ \mathsf{FINITE}\ (\mathit{fs}/\!/E\ \mu)\ \Rightarrow\ \mathsf{FINITE}\ (\{\ f\ \mid\ \mathsf{IBC}\, f\ \mathit{fs}\ \}/\!/E\ \mu)$$

Note that it suffices to prove the set of formulas obtained by boolean combination on a finite set is finite up to equivalence. Since then as fs have only finitely many non-equivalent formulas, the set $\mathsf{IMAGE\ CHOICE}\ (\mathit{fs}/\!/E\ \mu)$ of representatives of the equivalence classes is finite. There is a surjection $\lambda s.\{y \mid IBC\ y\ fs \wedge !f.f \in s \implies equiv0\ \mu\ y\ f\}$ from $\{f \mid IBC\ f\ (IMAGE\ CHOICE\ (fs//E\ \mu))\}//E\ \mu$ to $\{f \mid IBC\ f\ fs\}//E\ \mu$ is a surjection, showing the codomain is finite.

The strategy we used to prove this is to prove that any formula obtained by a boolean combination on a set is equivalent to a formula in disjunction normal form on the same set. The disjunction normal form is defined by:

$$\mathsf{DNF\_OF}\, f\ \mathit{fs}\ \overset{\mathsf{def}}{=}\ \mathsf{DISJ\_OF}\, f\ \{\ c\ \mid\ \mathsf{CONJ\_OF}\ c\ \mathit{fs}\ \}$$

where CONJ_OF is defined by:

$$\begin{array}{l}
\vdash (\forall c_0\ c.\ c\ =\ c_0\ \vee\ c\ =\ \neg c_0\ \Rightarrow\ \mathsf{CONJ\_OF}\ c\ \{\ c_0\ \})\ \wedge \\
\quad \forall f_0\, f_1\, f_2\, \mathit{fs}. \\
\quad\quad (f_1\ =\ f_0\ \vee\ f_1\ =\ \neg f_0)\ \wedge\ f_0\ \in\ \mathit{fs}\ \wedge \\
\quad\quad \mathsf{CONJ\_OF}\, f_2\ (\mathit{fs}\ \mathsf{DELETE}\ f_0)\ \Rightarrow \\
\quad\quad \mathsf{CONJ\_OF}\ (\mathsf{AND}\, f_1\, f_2)\ \mathit{fs}\ \vdash (\forall c_0\ c.\ c\ =\ c_0\ \vee\ c\ =\ \neg c_0\ \Rightarrow\ \mathit{CONJ\_OF}'\ c\ \{\ c_0\ \})\ \wedge \\
\quad (\forall f_0\, f_1\, f_2\, \mathit{fs}. \\
\quad\quad (f_1\ =\ f_0\ \vee\ f_1\ =\ \neg f_0)\ \wedge\ f_0\ \in\ \mathit{fs}\ \wedge \\
\quad\quad \mathit{CONJ\_OF}'\, f_2\ (\mathit{fs}\ \mathsf{DELETE}\ f_0)\ \Rightarrow \\
\quad\quad \mathit{CONJ\_OF}'\ (\mathsf{AND}\, f_1\, f_2)\ \mathit{fs})\ \Rightarrow \\
\quad \forall a_0\ a_1.\, \mathsf{CONJ\_OF}\ a_0\ a_1\ \Rightarrow\ \mathit{CONJ\_OF}'\ a_0\ a_1\ \vdash \mathsf{CONJ\_OF}\ a_0\ a_1\ \iff \\
\quad\ (\exists c_0.\ a_1\ =\ \{\ c_0\ \}\ \wedge\ (a_0\ =\ c_0\ \vee\ a_0\ =\ \neg c_0))\ \vee \\
\quad\quad \exists f_0\, f_1\, f_2. \\
\quad\quad\quad a_0\ =\ \mathsf{AND}\, f_1\, f_2\ \wedge\ (f_1\ =\ f_0\ \vee\ f_1\ =\ \neg f_0)\ \wedge\ f_0\ \in\ a_1\ \wedge \\
\quad\quad\quad \mathsf{CONJ\_OF}\, f_2\ (a_1\ \mathsf{DELETE}\ f_0)
\end{array}$$

and for a formula f, we say $\mathsf{DISJ\_OF}\, f\ \mathit{fs}$ when f is either the '$\bot$' or a formula such that $\mathsf{DISJ\_OF0}\, f\ \mathit{fs}$

$$\begin{array}{l}
\vdash (\forall f\ fs.\ f\ \in\ fs\ \Rightarrow\ \mathsf{DISJ\_OF0}\ f\ fs)\ \wedge \\
\quad \forall f_1\ f_2\ fs. \\
\qquad f_1\ \in\ fs\ \wedge\ \mathsf{DISJ\_OF0}\ f_2\ (fs\ \mathsf{DELETE}\ f_1)\ \Rightarrow \\
\qquad \mathsf{DISJ\_OF0}\ (\mathsf{DISJ}\ f_1\ f_2)\ fs\ \ \vdash (\forall f\ fs.\ f\ \in\ fs\ \Rightarrow\ DISJ\_OF_0'\ f\ fs)\ \wedge \\
\quad (\forall f_1\ f_2\ fs. \\
\qquad f_1\ \in\ fs\ \wedge\ DISJ\_OF_0'\ f_2\ (fs\ \mathsf{DELETE}\ f_1)\ \Rightarrow \\
\qquad DISJ\_OF_0'\ (\mathsf{DISJ}\ f_1\ f_2)\ fs)\ \Rightarrow \\
\quad \forall a_0\ a_1.\ \mathsf{DISJ\_OF0}\ a_0\ a_1\ \Rightarrow\ DISJ\_OF_0'\ a_0\ a_1\ \ \vdash \mathsf{DISJ\_OF0}\ a_0\ a_1\ \iff \\
\quad\ a_0\ \in\ a_1\ \vee \\
\qquad \exists f_1\ f_2. \\
\qquad\quad a_0\ =\ \mathsf{DISJ}\ f_1\ f_2\ \wedge\ f_1\ \in\ a_1\ \wedge \\
\qquad\quad \mathsf{DISJ\_OF0}\ f_2\ (a_1\ \mathsf{DELETE}\ f_1)\mathsf{DISJ\_OF}\ f\ fs\ \overset{\mathsf{def}}{=}\ f\ =\ \bot\ \vee\ \mathsf{DISJ\_OF0}\ f\ fs
\end{array}$$

As an example, the four $\mathsf{CONJ\_OF}$ formulas on the set $\{f1, f2\}$ are $\mathsf{AND}\ f_1\ f_2$, $\mathsf{AND}\ (\neg f_1)\ f_2$ $\mathsf{AND}\ f_1\ (\neg f_2)$ and $\mathsf{AND}\ (\neg f_1)\ (\neg f_2)$. For another example, consider the set $fs := \{f1, f2, f3\}$ of three formulas, we can say $\mathsf{CONJ\_OF}\ (\mathsf{AND}\ f_3\ (\mathsf{AND}\ f_1\ f_2))\ fs$, $\mathsf{CONJ\_OF}\ (\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3))\ fs$, $\mathsf{CONJ\_OF}\ (\mathsf{AND}\ (\neg f_3)\ (\mathsf{AND}\ f_2\ f_1))\ fs$ and so on. But it is not the fact that $\mathsf{CONJ\_OF}\ f_1\ fs$, since from the definition, any element of fs or its negation must appear exactly once in a $\mathsf{CONJ\_OF}$ formula on fs. Let $CO := \{f\ |\mathsf{CONJ\_OF}\ f\ fs\}$, then we can say $\mathsf{DISJ\_OF}$
$(\mathsf{DISJ}\ (\mathsf{AND}\ f_3\ (\mathsf{AND}\ f_1\ f_2))\ (\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3)))\ CO$, $\mathsf{DISJ\_OF}\ (\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3))\ CO$ and so on, so both these two formulas are disjunction normal form on fs, but we cannot say $\mathsf{DISJ\_OF}$
$(\mathsf{DISJ}\ (\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3))\ (\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3)))\ CO$ or $\mathsf{DISJ\_OF}$
$(\mathsf{DISJ}\ (\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3))\ (\neg\mathsf{AND}\ f_2\ (\mathsf{AND}\ (\neg f_1)\ f_3)))\ CO$, since any formula in CO or its negation is only allowed to appear in a $\mathsf{DISJ\_OF}$ formula on CO for at most once, so these two are not in disjunction normal form.

If we start with a finite set, by induction on finiteness using `FINITE_COMPLETE_INDUCTION` , we conclude the set of $\mathsf{CONJ\_OF}$ formulas and $\mathsf{DISJ\_OF0}$, hence the $\mathsf{DNF\_OF}$ formulas on this set is finite:

$$\vdash \mathsf{FINITE}\ s\ \Rightarrow\ \mathsf{FINITE}\ \{\ f\ \mid\ \mathsf{CONJ\_OF}\ f\ s\ \}\ \ \vdash \mathsf{FINITE}\ s\ \Rightarrow\ \mathsf{FINITE}\ \{\ f\ \mid\ \mathsf{DISJ\_OF0}\ f\ s\ \}\ \ \vdash \mathsf{FINITE}\ fs\ \Rightarrow\ \mathsf{FINITE}\ \{\ f\ \mid\ \mathsf{DN}$$

Once we prove:

$$\begin{array}{l}
\vdash \mathsf{IBC}\ f\ fs\ \Rightarrow \\
\quad \mathsf{FINITE}\ fs\ \wedge\ fs\ \neq\ \emptyset\ \Rightarrow \\
\quad \exists\, p.\ \mathsf{DNF\_OF}\ p\ fs\ \wedge\ \mathsf{equiv0}\ \mu\ f\ p
\end{array}$$

we get `FINITE_FINITE_IBC`, which leads to a proof of the step case of **`prop_2_29_strengthen`** goes as follows: Suppose $\{f|DEG\ f\ \leq\ n \wedge \forall a.VAR\ a\ \in\ subforms\ f\ \implies\ a\ \in\ s\}//E\ \mu$ is finite and let $A = (\{VAR\ v|v\ \in\ s\} \cup \{\Diamond psi|DEG\ psi \leq n \wedge \forall a.(VAR\ a) \in subforms\ psi\ \implies\ a \in s\})$. By **`DEG_IBC_-strengthen`**, we have $B := \{f|DEG\ f \leq n+1 \wedge \forall a.VAR\ a \in subforms\ f\ \implies\ a \in s\} = \{phi|IBC\ phi\ A\}$, we prove B is finite up to equivalence. If $A = \emptyset$, then by `IBC_EMPTY_lemma`, B contains only two non-equivaelent formulas $\bot$ and $\top$. Otherwise, by `FINITE_FINITE_IBC`, it suffices to prove A is finite up to equivalence. Under the assumption $\mathsf{INFINITE}\ \mathcal{U}(:\beta)$, $\mathsf{equiv0}\ \mu\ (\Diamond\ f)\ (\Diamond\ g)$ iff $\mathsf{equiv0}\ \mu\ f\ g$, so the inductive hypothesis together with the assumption $\mathsf{FINITE}\ s$ gives the finiteness of A.

Therefore, it remains to give a proof of `IBC_DNF_EXISTS`. The proof is by rule induction on IBC, the things to prove are:

Base case:

- The falsity $\bot$ is equivalent to a disjunction normal form on fs.
- For any element $f \in fs$, it is equivalent to a disjunction normal form on fs.

