Lyndon Interpolation for Modal Logic via Type Elimination Sequences

Balder ten Cate[1]

[a]ILLC, University of Amsterdam, Postbus 94242, Amsterdam, 1090 GE, , The Netherlands

Abstract

This note describes a method for constructing Lyndon interpolants based on quasi-models and type elimination sequences. The same method was employed in [2] (using mosaics) to compute optimal-size Lyndon interpolants for formulas in the guarded-fragment and the guarded-negation fragment. This note serves to showcase the method in a simpler setting, namely that of the basic modal language. To provide context, I also briefly survey some other general approaches that have been used to prove interpolation for modal logic in the past. We finish with a list of questions.

Keywords: Modal Logic, Craig Interpolation, Lyndon Interpolation, Mosaics, Type Models

1. Introduction: interpolation for modal logic and its many proofs

A logic has the Craig interpolation property if, whenever an implication $\phi \to \psi$ is valid, there is a formula θ (which we will call an "interpolant") such that $\phi \to \theta$ and $\theta \to \psi$ are valid, and such that all the non-logical symbols occurring in θ occur both in ϕ and in ψ. Craig [6] proved that first-order logic has this property. This result was later refined by Lyndon [11], who proved that every valid implication has an interpolant, such that every non-logical symbol occurring positively in the interpolant occurs positively in both the antecedent and the consequent, and likewise for negative occurrences. Several further strengthenings of these interpolation theorems were later obtained for first-order logic, such as the relativized Lyndon interpolation theorem in [14] and the access interpolation theorem in [3].

Modal logic has Craig interpolation, and indeed, Lyndon interpolation, as well [8]. In fact, modal logic enjoys a strong form of interpolation that fails for first-order logic, called *uniform interpolation*, whereby the interpolant can be constructed in a uniform way, such that it does not depend on the consequent, but only on the antecedent and shared language (i.e., the set of non-logical symbols shared by the antecedent and consequent). There are many ways to prove interpolation theorems for modal logic. We briefly describe a few approaches, in order to provide some context. For simplicity, we focus on Craig interpolation here.

1.1. Model theoretic methods

Here, one proves interpolation by contrapositive: if there is no Craig interpolant for $\phi \to \psi$ then $\phi \to \psi$ is not valid. The main idea here is that, when $\phi \to \psi$ does not have a Craig interpolant, we can construct (using compactness and ω-saturated extensions) a pair of structures, one satisfying ϕ and the other satisfying $\neg\psi$, such that the two structures are bisimilar with respect to the common language. These two structures can then be combined with each other by taking a *bisimulation product* to obtain a model for $\phi \land \neg\psi$ [13]. This approach generalizes to the modal logic of any elementary frame class closed under bisimulation products and generated subframes. A self-contained exposition can be found in [5], which also gives a precise syntactic characterization of first-order frame conditions that are closed under bisimulation products and generated subframes.

Gabbay [8]'s original proof of Craig interpolation for modal logic falls in this category of model-theoretic proofs, even though the amalgamation operation was not yet as clearly spelled out yet. Amalgamation operations can also be studied in an algebraic setting, cf. the long line of work by Maksimova [12]. See also [9].

1.2. Proof theoretic methods

An interpolant for $\phi \to \psi$ can also be constructed from a proof of the validity of $\phi \to \psi$, in a suitable proof system such as a tableau calculus. See for instance [17]. Proof theoretic techniques have also been used to prove uniform interpolation results (cf. for instance [15] for intuitionistic logic, and [1] for the modal μ-calculus).

[1]Thanks to Johan van Benthem for interesting comments of a draft of this note, and suggesting many of the questions listed at the end of the document. Thanks to anonymous reviewers for identifying a mistake in a previous version.
A preliminary version of this paper was published as a technical report [18]. Since then, the proof techniques described here were put to use in [10] to prove new results, prompting this more polished version.
Research supported by the European Union's Horizon 2020 research and innovation programme (MSCA-101031081).

1.3. Syntactic (automata theoretic) methods

A modal formula is in ∇-normal form if it is a (possibly empty) disjunction of formulas generated by the following grammar:

$$\phi \ ::= \ (\neg)p_1\wedge\cdots\wedge(\neg)p_n \ \mid \ (\neg)p_1\wedge\cdots\wedge(\neg)p_n\wedge\nabla(\phi_1,\ldots,\phi_m)$$

where $m,n\geq 0$ and where $p_1,\ldots,p_n$ are distinct proposition letters. The semantics of the ∇ operator as follows: $M,w\models\nabla(\phi_1,\ldots,\phi_m)$ holds if *each ϕ_i is true in some successor and each successor satisfies some ϕ_i*. Note that ∇ can be expressed in terms of $\Diamond$ and $\Box$, and similarly vice versa. Every modal formula is equivalent to one in ∇-normal form (by means of simple syntactic rewrite rules).

If $\phi\rightarrow\psi$ is valid, we can construct a Craig interpolant by rewriting ϕ to ∇-normal form and then simply dropping all (positive and negative) occurrences of proposition letters from it that are not in the common language. (Lyndon interpolants can be obtained similarly by dropping positive (negative) occurrences of proposition letters that do not occur positively (negatively) in ψ). This surprisingly simple construction, in fact, yields something stronger than Craig interpolation, namely uniform interpolation. Furthermore, since normalization to ∇-normal form can be performed in exponential time, this yields Craig interpolants of single exponential size (even without requiring a DAG-representation for the interpolant). Details can be found for instance in [19] (the results there are phrased in terms of the description logic $\mathcal{ALC}$ which is simply a notational variant of modal logic with multiple modalities).

Although it may not be apparent from this purely syntactic presentation, there is a strong automata-theoretic underpinning to this approach. Modal formulas in ∇-normal form can be naturally viewed as tree automata, and the operation of "dropping all occurrences of proposition letters" corresponds to *projection* from the view point of automata theory (which can also be characterized in terms of *bisimulation quantifiers* [7]).

The same construction also works for the modal μ-calculus (although the proof of the ∇-normal form theorem is more involved here, cf. Yde Venema's *lectures on the modal μ-calculus*). On the other hand, since this technique yields uniform interpolation, it does not allow us to prove Craig interpolation