Step case:

- Under the assumption that fs is finite and non-empty, if $f1, f2$ are boolean combination of the set fs which are equivalent to $p1, p2$ of disjunction normal form respectively, then $\mathsf{DISJ}\ f_1\ f_2$ is equivalent to a formula in disjunction normal form.
- With the same assumption on fs, if f is a boolean combination of fs and equivalent to p in disjunction normal form, then $\neg f$ is equivalent to a formula in disjunction normal form.

The first item of base case is trivial since $\bot$ itself is in disjunction normal form, we begin by proving the second item of the base case.

4.2.1 Case for $f \in fs$

We aim to prove:

$$\begin{array}{l}\vdash \mathsf{FINITE}\ fs\ \wedge\ fs\ \neq\ \emptyset\ \Rightarrow \\ \quad \forall f.\ f\ \in\ fs\ \Rightarrow\ \exists p.\ \mathsf{DNF\_OF}\ p\ fs\ \wedge\ \mathsf{equiv0}\ \mu\ f\ p\end{array}$$

Let us consider what does the DNF_OF formula p on fs that is equivalent to and element of fs look like: We require p to be satisfied if and only if f is satisfied. As p is of disjunction normal form, p is satisfied once some of its disjuncts is satisfied. Hence we require the satisfication of each disjuncts of p to imply the satisfication of f, that is, f is a conjunct of each disjunct of p. So up to rearrangement, p is a disjunction of conjunctions $f \wedge f_1 \cdots f_n$ where each formula in $fs/\{f\}$ or its negation appears exactly once in these f_i's. Such a formula is equivalent to $f \wedge g$, where g is the disjunction of the formulas obtained by taking f out of each conjunct of p. And $f \wedge g$ is equivalent to f if and only if g is equivalent to $\top$. Hence, g must be the disjunction of all the CONJ_OF formulas on $fs/\{f\}$. By conclusion, a disjunction normal form we need can be taken as the disjunction of conjunctions starting with f, with its tail ranging over all possible combination of negated and unnegated formulas in $fs/\{f\}$. Use the set $\{f_1, f_2, f_3\}$ as example again, an example of disjunction normal form equivalent to f_1 is the formula $(f_1 \wedge f_2 \wedge f_3) \vee (f_1 \wedge \neg f_2 \wedge f_3) \vee (f_1 \wedge f_2 \wedge \neg f_3) \vee (f_1 \wedge \neg f_2 \wedge \neg f_3)$, note that such a formula is equivalent to $f_1 \wedge ((f_2 \wedge f_3) \vee (\neg f_2 \wedge f_3) \vee (f_2 \wedge \neg f_3) \vee (\neg n_2 \wedge \neg f_3))$, where $(f_2 \wedge f_3) \vee (\neg f_2 \wedge f_3) \vee (f_2 \wedge \neg f_3) \vee (\neg f_2 \wedge \neg f_3)$ is equivalent to TRUE.

To formalise the idea above, we want to be able to building disjunction normal form piece by piece. We use the following definitions as our tool:

$$\begin{array}{l}\mathsf{negf}\ (f, \mathsf{T})\ \stackrel{\text{def}}{=}\ f \\ \mathsf{negf}\ (f, \mathsf{F})\ \stackrel{\text{def}}{=}\ \neg f \mathsf{lit\_list\_to\_form}\ []\ \stackrel{\text{def}}{=}\ \mathsf{TRUE} \\ \mathsf{lit\_list\_to\_form}\ [fb]\ \stackrel{\text{def}}{=}\ \mathsf{negf}\ fb \\ \mathsf{lit\_list\_to\_form}\ (fb :: v_2 :: v_3)\ \stackrel{\text{def}}{=} \\ \quad \mathsf{AND}\ (\mathsf{negf}\ fb)\ (\mathsf{lit\_list\_to\_form}\ (v_2 :: v_3))\mathsf{lit\_list\_to\_form2}\ []\ \stackrel{\text{def}}{=}\ \bot \\ \mathsf{lit\_list\_to\_form2}\ [fb]\ \stackrel{\text{def}}{=}\ fb \\ \mathsf{lit\_list\_to\_form2}\ (fb :: v_2 :: v_3)\ \stackrel{\text{def}}{=} \\ \quad \mathsf{DISJ}\ fb\ (\mathsf{lit\_list\_to\_form2}\ (v_2 :: v_3))\end{array}$$

A pair (f, tv) where $f \in fs$ and tv is a truth value encodes a formula in fs or its negation. Such a pair is turned to a formula f or $\neg f$ by negf, depends on the truth value given in its second coordinate. For a finite fs, we are interested in non-empty lists of such pairs with the condition that set (MAP FST l) $=$ fs and ALL_DISTINCT (MAP FST l), where MAP takes a function f and a list $[a_1; \cdots, a_n]$, and return the list $[f(a_1); \cdots; f(a_n)]$. The function set takes a list and gives the set of its members, and ALL_DISTINCT takes a list and return the truth value of whether all the member of the list. These conditions precisely say that each member of fs occurs exactly once in the list. As a consequence, we can build CONJ_OF formulas from such list using `lit_list_to_form`.

$$\begin{array}{l}\vdash l\ \neq\ []\ \Rightarrow \\ \quad \mathsf{set}\ (\mathsf{MAP\ FST}\ l)\ =\ fs\ \wedge\ \mathsf{ALL\_DISTINCT}\ (\mathsf{MAP\ FST}\ l)\ \Rightarrow \\ \quad \mathsf{CONJ\_OF}\ (\mathsf{lit\_list\_to\_form}\ l)\ fs\end{array}$$

Under the same condition on the list, we obtain DISJ_OF0 formulas using `lit_list_to_form2`.

$$\begin{array}{l}\vdash l\ \neq\ []\ \wedge\ \mathsf{set}\ l\ \subseteq\ fs\ \wedge\ \mathsf{ALL\_DISTINCT}\ l\ \Rightarrow \\ \quad \mathsf{DISJ\_OF0}\ (\mathsf{lit\_list\_to\_form2}\ l)\ fs\end{array}$$

Put the two things together, we can build DNF_OF formulas using `lit_list_to_form` and `list_list_-to_form2`

$$\begin{array}{l}\vdash ld\ \neq\ []\ \wedge\ \mathsf{ALL\_DISTINCT}\ ld\ \wedge \\ \quad (\forall d. \\ \quad\quad \mathsf{MEM}\ d\ ld\ \Rightarrow \\ \quad\quad \exists lc. \\ \quad\quad\quad lc\ \neq\ []\ \wedge\ d\ =\ \mathsf{lit\_list\_to\_form}\ lc\ \wedge \\ \quad\quad\quad \mathsf{set}\ (\mathsf{MAP\ FST}\ lc)\ =\ fs\ \wedge\ \mathsf{ALL\_DISTINCT}\ (\mathsf{MAP\ FST}\ lc))\ \Rightarrow \\ \quad \mathsf{DNF\_OF}\ (\mathsf{lit\_list\_to\_form2}\ ld)\ fs\end{array}$$

This is a report for summarizing my learning of interactive theorem proving in the HOL. This paper consists of two parts, the first part it about the correspondence theorem on group theory, through the process of searching for methods of proving it, I get known to some basics tools, mainly the simplifiers of the HOL. The second part is about modal logic, where I think is my starting point of being exposed to more advanced and useful tools.

5 Correspondence Theorem in the HOL

The main theorem is, stated in usual-sense mathematical language (version on Artin's Algebra):

For a homomorphism $\phi : G \to \mathcal{G}$, call its kernel K. Let H be a subgroup of G that contains K, and let $\mathcal{H}$ be a subgroup of $\mathcal{G}$. Then:
(1) $\phi(H)$ is a subgroup of $\mathcal{G}$.
(2) $\phi^{-1}(\mathcal{H})$ is a subgroup of G.
(3) $\phi^{-1}(\mathcal{H})$ contains K.
(4) $\mathcal{H}$ is a normal subgroup of $\mathcal{G}$ if and only if $\phi^{-1}(\mathcal{H})$ is a normal subgroup of G.
(5) $\phi(\phi^{-1}(\mathcal{H})) = \mathcal{H}$.
(6) $\phi^{-1}(\phi(H)) = H$.
(7) $|\phi^{-1}(\mathcal{H})| = |\mathcal{H}||K|$

In the language of the HOL, the correspondence theorem is stated as:

The $h << g$ denotes the fact that h is a normal subgroup of g. And $h \leq g$ denotes the fact that h is a subgroup of g. All the definitions about group theory are given as functions that take an "object" and returns a boolean to judge whether it satisfies its definition. For instance, *Group g* reads "g is a group".

The right hand side of the arrow is a conjunction, which means that we can split our main goal as smaller subgoals for each conjuncts using `rpt strip_tac` and then prove each subgoal as a lemma. Once we have proved all these lemmas, we can give the remaining work to `metis_tac`.

5.1 lemma 1: $\phi(H)$ is a subgroup of $\mathcal{G}$

My Goal:

In the algebra library there are two versions of definition of subgroup, namely `Subgroup` and `subgroup`. Our goal only involves the version `Subgroup`, which it our $\leq$, but one of the theorems in the library which would be used to prove this goal uses the definition `subgroup`. In order to use it, we need to prove that version of definition of subgroup we use implies the other verision. That is:

As this is proved, we can use the existing theorems

The proof would then be strightforward: As $h \leq g1$ then *subgroup* h g_1, then together with the fact *GroupHomo* f g_1 g_2, by the first lemma *GroupHomo* f h k. And also $h \leq g$ gives *Group* h and *Group* g, then by the last lemma above the goal is proved.

5.2 lemma 2: $\phi^{-1}(\mathcal{H})$ is a subgroup of G

Goal: Note the usage of *preimage_group*, which is a function defined as:

It is a function that take a function f, two "groups" g_1, g_2 of types α, β respectively, and a subset h of g_2.

The word "groups" has a quotation mark because it does not denote the group in mathematical sence, instead, it denotes the datatype "group", which consists by three pieces of informations: the underlying set, which it also called the carrier, the identity element, and the operation function defined for every pair of elements in the carrier and returns an element in the carrier.

We cannot encode the information that f is a homomorphism and the gs are groups directly in the definition of *preimage_group*. This is because we are only allowed to define anything on the whole universe of a type. Also the function *preimage_group* does not directly return a group. We only know that the output of it consists of the three pieces of information that a group needs as mentioned above, and to conclude that it is indeed a group when g_1, g_2 are real groups, f is a homomorphism and h is an actual subgroup, we need to prove that the three pieces of information defined by this sort of inputs of *preimage_group* satisfies the group axioms.

To prove this goal, we mainly want to prove that the preimage of a subgroup of a group under a homomorphism is a group. Once we prove

then the main goal then easily followes from here.

5.3 lemma 3: $K \subseteq \phi^{-1}(\mathcal{H})$

6 Modal logic in the HOL

My Script files on modal logic basically follow the book of Blackburn's book "Modal Logic".

6.1 Chapter 1

6.1.1 Basic Constructions

Everything discussed in this report is with respect to the following modal language-called the basic modal language. In this language, a formulas of a type α are:

$$\begin{aligned}&\alpha\ \texttt{chap1\$form} =\\ &\quad \mathsf{VAR}\ \alpha\\ &\ \mid \mathsf{DISJ}\ (\alpha\ \texttt{chap1\$form})\ (\alpha\ \texttt{chap1\$form})\\ &\ \mid \bot\\ &\ \mid (\neg)\ (\alpha\ \texttt{chap1\$form})\\ &\ \mid \Diamond\ (\alpha\ \texttt{chap1\$form})\end{aligned}$$

The $\Box, \rightarrow, \wedge, \leftrightarrow$ and the tautology are defined as:

$$\Box\,\phi \overset{\text{def}}{=} \neg\Diamond\,(\neg\phi)$$

$$f_1 \rightarrow f_2 \overset{\text{def}}{=} \mathsf{DISJ}\ (\neg f_1)\ f_2$$

$$\mathsf{AND}\ f_1\ f_2 \overset{\text{def}}{=} \neg\mathsf{DISJ}\ (\neg f_1)\ (\neg f_2)$$

$$\mathsf{DOUBLE\_IMP}\ f_1\ f_2 \overset{\text{def}}{=} \mathsf{AND}\ (f_1 \rightarrow f_2)\ (f_2 \rightarrow f_1)$$

$$\mathsf{TRUE} \overset{\text{def}}{=} \neg\bot$$

respectively.

We can define any function that sends each element in the universal set of type α to a β-formula, such a function f induces a map from the collection of all the α-formulas to the collections of β-formulas. It is called the substitution map induced by the function f, it is defined as:

$$\begin{aligned}&\mathsf{subst}\ f\ \bot \overset{\text{def}}{=} \bot\\ &\mathsf{subst}\ f\ (\mathsf{VAR}\ p) \overset{\text{def}}{=} f\ p\\ &\mathsf{subst}\ f\ (\mathsf{DISJ}\ \phi_1\ \phi_2) \overset{\text{def}}{=} \mathsf{DISJ}\ (\mathsf{subst}\ f\ \phi_1)\ (\mathsf{subst}\ f\ \phi_2)\\ &\mathsf{subst}\ f\ (\neg\phi) \overset{\text{def}}{=} \neg\mathsf{subst}\ f\ \phi\\ &\mathsf{subst}\ f\ (\Diamond\ \phi) \overset{\text{def}}{=} \Diamond\ (\mathsf{subst}\ f\ \phi)\end{aligned}$$

Directly from its definition, the substitution rule on a formula involved the $\Box$ is given by:

$$\vdash \mathsf{subst}\ f\ (\Box\ \phi) = \Box\ (\mathsf{subst}\ f\ \phi)$$

We would study formulas in basic modal language on models. A model of type α consists of three pieces of informations: a set of worlds, a relation defined on the set of worlds, and a evaluation map from the set of propositional letters of type α to the power set of the set of worlds. The first two pieces together is called a frame. It is created as a type(for further convenience, we only allow the worlds in a model to be natural numbers):

$$\alpha\ \texttt{frame} = \texttt{<|}\ \mathsf{world} : \alpha \rightarrow \texttt{bool};\ \mathsf{rel} : \alpha \rightarrow \alpha \rightarrow \texttt{bool}\ \texttt{|>}$$

And a model is defined as a type based on a frame, it is:

$$\begin{aligned}&(\alpha,\ \beta)\ \texttt{chap1\$model} = \texttt{<|}\\ &\quad \mathsf{frame} : \beta\ \texttt{frame};\\ &\quad \mathsf{valt} : \alpha \rightarrow \beta \rightarrow \texttt{bool}\\ &\texttt{|>}\end{aligned}$$

A large part of study of formulas of basic modal language is about modal satisfication. Satisfication is defined on the elements of the world set of a model, so we need to ensure that once we say “a formula is satisfied at a world of a model”, we must ensure that the world we are talking about is actually in model. So the definition of satisfication is encoded as:

$$\begin{array}{l}
\mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{VAR}\ p) \overset{\mathsf{def}}{=} w \in \mathfrak{M}.\mathsf{frame}.\mathsf{world} \wedge w \in \mathfrak{M}.\mathsf{valt}\ p \\
\mathsf{satis}\ \mathfrak{M}\ w\ \bot \overset{\mathsf{def}}{=} w \in \mathfrak{M}.\mathsf{frame}.\mathsf{world} \wedge \mathsf{F} \\
\mathsf{satis}\ \mathfrak{M}\ w\ (\neg\phi) \overset{\mathsf{def}}{=} w \in \mathfrak{M}.\mathsf{frame}.\mathsf{world} \wedge \neg\mathsf{satis}\ \mathfrak{M}\ w\ \phi \\
\mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{DISJ}\ \phi_1\ \phi_2) \overset{\mathsf{def}}{=} \mathsf{satis}\ \mathfrak{M}\ w\ \phi_1 \vee \mathsf{satis}\ \mathfrak{M}\ w\ \phi_2 \\
\mathsf{satis}\ \mathfrak{M}\ w\ (\Diamond\ \phi) \overset{\mathsf{def}}{=} \\
\quad w \in \mathfrak{M}.\mathsf{frame}.\mathsf{world} \wedge \\
\quad \exists\, v.\ \mathfrak{M}.\mathsf{frame}.\mathsf{rel}\ w\ v \wedge v \in \mathfrak{M}.\mathsf{frame}.\mathsf{world} \wedge \mathsf{satis}\ \mathfrak{M}\ v\ \phi
\end{array}$$

Thus latter if we want to to prove a formula is satisfied at some point in the model, firstly we need the fact that the point is indeed in the world set of the model.

Instead of regarding a single formula, we can also say a set of formulas of some fixed type is satisfied at a world. We say a set of formulas is satisfied at a point of a model when each formula in the set is satisfied in at this point. In the HOL, it looks like:

$$\mathsf{satis\_set}\ \mathfrak{M}\ w\ \Sigma \overset{\mathsf{def}}{=} \forall\, a.\ a \in \Sigma \Rightarrow \mathsf{satis}\ \mathfrak{M}\ w\ a$$

Regarding a formula and its satisfiability on a given model, it is universally true if it is satisfied at every world in the model, it is satisfiable if there exists some worlds in the model where it is satisfied, and it is refutable if there exists some worlds in the model where its negation is satisfied, just as:

$$\mathsf{stfable}\ \mathfrak{M}\ \phi \overset{\mathsf{def}}{=} \exists\, w.\ \mathsf{satis}\ \mathfrak{M}\ w\ \phi$$

$$\mathsf{refutable}\ \mathfrak{M}\ \phi \overset{\mathsf{def}}{=} \exists\, w.\ \mathsf{satis}\ \mathfrak{M}\ w\ (\neg\phi)$$

We use a bunch of definition regarding “validness” to describe the satisfiability of a formula over a frame. In fact, the validness of a formula can be defined on the level of a state of a frame, a frame or a class of frames. A formula ϕ is valid at a state w in a frame if for any model over this frame, ϕ is satisfied at w. ϕ is valid in a frame if it is valid at every state in the frame, and is valid in a class C of frames if for any frame f in C, ϕ is valid in f. Finally, to say a formula is valid is to say that for any frame f, it is valid in f. These definitions are expressed as follows:

$$\mathsf{valid\_frame\_state}\ f\ w\ \phi \overset{\mathsf{def}}{=} \forall\,\mathfrak{M}.\ \mathfrak{M}.\mathsf{frame} = f \Rightarrow \mathsf{satis}\ \mathfrak{M}\ w\ \phi$$

$$\mathsf{valid\_frame}\ f\ \phi \overset{\mathsf{def}}{=} \forall\, w.\ \mathsf{valid\_frame\_state}\ f\ w\ \phi$$

$$\mathsf{valid\_class}\ C\ \phi \overset{\mathsf{def}}{=} \forall\, f.\ f \in C \Rightarrow \mathsf{valid\_frame}\ f\ \phi$$

$$\vdash \mathsf{valid}\ \phi \iff \forall\, f.\ \mathsf{valid\_frame}\ f\ \phi$$

For a class of frames, the “logic” of it is the collection of all formulas that are valid in every frames in this class:

$$\mathsf{LOGIC}\ C \overset{\mathsf{def}}{=} \{\ \phi \mid \mathsf{valid\_class}\ C\ \phi\ \}$$

For a set S of α-models and a set Σ for α-formulas, saying an α-formula ϕ is a local semantic consequence of Σ over S is saying that for any model M in S, and any point w in M, if Σ is satisfied, then ϕ must be satisfied. Stated in the language of HOL, it is:

$$\mathsf{LSC}\ \Sigma\ S\ \phi \overset{\mathsf{def}}{=} \forall\,\mathfrak{M}\ w.\ \mathfrak{M} \in S \wedge \mathsf{satis\_set}\ \mathfrak{M}\ w\ \Sigma \Rightarrow \mathsf{satis}\ \mathfrak{M}\ w\ \phi$$

So far what we have already defined are all about semantics, and in fact, there is a syntactic mechanism of obtaining the collection of valid formulas. This syntactic mechanism involves the usage of K-proof. In usual language, it is defined as:

A K-proof is a finite sequence of formulas, each of which is an axiom, or follows from one or more earlier items in the sequence by applying a rule of proof.

For the axioms, they are all propositional tautologies plus:
$\Box(p \to q) \to (\Box p \to \Box q)$
$\Diamond p \leftrightarrow \neg\Box\neg p$

And the rules are: (1) Modens ponens: if ϕ and $\phi \to \psi$ both occurs as lines of some line of a K-proof, then we can append a line ψ so the resulting list is also a K-proof.
(2) Uniform substitution: if ϕ occurs as a line of a K-proof, then we can append a line θ where θ is obtained from ϕ by applying some substitution function.
(3) Generalization: if ϕ occurs as a line of a K-proof, then we can append a line $\Box\phi$ to the original list and obtain a K-proof.

Given the rules of the K-proof, we can know about three things: the rule itself which defines what sort of list is called a K-proof, a induction principle according to the rule which allows us to use in proving things about Kproofs by induction, and if a given list is a K-proof, what are the possible cases of its form. So we can define these three things altogether, with the application of `Hol_reln`:

It gives three statements at once, namely:

$$\begin{array}{l}
\vdash \mathsf{Kproof}\ []\ \wedge \\
\quad (\forall p\ \phi_1\ \phi_2. \\
\qquad \mathsf{Kproof}\ p\ \wedge\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p\ \wedge\ \mathsf{MEM}\ \phi_1\ p\ \Rightarrow \\
\qquad \mathsf{Kproof}\ (p\ +\!\!+\ [\phi_2]))\ \wedge \\
\quad (\forall p\ \phi\ f.\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathsf{Kproof}\ (p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]))\ \wedge \\
\quad (\forall p\ \phi.\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathsf{Kproof}\ (p\ +\!\!+\ [\Box\ \phi]))\ \wedge \\
\quad (\forall p\ \phi_1\ \phi_2. \\
\qquad \mathsf{Kproof}\ p\ \Rightarrow\ \mathsf{Kproof}\ (p\ +\!\!+\ [\Box\ (\phi_1\ \to\ \phi_2)\ \to\ \Box\ \phi_1\ \to\ \Box\ \phi_2]))\ \wedge \\
\quad (\forall p\ \phi.\ \mathsf{Kproof}\ p\ \Rightarrow\ \mathsf{Kproof}\ (p\ +\!\!+\ [\Diamond\ \phi\ \to\ \neg\Box\ (\neg\phi)]))\ \wedge \\
\quad (\forall p\ \phi.\ \mathsf{Kproof}\ p\ \Rightarrow\ \mathsf{Kproof}\ (p\ +\!\!+\ [\neg\Box\ (\neg\phi)\ \to\ \Diamond\ \phi]))\ \wedge \\
\quad \forall p\ f.\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{ptaut}\ f\ \Rightarrow\ \mathsf{Kproof}\ (p\ +\!\!+\ [f])
\end{array}$$

$$\begin{array}{l}
\vdash \mathit{Kproof}'\ []\ \wedge \\
\quad (\forall p\ \phi_1\ \phi_2. \\
\qquad \mathit{Kproof}'\ p\ \wedge\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p\ \wedge\ \mathsf{MEM}\ \phi_1\ p\ \Rightarrow \\
\qquad \mathit{Kproof}'\ (p\ +\!\!+\ [\phi_2]))\ \wedge \\
\quad (\forall p\ \phi\ f.\ \mathit{Kproof}'\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathit{Kproof}'\ (p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]))\ \wedge \\
\quad (\forall p\ \phi.\ \mathit{Kproof}'\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathit{Kproof}'\ (p\ +\!\!+\ [\Box\ \phi]))\ \wedge \\
\quad (\forall p\ \phi_1\ \phi_2. \\
\qquad \mathit{Kproof}'\ p\ \Rightarrow\ \mathit{Kproof}'\ (p\ +\!\!+\ [\Box\ (\phi_1\ \to\ \phi_2)\ \to\ \Box\ \phi_1\ \to\ \Box\ \phi_2]))\ \wedge \\
\quad (\forall p\ \phi.\ \mathit{Kproof}'\ p\ \Rightarrow\ \mathit{Kproof}'\ (p\ +\!\!+\ [\Diamond\ \phi\ \to\ \neg\Box\ (\neg\phi)]))\ \wedge \\
\quad (\forall p\ \phi.\ \mathit{Kproof}'\ p\ \Rightarrow\ \mathit{Kproof}'\ (p\ +\!\!+\ [\neg\Box\ (\neg\phi)\ \to\ \Diamond\ \phi]))\ \wedge \\
\quad (\forall p\ f.\ \mathit{Kproof}'\ p\ \wedge\ \mathsf{ptaut}\ f\ \Rightarrow\ \mathit{Kproof}'\ (p\ +\!\!+\ [f]))\ \Rightarrow \\
\quad \forall a_0.\ \mathsf{Kproof}\ a_0\ \Rightarrow\ \mathit{Kproof}'\ a_0
\end{array}$$

$$\begin{array}{l}
\vdash \mathsf{Kproof}\ a_0\ \iff \\
\quad a_0\ =\ []\ \vee \\
\quad (\exists p\ \phi_1\ \phi_2. \\
\qquad a_0\ =\ p\ +\!\!+\ [\phi_2]\ \wedge\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p\ \wedge \\
\qquad \mathsf{MEM}\ \phi_1\ p)\ \vee \\
\quad (\exists p\ \phi\ f.\ a_0\ =\ p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]\ \wedge\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{MEM}\ \phi\ p)\ \vee \\
\quad (\exists p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\Box\ \phi]\ \wedge\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{MEM}\ \phi\ p)\ \vee \\
\quad (\exists p\ \phi_1\ \phi_2. \\
\qquad a_0\ =\ p\ +\!\!+\ [\Box\ (\phi_1\ \to\ \phi_2)\ \to\ \Box\ \phi_1\ \to\ \Box\ \phi_2]\ \wedge\ \mathsf{Kproof}\ p)\ \vee \\
\quad (\exists p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\Diamond\ \phi\ \to\ \neg\Box\ (\neg\phi)]\ \wedge\ \mathsf{Kproof}\ p)\ \vee \\
\quad (\exists p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\neg\Box\ (\neg\phi)\ \to\ \Diamond\ \phi]\ \wedge\ \mathsf{Kproof}\ p)\ \vee \\
\quad \exists p\ f.\ a_0\ =\ p\ +\!\!+\ [f]\ \wedge\ \mathsf{Kproof}\ p\ \wedge\ \mathsf{ptaut}\ f
\end{array}$$

The significance of the K-proof system can be seen through this fact: A basic modal formula is K-provable if and only if it is valid.

The definition of normal modal logic is a direct abstraction from the definition of the K-proof system. It is defined as a set of formulas closed under modens ponens, uniform substitution and generalization, Which is expressed as:

$$\begin{aligned}
&\mathsf{NML}\ S \overset{\text{def}}{=}\\
&\quad \forall A\ B\ p\ q\ f\ \phi.\\
&\qquad (\mathsf{ptaut}\ \phi \Rightarrow \phi \in S) \land (\Box\,(p \to q) \to \Box\, p \to \Box\, q) \in S \land\\
&\qquad (\Diamond\, p \to \neg\Box\,(\neg p)) \in S \land (\neg\Box\,(\neg p) \to \Diamond\, p) \in S \land\\
&\qquad (A \in S \Rightarrow \mathsf{subst}\, f\ A \in S) \land (A \in S \Rightarrow \Box\, A \in S) \land\\
&\qquad ((A \to B) \in S \land A \in S \Rightarrow B \in S)
\end{aligned}$$

6.1.2 Exercises

There are two exercises in chapter 1.6 which have been proved in my file.

Exercise 1.6.2 Goal: Let ϕ^- be the 'demodalized' version of a modal formula ϕ, that is, ϕ^- is obtained from ϕ by erasing all the diamonds. Then if ϕ is K-provable, then ϕ^- is a propositional tautology.

To solve this exercise, firstly we need to define the demodalization for a formula. Demodalizing is the operation that for diamonded formulas, removing the diamond, and extending to all the formulas, that is(quite similar to the definition of substitution map):

$$\begin{aligned}
&\mathsf{demodalize}\ \bot \overset{\text{def}}{=} \bot\\
&\mathsf{demodalize}\ (\mathsf{VAR}\ p) \overset{\text{def}}{=} \mathsf{VAR}\ p\\
&\mathsf{demodalize}\ (\mathsf{DISJ}\ \phi_1\ \phi_2) \overset{\text{def}}{=} \mathsf{DISJ}\ (\mathsf{demodalize}\ \phi_1)\ (\mathsf{demodalize}\ \phi_2)\\
&\mathsf{demodalize}\ (\neg\phi) \overset{\text{def}}{=} \neg\mathsf{demodalize}\ \phi\\
&\mathsf{demodalize}\ (\Diamond\, \phi) \overset{\text{def}}{=} \mathsf{demodalize}\ \phi
\end{aligned}$$

Secondly, we need to state clearly which kind of formula is a propositional tautology. To do this, firstly we need to define what is a propositional formula, then the definition of propositional tautology would be given by: for a formula ϕ, ϕ is a propositional tautology if and only if ϕ is a propositional tautology, and ϕ 'always holds'.

A propositional formula is just a modal formula without the diamond, so it can be defined as:

$$\begin{aligned}
&\mathsf{propform}\ (\mathsf{VAR}\ p) \overset{\text{def}}{=} \mathsf{T}\\
&\mathsf{propform}\ (\mathsf{DISJ}\ \phi_1\ \phi_2) \overset{\text{def}}{=} \mathsf{propform}\ \phi_1 \land \mathsf{propform}\ \phi_2\\
&\mathsf{propform}\ (\neg f) \overset{\text{def}}{=} \mathsf{propform}\ f\\
&\mathsf{propform}\ \bot \overset{\text{def}}{=} \mathsf{T}\\
&\mathsf{propform}\ (\Diamond\, f) \overset{\text{def}}{=} \mathsf{F}
\end{aligned}$$

Similar as mentioned in section on the correspondence theorem. propform is a function that takes a formula and returns a truth-value, that is, it judges whether a modal formula is a propositional formula.

Then we will need to describe what is 'always holds' for a propositional formula. Here we describe this property as: for any assignment of truth values of propositional symbols involved in this formula, the truth value of the whole formula must be T. Here we need to define a evaluation function that can give us a truth value of a propositional formula for each assignment of the propositional symbols. it is defined as:

$$\begin{aligned}
&\mathsf{peval}\ \sigma\ (\mathsf{VAR}\ p) \overset{\text{def}}{=} \sigma\ p\\
&\mathsf{peval}\ \sigma\ (\mathsf{DISJ}\ f_1\ f_2) \overset{\text{def}}{=} \mathsf{peval}\ \sigma\ f_1 \lor \mathsf{peval}\ \sigma\ f_2\\
&\mathsf{peval}\ \sigma\ \bot \overset{\text{def}}{=} \mathsf{F}\\
&\mathsf{peval}\ \sigma\ (\neg f) \overset{\text{def}}{=} \neg\mathsf{peval}\ \sigma\ f\\
&\mathsf{peval}\ \sigma\ (\Diamond\, f) \overset{\text{def}}{=} \mathsf{F}
\end{aligned}$$

Here σ is an assignment of truth-values on the propositional symbols, it is a function sending each propositional symbol to its assigned truth-value.

Thus as explained above, we can define propositional tautology as:

$$\mathsf{ptaut}\ f \overset{\text{def}}{=} \mathsf{propform}\ f \land \forall\sigma.\ \mathsf{peval}\ \sigma\ f \iff \mathsf{T}$$

Have these definitions constructed, now we are aim to prove that once a formula ϕ is K-provable, means that it occurs at the bottom line of a K-proof, then ϕ^- is a propostional tautology. Noting that almost everything here is defined inductively, so this goal is suitable to be prove by induction on 'K-proof'. The trick is a stronger goal gives a stronger inductive hypothesis, which can be more useful. Here to make the inductive hypothesis more useful, we strengthen the goal as:

$$\vdash \mathsf{Kproof}\ p \ \Rightarrow\ \forall f.\ \mathsf{MEM}\ f\ p\ \Rightarrow\ \mathsf{ptaut}\ (\mathsf{demodalize}\ f)$$

That is, for every line in a K-proof, it can be demodalized to give a propositional tautology.

The induction gives subgoals looks like:

$$\begin{array}{l} \mathsf{ptaut}\ (\mathsf{demodalize}\ f) \\ \hline 0.\ \mathsf{Kproof}\ p \\ 1.\ \forall f.\ \mathsf{MEM}\ f\ p\ \Rightarrow\ \mathsf{ptaut}\ (\mathsf{demodalize}\ f) \\ 2.\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p \\ 3.\ \mathsf{MEM}\ \phi_1\ p \\ 4.\ \mathsf{MEM}\ f\ (p\ +\!\!+\ [\phi_2]) \end{array}$$

The line above the solid line is the goal, and the lines under the line are the assumptions. To prove it, observing that by assumption 4, either f in p or $f\ =\ \phi_2$. For the first case, the inductive hypothesis, which is assumption 1, applies. For the second case, by assumption 3 and the inductive hypothesis, $\phi_1\ \to\ \phi_2$ is a propositional tautology, and then we can prove

$$\vdash \mathsf{ptaut}\ (f_1\ \to\ f_2)\ \wedge\ \mathsf{ptaut}\ f_1\ \Rightarrow\ \mathsf{ptaut}\ f_2$$

from definitions "peval" and "ptaut".

Then by the lemma above, the subgoal is solved. All the subgoals except the following one can be solved in a similar way.

The trickest subgoal is:

$$\begin{array}{l} \mathsf{ptaut}\ (\mathsf{demodalize}\ f') \\ \hline 0.\ \mathsf{Kproof}\ p \\ 1.\ \forall f.\ \mathsf{MEM}\ f\ p\ \Rightarrow\ \mathsf{ptaut}\ (\mathsf{demodalize}\ f) \\ 2.\ \mathsf{MEM}\ \phi\ p \\ 3.\ \mathsf{MEM}\ f'\ (p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]) \end{array}$$

As before, f' is either in p or is $\mathsf{subst}\ f\ \phi$. The former case is trivial by assumption 1. And for the later case, we aim to prove that $\mathsf{demodalize}\ (\mathsf{subst}\ f\ \phi)$ is a propositional tautology, means that $\forall \sigma.\ \mathsf{peval}\ \sigma\ (\mathsf{demodalize}\ (\mathsf{subst}\ f\ \phi))\ \iff\ \mathsf{T}$. As as ϕ is in the list p, by the inductive hypothesis, $\mathsf{demodalize}\ \phi$ is a tautology. So once we prove that
$\forall \sigma.$
$\exists \sigma'.$
$\mathit{pevel}\ \sigma\ \mathsf{demodalize}\ (\mathsf{subst}\ f\ \phi)\ \iff$
$\mathsf{peval}\ \sigma'\ (\mathsf{demodalize}\ \phi)$, as the right hand side has truth value T, then we are done.

So we prove:

$$\begin{array}{l} \vdash \mathsf{propform}\ \phi\ \Rightarrow \\ \quad (\mathsf{peval}\ \sigma\ (\mathsf{demodalize}\ (\mathsf{subst}\ f\ \phi))\ \iff \\ \qquad \mathsf{peval}\ (\mathsf{peval}\ \sigma\ \circ\ \mathsf{demodalize}\ \circ\ f)\ \phi) \end{array}$$

by inducting on ϕ and using the definition of demodalization.

Note that $\mathsf{demodalize}\ f$ is a propositional formula so the above lemma applies, and after proving:

$$\begin{array}{l} \vdash \mathsf{demodalize}\ (\mathsf{subst}\ f\ \phi)\ = \\ \quad \mathsf{demodalize}\ (\mathsf{subst}\ f\ (\mathsf{demodalize}\ \phi)) \end{array}$$

We can conclude $\mathsf{peval}\ \sigma\ (\mathsf{demodalize}\ (\mathsf{subst}\ f\ (\mathsf{demodalize}\ \phi)))\ \iff$
$\mathsf{peval}\ \sigma\ (\mathsf{demodalize}\ (\mathsf{subst}\ f\ \phi)) = \mathsf{peval}\ (\mathsf{peval}\ \sigma\ \circ\ \mathsf{demodalize}\ \circ\ f)\ \phi\ \iff\ \mathsf{T}$.

Exercise 1.6.6 Goal: Given a set of formulas Γ. Suppose we form the axiom system $K\Gamma$ by adding as axioms all the formulas in Γ to K. Then the Hilbert system $K\Gamma$ proves precisely the formulas contained in the normal modal logic $K\Gamma$.

Firstly, we setup the Hilbert system KΓ by defining the rule of proving things in $K\Gamma$. Almost the same as the definition of K-proof, we use `Hol_reln` to give:

$$\begin{array}{l}
\vdash \mathsf{KGproof}\ \Gamma\ []\ \wedge \\
\quad (\forall\, p\ \phi_1\ \phi_2. \\
\qquad \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p\ \wedge\ \mathsf{MEM}\ \phi_1\ p\ \Rightarrow \\
\qquad \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\phi_2]))\ \wedge \\
\quad (\forall\, p\ \phi\ f. \\
\qquad \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\Box\ \phi]))\ \wedge \\
\quad (\forall\, p\ \phi_1\ \phi_2. \\
\qquad \mathsf{KGproof}\ \Gamma\ p\ \Rightarrow \\
\qquad \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\Box\ (\phi_1\ \to\ \phi_2)\ \to\ \Box\ \phi_1\ \to\ \Box\ \phi_2]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathsf{KGproof}\ \Gamma\ p\ \Rightarrow\ \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\Diamond\ \phi\ \to\ \neg\Box\ (\neg\phi)]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathsf{KGproof}\ \Gamma\ p\ \Rightarrow\ \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\neg\Box\ (\neg\phi)\ \to\ \Diamond\ \phi]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{ptaut}\ \phi\ \Rightarrow\ \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\phi]))\ \wedge \\
\quad \forall\, p\ \phi.\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \phi\ \in\ \Gamma\ \Rightarrow\ \mathsf{KGproof}\ \Gamma\ (p\ +\!\!+\ [\phi])
\end{array}$$

$$\begin{array}{l}
\vdash \mathit{KGproof}'\ []\ \wedge \\
\quad (\forall\, p\ \phi_1\ \phi_2. \\
\qquad \mathit{KGproof}'\ p\ \wedge\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p\ \wedge\ \mathsf{MEM}\ \phi_1\ p\ \Rightarrow \\
\qquad \mathit{KGproof}'\ (p\ +\!\!+\ [\phi_2]))\ \wedge \\
\quad (\forall\, p\ \phi\ f.\ \mathit{KGproof}'\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathit{KGproof}'\ (p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathit{KGproof}'\ p\ \wedge\ \mathsf{MEM}\ \phi\ p\ \Rightarrow\ \mathit{KGproof}'\ (p\ +\!\!+\ [\Box\ \phi]))\ \wedge \\
\quad (\forall\, p\ \phi_1\ \phi_2. \\
\qquad \mathit{KGproof}'\ p\ \Rightarrow \\
\qquad \mathit{KGproof}'\ (p\ +\!\!+\ [\Box\ (\phi_1\ \to\ \phi_2)\ \to\ \Box\ \phi_1\ \to\ \Box\ \phi_2]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathit{KGproof}'\ p\ \Rightarrow\ \mathit{KGproof}'\ (p\ +\!\!+\ [\Diamond\ \phi\ \to\ \neg\Box\ (\neg\phi)]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathit{KGproof}'\ p\ \Rightarrow\ \mathit{KGproof}'\ (p\ +\!\!+\ [\neg\Box\ (\neg\phi)\ \to\ \Diamond\ \phi]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathit{KGproof}'\ p\ \wedge\ \mathsf{ptaut}\ \phi\ \Rightarrow\ \mathit{KGproof}'\ (p\ +\!\!+\ [\phi]))\ \wedge \\
\quad (\forall\, p\ \phi.\ \mathit{KGproof}'\ p\ \wedge\ \phi\ \in\ \Gamma\ \Rightarrow\ \mathit{KGproof}'\ (p\ +\!\!+\ [\phi]))\ \Rightarrow \\
\quad \forall\, a_0.\ \mathsf{KGproof}\ \Gamma\ a_0\ \Rightarrow\ \mathit{KGproof}'\ a_0
\end{array}$$

$$\begin{array}{l}
\vdash \mathsf{KGproof}\ \Gamma\ a_0\ \iff \\
\quad a_0\ =\ []\ \vee \\
\quad (\exists\, p\ \phi_1\ \phi_2. \\
\qquad a_0\ =\ p\ +\!\!+\ [\phi_2]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{MEM}\ (\phi_1\ \to\ \phi_2)\ p\ \wedge \\
\qquad \mathsf{MEM}\ \phi_1\ p)\ \vee \\
\quad (\exists\, p\ \phi\ f.\ a_0\ =\ p\ +\!\!+\ [\mathsf{subst}\ f\ \phi]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{MEM}\ \phi\ p)\ \vee \\
\quad (\exists\, p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\Box\ \phi]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{MEM}\ \phi\ p)\ \vee \\
\quad (\exists\, p\ \phi_1\ \phi_2. \\
\qquad a_0\ =\ p\ +\!\!+\ [\Box\ (\phi_1\ \to\ \phi_2)\ \to\ \Box\ \phi_1\ \to\ \Box\ \phi_2]\ \wedge \\
\qquad \mathsf{KGproof}\ \Gamma\ p)\ \vee \\
\quad (\exists\, p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\Diamond\ \phi\ \to\ \neg\Box\ (\neg\phi)]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p)\ \vee \\
\quad (\exists\, p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\neg\Box\ (\neg\phi)\ \to\ \Diamond\ \phi]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p)\ \vee \\
\quad (\exists\, p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\phi]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \mathsf{ptaut}\ \phi)\ \vee \\
\quad \exists\, p\ \phi.\ a_0\ =\ p\ +\!\!+\ [\phi]\ \wedge\ \mathsf{KGproof}\ \Gamma\ p\ \wedge\ \phi\ \in\ \Gamma
\end{array}$$

Also similar to the mentioned definition of normal modal logic, we define the normal modal logic KΓ as:

$$\mathsf{NMLG}\ \Gamma\ \overset{\mathsf{def}}{=}\ \mathsf{BIGINTER}\ \{\ A\ \mid\ \mathsf{NML}\ A\ \wedge\ \Gamma\ \subseteq\ A\ \}$$

$\mathsf{NMLG}\ \Gamma$ is the smallest normal modal logic containing Γ. Once we mention the word "smallest", we know that there is a induction principle here, we can derive this principle as:

$$\begin{array}{l}
\vdash (\forall\, A\ B\ p\ q\ f\ \phi. \\
\qquad (\mathsf{ptaut}\ \phi\ \Rightarrow\ \phi\ \in\ P)\ \wedge\ (\Box\ (p\ \to\ q)\ \to\ \Box\ p\ \to\ \Box\ q)\ \in\ P\ \wedge \\
\qquad (\Diamond\ p\ \to\ \neg\Box\ (\neg p))\ \in\ P\ \wedge\ (\neg\Box\ (\neg p)\ \to\ \Diamond\ p)\ \in\ P\ \wedge \\
\qquad (A\ \in\ P\ \Rightarrow\ \mathsf{subst}\ f\ A\ \in\ P)\ \wedge\ (A\ \in\ P\ \Rightarrow\ \Box\ A\ \in\ P)\ \wedge \\
\qquad ((A\ \to\ B)\ \in\ P\ \wedge\ A\ \in\ P\ \Rightarrow\ B\ \in\ P))\ \wedge\ G\ \subseteq\ P\ \Rightarrow \\
\quad \forall\,\phi.\ \phi\ \in\ \mathsf{NMLG}\ G\ \Rightarrow\ \phi\ \in\ P
\end{array}$$

This principle is used to derive a conclusion that once a formula is in the normal modal logic, the predicate P holds for this formula. In one direction, we are proving that all the formulas in the normal modal logic is $K\Gamma$-provable. The goal looks like:

$$\vdash f \in \mathsf{NMLG}\ \Gamma \Rightarrow f \in \{\ \phi \mid \mathsf{KG\_provable}\ \Gamma\ \phi\ \}$$

It matchs the conclusion of our induction principle exactly, so we can use `ho_match_mp_tac` to match it onto the induction principle above, and the remaining part of the proof would become straightforward.

The other half the goal, which is saying any $K\Gamma$-provable formula is in the normal modal logic $K\Gamma$:

$$\vdash \mathsf{KGproof}\ \Gamma\ p \Rightarrow \forall \phi.\ \mathsf{MEM}\ \phi\ p \Rightarrow \phi \in \mathsf{NMLG}\ \Gamma$$

is proved by induct on KGproof.

6.2 Chapter 2

6.2.1 2.1 Invariance Results

Section 2.1 of Blackburn's textbook is talking about obtaining new models from old ones without affecting modal satisfication. There are basically 3 ways to do it:

Disjoint union To define disjoint union in the HOL, noting that the disjoint union is defined on a family of models. For further convenience, we will only allow the elements of the index set to be natural numbers, so a family of models indexed by natural numbers is just a function that assigns each natural number in the index set a model. So taking the disjoint union is taking such a function and a index set as inputs, and returns a model:

```
DU (f, dom) ≝
  <|frame :=
    <|world :=
       { w | FST w ∈ dom ∧ SND w ∈ (f (FST w)).frame.world };
      rel :=
       (λ w₁ w₂.
          FST w₁ = FST w₂ ∧ FST w₁ ∈ dom ∧
          (f (FST w₁)).frame.rel (SND w₁) (SND w₂))|>;
   valt := (λ v w. (f (FST w)).valt v (SND w))|>
```

As for a model in a disjoint union, both the index of the model any world in the model are natural numbers, if w is a world of a model M_n, then the copy of the world w in the disjoint union is expressed as $n \otimes w$.So the worlds in the disjoint union are all "npair"s. And the two worlds in the disjoint union is related iff they come from the same model, and in the model that they come from they are related. The fact that two worlds n_1, n_2 in the disjoint union come from the same model is encoded as the first coordinate of them, which corresponds to the index of the model they comes from, are the same, that is, $\mathsf{nfst}\ n_1 = \mathsf{nfst}\ n_2$. And we can see from the code above that the for a propostional letter v, an element n in the disjoint union is in $V(v)$ if and only if in the world indexed by $nfstn$ that n comes from, the point $nsndn$ in that model that corresponds to n is in the evalution set of v.

The special case of the union of two models M_1, M_2 is given by:

$$\mathsf{M\_union}\ \mathfrak{M}_1\ \mathfrak{M}_2 \stackrel{\text{def}}{=} \mathsf{DU}\ ((\lambda n.\ \mathtt{if}\ n = 0\ \mathtt{then}\ \mathfrak{M}_1\ \mathtt{else}\ \mathfrak{M}_2), \{\ x \mid x = 0 \lor x = 1\ \})$$

That is, taking the dom to be the two-element set.

We can induct on the structure of formula to prove the invarience result for disjoint unions, which says a formula is satisfies in a world in a disjoint union iff is it satisfied at the corresponding world in the model that this world comes from:

$$\begin{aligned}&\vdash \mathsf{FST}\ w \in dom \Rightarrow\\ &\quad (\mathsf{satis}\ (f\ (\mathsf{FST}\ w))\ (\mathsf{SND}\ w)\ phi \iff\\ &\quad\ \ \mathsf{satis}\ (\mathsf{DU}\ (f, dom))\ w\ phi)\end{aligned}$$

generated submodel For two models $(\mathfrak{M}_1, \mathfrak{M}_2)$, we say $\mathfrak{M}_1$ is a submodel of $\mathfrak{M}_2$ if the world set of $\mathfrak{M}_1$ is a subset of that of $\mathfrak{M}_2$, and the relation and evaluation map of $\mathfrak{M}_1$ is given by that of $\mathfrak{M}_2$ restricted on the world set of $\mathfrak{M}_1$.

```
SUBMODEL $\mathfrak{M}_1$ $\mathfrak{M}_2$ $\overset{\text{def}}{=}$
 $\mathfrak{M}_1$.frame.world $\subseteq$ $\mathfrak{M}_2$.frame.world $\wedge$
 $\forall w_1$.
   $w_1 \in \mathfrak{M}_1$.frame.world $\Rightarrow$
   ($\forall v$. $\mathfrak{M}_1$.valt $v\ w_1 \iff \mathfrak{M}_2$.valt $v\ w_1$) $\wedge$
   $\forall w_2$.
     $w_2 \in \mathfrak{M}_1$.frame.world $\Rightarrow$
     ($\mathfrak{M}_1$.frame.rel $w_1\ w_2 \iff \mathfrak{M}_2$.frame.rel $w_1\ w_2$)
```

The generated submodel is a concept defined upon submodels. Saying a submodel $\mathfrak{M}_1$ is a generating submodel of $\mathfrak{M}_2$, we mean that for any point v in $\mathfrak{M}_1$, if the relation defined on $\mathfrak{M}_2$ relates it to a point w of $\mathfrak{M}_2$, then the w must be in the world set of $\mathfrak{M}_1$ as well:

```
GENSUBMODEL $\mathfrak{M}_1$ $\mathfrak{M}_2$ $\overset{\text{def}}{=}$
 SUBMODEL $\mathfrak{M}_1$ $\mathfrak{M}_2$ $\wedge$
 $\forall w_1$.
   $w_1 \in \mathfrak{M}_1$.frame.world $\Rightarrow$
   $\forall w_2$.
     $w_2 \in \mathfrak{M}_2$.frame.world $\wedge$ $\mathfrak{M}_2$.frame.rel $w_1\ w_2 \Rightarrow$
     $w_2 \in \mathfrak{M}_1$.frame.world
```

The invarience theorem of generated submodel stated in the HOL is:

$$\vdash \text{GENSUBMODEL}\ \mathfrak{M}_1\ \mathfrak{M}_2 \wedge n \in \mathfrak{M}_1.\text{frame.world} \Rightarrow (\text{satis}\ \mathfrak{M}_1\ n\ phi \iff \text{satis}\ \mathfrak{M}_2\ n\ phi)$$

It says that if M_1 is a generated submodel of M_2, then for any formula ϕ, ϕ is satisfied at the world n in M_1 if and only if ϕ is satisfied at n in M_2. As from the definition of generated submodel, the world set of a generated submodel M_1 of M_2 is a subset of the world set of M_2, the corresponding point of n in M_2 is just n itself.

morphisms for modalities In this section we firstly give a bunch of definitions of morphisms between models, they are all maps between elements of world sets satisfying some extra conditions:

If $f : W \to W'$ is a function from the world set of a model $\mathfrak{M}$ to the world set of a model $\mathfrak{M}'$ then it is a homomorphism if

(i) For each propositional letter p and each element w from $\mathfrak{M}$, if $w \in V(p)$, then $f(w) \in V'(p)$.
(ii) If Rwu then $R'f(w)f(u)$.

```
hom $f$ $\mathfrak{M}_1$ $\mathfrak{M}_2$ $\overset{\text{def}}{=}$
 $\forall w$.
   $w \in \mathfrak{M}_1$.frame.world $\Rightarrow$
   $f\ w \in \mathfrak{M}_2$.frame.world $\wedge$
   ($\forall p$. $w \in \mathfrak{M}_1$.valt $p \Rightarrow f\ w \in \mathfrak{M}_2$.valt $p$) $\wedge$
   $\forall u$.
     $u \in \mathfrak{M}_1$.frame.world $\Rightarrow$
     $\mathfrak{M}_1$.frame.rel $w\ u \Rightarrow$
     $\mathfrak{M}_2$.frame.rel ($f\ w$) ($f\ u$)
```

It is a strong homomorphism if

(i) For each propositional letter p and each element w from $\mathfrak{M}$, $w \in V(p)$ iff $f(w) \in V'(p)$.
(ii) If Rwu then $R'f(w)f(u)$.

$$\begin{array}{l}
\mathsf{strong\_hom}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
\quad \forall w. \\
\qquad w \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow \\
\qquad f\ w \in \mathfrak{M}_2.\mathsf{frame.world} \land \\
\qquad (\forall p.\ w \in \mathfrak{M}_1.\mathsf{valt}\ p \iff f\ w \in \mathfrak{M}_2.\mathsf{valt}\ p) \land \\
\qquad \forall u. \\
\qquad\quad u \in \mathfrak{M}_1.\mathsf{frame.world} \Rightarrow \\
\qquad\quad (\mathfrak{M}_1.\mathsf{frame.rel}\ w\ u \iff \mathfrak{M}_2.\mathsf{frame.rel}\ (f\ w)\ (f\ u))
\end{array}$$

A strong homomorphism is called an embedding if it is injective.

$$\begin{array}{l}
\mathsf{embedding}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
\quad \mathsf{strong\_hom}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \land \mathsf{INJ}\ f\ \mathfrak{M}_1.\mathsf{frame.world}\ \mathfrak{M}_2.\mathsf{frame.world}
\end{array}$$

And a strong morphism which is bijective is called an isomorphism.

$$\begin{array}{l}
\mathsf{iso}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \overset{\mathsf{def}}{=} \\
\quad \mathsf{strong\_hom}\ f\ \mathfrak{M}_1\ \mathfrak{M}_2 \land \mathsf{BIJ}\ f\ \mathfrak{M}_1.\mathsf{frame.world}\ \mathfrak{M}_2.\mathsf{frame.world}
\end{array}$$

For descibing the concept "equivalence of modalities", we introduce a bunch of definitions about τ-theory. The τ-theory of a point in a model is the set of formulas satisfied at this point. And the τ-theory of a model is the set of the the formula that is satisfied at any point in the model.

$$\mathsf{tau\_theory}\ \mathfrak{M}\ w \overset{\mathsf{def}}{=} \{\ \phi \mid \mathsf{satis}\ \mathfrak{M}\ w\ \phi\ \}$$

$$\mathsf{tau\_theory\_model}\ \mathfrak{M} \overset{\mathsf{def}}{=} \{\ \phi \mid \forall w.\ w \in \mathfrak{M}.\mathsf{frame.world} \Rightarrow \mathsf{satis}\ \mathfrak{M}\ w\ \phi\ \}$$

The above definitions enables us to define modal equivalence on two levels: for points or for models: Two points from two models are called modally equivalent if their τ-theory of points are the same.

$$\mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w' \overset{\mathsf{def}}{=} \mathsf{tau\_theory}\ \mathfrak{M}\ w = \mathsf{tau\_theory}\ \mathfrak{M}'\ w'$$

Similarly, two models are called modally equivalent if their τ-theory of models are the same.

$$\mathsf{modal\_eq\_model}\ \mathfrak{M}\ \mathfrak{M}' \overset{\mathsf{def}}{=} \mathsf{tau\_theory\_model}\ \mathfrak{M} = \mathsf{tau\_theory\_model}\ \mathfrak{M}'$$

Here comes our invarience theorem for strong homomorphisms

$$\begin{array}{l}
\vdash \mathsf{strong\_hom}\ f\ \mathfrak{M}\ \mathfrak{M}' \land f\ w = w' \land w \in \mathfrak{M}.\mathsf{frame.world} \land \\
\quad \mathsf{SURJ}\ f\ \mathfrak{M}.\mathsf{frame.world}\ \mathfrak{M}'.\mathsf{frame.world} \Rightarrow \\
\quad \mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'
\end{array}$$

which says if we have a surjective from a model $\mathfrak{M}$ to a model $\mathfrak{M}'$, then if the point w in $\mathfrak{M}$ is sent to w' under f, then w and w' are modally equivalent, means that they have the same τ-theory.

And the invarience theorem for the isomorphisms

$$\vdash \mathsf{iso}\ f\ \mathfrak{M}\ \mathfrak{M}' \Rightarrow \mathsf{modal\_eq\_model}\ \mathfrak{M}\ \mathfrak{M}'$$

which says if two models are isomorphic, then they are modally equivalent. It is directly by the invarience theorem for strong homomorphisms.

We prove the first one by firstly proving a equivalent statement without mentioning τ-theory:

$$\begin{array}{l}
\vdash \mathsf{strong\_hom}\ f\ \mathfrak{M}\ \mathfrak{M}' \land f\ w = w' \land w \in \mathfrak{M}.\mathsf{frame.world} \land \\
\quad \mathsf{SURJ}\ f\ \mathfrak{M}.\mathsf{frame.world}\ \mathfrak{M}'.\mathsf{frame.world} \Rightarrow \\
\quad (\mathsf{satis}\ \mathfrak{M}\ w\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ w'\ \phi)
\end{array}$$

Then we quote this lemma and expand the definition of modal equivalence and τ-theory to finish the proof.

We also have the notion of bounded morphism. It is just a homomorphism plus a backward condition, (use the notations as the definition of all other morphisms)which is "If $R'f(w)v'$, then there exists a v in $\mathfrak{M}$ such that Rwv and $f(v) = v'$".

$$
\begin{array}{l}
\mathsf{bounded\_mor}\ f\ \mathfrak{M}\ \mathfrak{M}' \ \stackrel{\mathsf{def}}{=} \\
\quad \forall\, w. \\
\quad\quad w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow \\
\quad\quad f\ w\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \land \\
\quad\quad (\forall\, a.\ \mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{VAR}\ a)\ \iff\ \mathsf{satis}\ \mathfrak{M}'\ (f\ w)\ (\mathsf{VAR}\ a))\ \land \\
\quad\quad (\forall\, v. \\
\quad\quad\quad v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \land\ \mathfrak{M}.\mathsf{frame.rel}\ w\ v\ \Rightarrow \\
\quad\quad\quad \mathfrak{M}'.\mathsf{frame.rel}\ (f\ w)\ (f\ v))\ \land \\
\quad\quad \forall\, v'. \\
\quad\quad\quad v'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \land\ \mathfrak{M}'.\mathsf{frame.rel}\ (f\ w)\ v'\ \Rightarrow \\
\quad\quad\quad \exists\, v.\ v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \land\ \mathfrak{M}.\mathsf{frame.rel}\ w\ v\ \land\ f\ v\ =\ v'
\end{array}
$$

If we have a surjective bounded morphism from M to M', then M' is called the bounded morphic image of M under f. That is:

$$
\begin{array}{l}
\mathsf{bounded\_mor\_image}\ f\ \mathfrak{M}\ \mathfrak{M}' \ \stackrel{\mathsf{def}}{=} \\
\quad \mathsf{bounded\_mor}\ f\ \mathfrak{M}\ \mathfrak{M}'\ \land\ \mathsf{SURJ}\ f\ \mathfrak{M}.\mathsf{frame.world}\ \mathfrak{M}'.\mathsf{frame.world}
\end{array}
$$

The invarience theorem for bounded morphism is:

$$
\begin{array}{l}
\vdash \mathsf{bounded\_mor}\ f\ \mathfrak{M}\ \mathfrak{M}'\ \land\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow \\
\quad (\mathsf{satis}\ \mathfrak{M}\ w\ \phi\ \iff\ \mathsf{satis}\ \mathfrak{M}'\ (f\ w)\ \phi)
\end{array}
$$

It states that if f is a bounded morphism from $\mathfrak{M}$ to $\mathfrak{M}'$, then for any world w in $\mathfrak{M}$, w and $f(w)$ are modally equivalent. Same as the invarience theorem for the previous constructions, it is proved by structual induction on formulas.

Now we give a definition of a important class of models, the tree-like models. Here we define a tree as a type of frame, so a tree-like model is just a model defined upon a frame which is a tree.

We say a frame S is a tree with the root r if the r is a point in the frame, satisfying:

(1) For any world in S, the root r is linked to it via the transitive closure of the relation defined on S.
(2) There is no point in S that is linked to the root via the relation on S.
(3) For any non-root point in S, there exists a unique point that is linked to it via the relation on S.

$$
\begin{array}{l}
\mathsf{tree}\ S\ r \ \stackrel{\mathsf{def}}{=} \\
\quad r\ \in\ S.\mathsf{world}\ \land \\
\quad (\forall\, t.\ t\ \in\ S.\mathsf{world}\ \Rightarrow\ (\mathsf{RESTRICT}\ S.\mathsf{rel}\ S.\mathsf{world})^*\ r\ t)\ \land \\
\quad (\forall\, r_0.\ r_0\ \in\ S.\mathsf{world}\ \Rightarrow\ \neg S.\mathsf{rel}\ r_0\ r)\ \land \\
\quad \forall\, t.\ t\ \in\ S.\mathsf{world}\ \land\ t\ \neq\ r\ \Rightarrow\ \exists!t_0.\ t_0\ \in\ S.\mathsf{world}\ \land\ S.\mathsf{rel}\ t_0\ t
\end{array}
$$

This definition is slightly different from the definition of the tree in the textbook, which does not say $(\forall r_0 \in S.world)(\neg S.rel\ r_0 r)$, but states that a tree has no loop as $(\forall t)\neg S^+\ t\ t$. To unify these two versions of definitions, we need to prove a tree has no loop as the following lemma:

$$
\vdash \mathsf{tree}\ s\ r\ \Rightarrow\ \forall\, t_0\ t.\ (\mathsf{RESTRICT}\ s.\mathsf{rel}\ s.\mathsf{world})^+\ t_0\ t\ \Rightarrow\ t_0\ \neq\ t
$$

This is proved by using `ho_match_mp_tac` with the induction principle `RTC_STRONG_INDUCT_RIGHT1`:

$$
\begin{array}{l}
\vdash (\forall\, x.\ P\ x\ x)\ \land\ (\forall\, x\ y\ z.\ P\ x\ y\ \land\ R^*\ x\ y\ \land\ R\ y\ z\ \Rightarrow\ P\ x\ z)\ \Rightarrow \\
\quad \forall\, x\ y.\ R^*\ x\ y\ \Rightarrow\ P\ x\ y
\end{array}
$$

Quite related, we have a definition of rooted model. A rooted model is a submodel of a model generated by a single point. For a rooted model, it would definitly be a submodel for some model, and the world set of a rooted model is exactly all the worlds that a world links to via the transitive closure of the relation of the model. $\mathsf{rooted\_model}\ \mathfrak{M}\ x\ \mathfrak{M}'$ reads "$\mathfrak{M}$ is a rooted model with root x, and it is a submodel of $\mathfrak{M}'$".

```
rooted_model $\mathfrak{M}$ $x$ $\mathfrak{M}'$ $\overset{\text{def}}{=}$
  $x \in \mathfrak{M}'$.frame.world $\wedge$
  ($\forall a$.
      $a \in \mathfrak{M}$.frame.world $\iff$
        $a \in \mathfrak{M}'$.frame.world $\wedge$
          (RESTRICT $\mathfrak{M}'$.frame.rel $\mathfrak{M}'$.frame.world)$^*$ $x$ $a$) $\wedge$
  ($\forall n_1\ n_2$.
      $n_1 \in \mathfrak{M}$.frame.world $\wedge$ $n_2 \in \mathfrak{M}$.frame.world $\Rightarrow$
      ($\mathfrak{M}$.frame.rel $n_1\ n_2$ $\iff$
        RESTRICT $\mathfrak{M}'$.frame.rel $\mathfrak{M}'$.frame.world $n_1\ n_2$)) $\wedge$
  $\forall v\ n$. $\mathfrak{M}$.valt $v\ n$ $\iff$ $\mathfrak{M}'$.valt $v\ n$
```

There is a significant result pointing out the importance of tree-like models and the connection between tree-like models, rooted models and bounded morphisms:

It says for any rooted model $\mathfrak{M}$, there is a tree-like model T such that $\mathfrak{M}$ is a bounded morphic image of T under some bounded morphism. So together with the invarience result for bounded morphisms, any satisfiable τ-formula is satisfiable in a tree-like model.

So to prove it, we need to specify such a bounded morphism and a tree-like model. For a rooted model $\mathfrak{M}$ whose root it x, the corresponding tree-like model is:

```
bounded_preimage_rooted $\mathfrak{M}$ $x$ $\overset{\text{def}}{=}$
  <|frame :=
    <|world :=
       { $l$ |
        HD $l = x$ $\wedge$ LENGTH $l > 0$ $\wedge$
        $\forall m$.
           $m <$ LENGTH $l - 1$ $\Rightarrow$
           RESTRICT $\mathfrak{M}$.frame.rel $\mathfrak{M}$.frame.world (EL $m$ $l$)
             (EL $(m + 1)$ $l$) } ;
     rel :=
       ($\lambda\, l_1\ l_2$.
           LENGTH $l_1 + 1 =$ LENGTH $l_2$ $\wedge$
           RESTRICT $\mathfrak{M}$.frame.rel $\mathfrak{M}$.frame.world (LAST $l_1$)
             (LAST $l_2$) $\wedge$
           $\forall m$. $m <$ LENGTH $l_1$ $\Rightarrow$ EL $m\ l_1 =$ EL $m\ l_2$)|>;
  valt := ($\lambda\, v\ n$. $\mathfrak{M}$.valt $v$ (LAST $n$))|>
```

That is, the worlds in this model are lists, encoded as natural numbers using `numlistTheory`. The definition actually says a list $(u_0, u_1, ..., u_n)$ is in the world of this model iff the list is nonempty, begining with the root x, and recording a path in M from the root x to u_n. A list l_1 is related to a list $l2$ iff $l2$ is one-element longer then l_1, if we put l_1 and l_2 in parallel and compare their members, they are all the same until the l_1 is end, and the last element of l_1 is linked to the last element of l_2 via the relation in $\mathfrak{M}$. And the evaluation map is defined by a list $l_1 \in V_{rooted}(p)$ iff the last element of $l_1 \in V(p)$ where V is the evaluation map of $\mathfrak{M}$. The model $\mathfrak{M}$ is then the bounded morphic image of it under the bounded morphism defined by taking the last element of the list, this is the function "nlast".

So we have two things to prove. One thing is that the model defined above is indeed a tree:

Beyond the works about numlistTheory(Here I have omitted all the works about the `numlistTheory`, otherwise the description of this part would be even longer then the already existing article.), The trickest part is to prove the unique existence of the predecessor of a non-root element. For a non-root element, its predecessor is the list obtained by discarding the last element of the list. discarding the last element of a numlist is defined as a function nfront, and the following lemmas proves that for a list t, nfront t is indeed a predecessor of t by proving that it is both in the world of our defined model, and it is linked to t by the relation on the model.

For proving that the predecessor is unique, note that an numlist is uniquely defined by the members in it. so once we conclude that if a list is a predecessor of a given list t, then each member of it is fixed, then we are done. And the information that "given a non-root element, each member of its predecessor is fixed" can be extracted from the third clauses of the definition of relation in `bounded_preimage_rooted_def`.

Having proved the first half. The other thing is that for the rooted model M with root x it is indeed the bounded morphic image of the model we construct above under the bounded morphism "nlast".

For the proof of this, rewriting with the definition of bounded morphic image gives four subgoals. Among them, a non-trivial one is the fact that if n is a list in the world set of the model we have defined, then $\mathsf{nlast}\ n$ is in the world set of M. To deal with it, note that $\mathsf{nlast}\ n\ =\ \mathsf{nel}\ (\mathsf{LENGTH}\ (\mathsf{listOfN}\ n)\ -\ 1)\ n$, then we can prove a lemma that for a numlist in our constructed model, every member of the list is in the world of $\mathfrak{M}$:

to proves the desired fact.

Another non-trivial part is the surjectiveness of the map nlast, this is proved as a lemma:

by induction on the reflexive transitive closure relation. To actually apply the above lemma into our proof. We also need to prove

which tells us for any element in a rooted model, it is linked to the root by the reflextive transitive closure. So we can get the anticedent of our lemma holds, which allows us to apply it.

With the two above theorems proved, the main theorem can be solved easily by applying `qexists_tac` and `metis_tac` with the above two lemmas.

6.2.2 2.2 Bisimulations

This section talks about bisimulations, which is defined as a relation between elements in the world sets of two models satisfying some conditions:

$$
\begin{array}{l}
\mathsf{bisim}\ Z\ \mathfrak{M}\ \mathfrak{M}'\ \stackrel{\text{def}}{=}\\
\quad\forall\, w\ w'.\\
\qquad w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ w'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ Z\ w\ w'\ \Rightarrow\\
\qquad (\forall\, a.\ \mathsf{satis}\ \mathfrak{M}\ w\ (\mathsf{VAR}\ a)\ \iff\ \mathsf{satis}\ \mathfrak{M}'\ w'\ (\mathsf{VAR}\ a))\ \wedge\\
\qquad (\forall\, v.\\
\qquad\quad v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ \mathfrak{M}.\mathsf{frame.rel}\ w\ v\ \Rightarrow\\
\qquad\quad \exists\, v'.\ v'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ Z\ v\ v'\ \wedge\ \mathfrak{M}'.\mathsf{frame.rel}\ w'\ v')\ \wedge\\
\qquad \forall\, v'.\\
\qquad\quad v'\ \in\ \mathfrak{M}'.\mathsf{frame.world}\ \wedge\ \mathfrak{M}'.\mathsf{frame.rel}\ w'\ v'\ \Rightarrow\\
\qquad\quad \exists\, v.\ v\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \wedge\ Z\ v\ v'\ \wedge\ \mathfrak{M}.\mathsf{frame.rel}\ w\ v
\end{array}
$$

As we can read from the code, for a relation Z between models M and M' to be a bisimulation, the worlds which are related must satisfies the same propositional letters, and the forward consition:

If wZw' and Rwv, then there exists a world v' in M' such that vZv' and $R'w'v'$.

And the backward condition:

If wZw' and $R'w'v'$, then there exists a world v in M such that vZv' and Rwv.

If two points in two models are related by a bisimulation, then we call the two points are bisimilar:

We can also say that two models are bisimilar if there is a bisimulation between them:

In fact, there are many situations where we can obtain a bisimulation:

(1) If $\mathfrak{M}\cong\mathfrak{M}'$, then $\mathfrak{M}$ is bisimilar to $\mathfrak{M}'$. (The bisimulation is given by relating a world to its image under the isomorphism.)

$$\vdash \mathsf{iso}\ f\ \mathfrak{M}\ \mathfrak{M}'\ \Rightarrow\ \mathsf{bisim\_model}\ \mathfrak{M}\ \mathfrak{M}'$$

(2) For a disjoint union $\mathsf{DU}\ (f, \mathit{dom})$ of models, any model $f\ i$ where i is in the dom is bisimilar to the unioned model. (The bisimulation is given by relating a world to its copy in the disjoint union.)

$$
\begin{array}{l}
\vdash i\ \in\ \mathit{dom}\ \wedge\ w\ \in\ (f\ i).\mathsf{frame.world}\ \Rightarrow\\
\quad \mathsf{bisim\_world}\ (f\ i)\ (\mathsf{DU}\ (f, \mathit{dom}))\ w\ (i, w)
\end{array}
$$

(3) If $\mathfrak{M}$ is a generated submodel of $\mathfrak{M}'$, then $\mathfrak{M}$ is bisimilar to $\mathfrak{M}'$. (The bisimulation is given by relating a world to itself.)

$$
\begin{array}{l}
\vdash \mathsf{GENSUBMODEL}\ \mathfrak{M}\ \mathfrak{M}'\ \Rightarrow\\
\quad \forall\, w.\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow\ \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w
\end{array}
$$

(4) If $\mathfrak{M}'$ is a bounded morphic image of $\mathfrak{M}$ under the bounded morphism f, then $\mathfrak{M}$ is bisimilar to $\mathfrak{M}'$. (The bisimulation is given by relating to a world to its image under the bounded morphism.)

$$
\begin{array}{l}
\vdash \mathsf{bounded\_mor\_image}\ f\ \mathfrak{M}\ \mathfrak{M}'\ \Rightarrow\\
\quad \forall\, w.\ w\ \in\ \mathfrak{M}.\mathsf{frame.world}\ \Rightarrow\ \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ (f\ w)
\end{array}
$$

Similar to the last section, there is also an invarience theorem for bisimulations. It says that if $\mathfrak{M}$ and $\mathfrak{M}'$ are models, then if two worlds w, w' in the two models are linked by a bisimulation, then they are modally equivalent.

$$\vdash \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w' \Rightarrow \mathsf{modal\_eq}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'$$

A natural question is that whether the converse of the above theorem holds. That is, if two worlds in two models are modally equivalent, whether they are bisimilar or not. In general, it does not. But the converse of the above theorem does hold for a special kind of models, called image-finite models. This kind of model is defined as the model has a special property that each point in the model is only related to a finite number of worlds, as defined below:

$$
\begin{array}{l}
\mathsf{image\_finite}\ \mathfrak{M} \overset{\mathsf{def}}{=} \\
\quad \forall x. \\
\qquad x \in \mathfrak{M}.\mathsf{frame.world} \Rightarrow \\
\qquad \mathsf{FINITE}\ \{\ y \mid y \in \mathfrak{M}.\mathsf{frame.world} \land \mathfrak{M}.\mathsf{frame.rel}\ x\ y\ \}
\end{array}
$$

The invarience theorem of this kind of model is called Hennessy-Milner Theorem:

It says that for worlds in image-finite models, modal equivalence is equivalent to being linked by a bisimulation. Half of this theorem is just a special case of the invarience theorem of bisimulations, the part that requires a proof is:

$$
\begin{array}{l}
\vdash \mathsf{image\_finite}\ \mathfrak{M} \land \mathsf{image\_finite}\ \mathfrak{M}' \land w \in \mathfrak{M}.\mathsf{frame.world} \land \\
\quad w' \in \mathfrak{M}'.\mathsf{frame.world} \Rightarrow \\
\quad (\forall \phi.\ \mathsf{satis}\ \mathfrak{M}\ w\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ w'\ \phi) \Rightarrow \\
\quad \mathsf{bisim\_world}\ \mathfrak{M}\ \mathfrak{M}'\ w\ w'
\end{array}
$$

We will prove that for two image-finite models, if two worlds are modally equivalent, then the relation Z that wZw' iff w and w' are modally equivalent is a bisimulation.

If we use `qexists_tac` to specify the bisimulation $\lambda\, n_1\ n_2.\ \forall \phi.\ \mathsf{satis}\ \mathfrak{M}\ n_1\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ n_2\ \phi$ and expand the definition of bisimulation of proving it, the HOL would ask us to prove two subgoals: both the forward condition and the backward consition. But the bisimulation we use here is actually special: it is a equivalence relation which is symmetric, so the forward and backward condition makes no difference, means that it suffices to prove only one of them. To convince the HOL that we just want prove only one of them, we prove this lemma:

The $P\ \mathfrak{M}\ N$ here is the $\mathsf{image\_finite}\ \mathfrak{M} \land \mathsf{image\_finite}\ \mathfrak{M}'$ in the theorem, and $Z\ \mathfrak{M}\ N\ u\ v$ corresponds to $\forall \phi.\ \mathsf{satis}\ \mathfrak{M}'\ n_2\ \phi \iff \mathsf{satis}\ \mathfrak{M}\ n_1\ \phi$. So once we prove this theorem, we can match it to the original statement with `ho_match_mp_tac` to give:

$$
\begin{array}{l}
(\forall \mathfrak{M}'\ \mathfrak{M}. \\
\quad \mathsf{image\_finite}\ \mathfrak{M}' \land \mathsf{image\_finite}\ \mathfrak{M} \Rightarrow \\
\quad \mathsf{image\_finite}\ \mathfrak{M} \land \mathsf{image\_finite}\ \mathfrak{M}') \land \\
(\forall \mathfrak{M}'\ \mathfrak{M}\ n_2\ n_1. \\
\quad (\forall \phi.\ \mathsf{satis}\ \mathfrak{M}'\ n_2\ \phi \iff \mathsf{satis}\ \mathfrak{M}\ n_1\ \phi) \Rightarrow \\
\quad \forall \phi.\ \mathsf{satis}\ \mathfrak{M}\ n_1\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ n_2\ \phi) \land \\
\forall \mathfrak{M}\ \mathfrak{M}'. \\
\quad \mathsf{image\_finite}\ \mathfrak{M} \land \mathsf{image\_finite}\ \mathfrak{M}' \Rightarrow \\
\quad \forall n_1\ n_2. \\
\qquad n_1 \in \mathfrak{M}.\mathsf{frame.world} \land n_2 \in \mathfrak{M}'.\mathsf{frame.world} \land \\
\qquad (\forall \phi.\ \mathsf{satis}\ \mathfrak{M}\ n_1\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ n_2\ \phi) \Rightarrow \\
\qquad (\forall a.\ \mathsf{satis}\ \mathfrak{M}\ n_1\ (\mathsf{VAR}\ a) \iff \mathsf{satis}\ \mathfrak{M}'\ n_2\ (\mathsf{VAR}\ a)) \land \\
\qquad \forall n_1'. \\
\qquad\quad n_1' \in \mathfrak{M}.\mathsf{frame.world} \land \mathfrak{M}.\mathsf{frame.rel}\ n_1\ n_1' \Rightarrow \\
\qquad\quad \exists n_2'. \\
\qquad\qquad n_2' \in \mathfrak{M}'.\mathsf{frame.world} \land \\
\qquad\qquad (\forall \phi.\ \mathsf{satis}\ \mathfrak{M}\ n_1'\ \phi \iff \mathsf{satis}\ \mathfrak{M}'\ n_2'\ \phi) \land \\
\qquad\qquad \mathfrak{M}'.\mathsf{frame.rel}\ n_2\ n_2'
\end{array}
$$

After `rpt strip_tac`, we can see many of the subgoals are trivial, and the only remaining task if prove one of the forward/backward condtion. So we can save labour on repeating the same proof.

The thing remains to prove to is that if n_1 in $\mathfrak{M}$,n_2 in $\mathfrak{M}'$ are modally equivalent, then for all n_1' in $\mathfrak{M}$ with Rn_1n_1', there exists an $n_2' \in M'$ such that $R'n_2n_2'$ and n_1' and n_2' are modally equivalent. It is proved by assuming that there exists a point in n_1' such that Rn_1n_1', and for all n_2' in $\mathfrak{M}'$ such that $R'n_2n_2'$, n_1' is not modally equivalent to n_2', then we will reach a contradiction by proving then n_1 and n_2 would not be modally equivalent. To conclude this, we need to find out a formula that is satisfied at n_1 but not n_2. It turns out that the formula we need is one with a diamond in the front. So it suffices to prove that there is a formula that is satisfied at n_1', but is not satisfied at any the world that is related to n_2. It turns out that we can find a formula which is a big conjunction that satisfies this condition. But the problem is: We have not got a definition of big conjunction by hand. To tackle this problem, we choose to do not use `qexists_tac` to specify an explict formula, but to prove its existence as a lemma.

We are given that $\mathfrak{M}$,$\mathfrak{M}'$ are both image finite, so the set of the worlds related to n_2 in $\mathfrak{M}'$ is finite. Also the set of such worlds is non-empty, since otherwise the formula $\Box\bot$ is vacuously true at n_2 but is false at n_1.

So once this lemma is proved, we can apply it to our case where s is the set of worlds that are related to n_2, v is the n_1', and then the "psi" in the conclusion gives the desire formula which gives the contradiction. Then by the argument above, we have proved the main theorem.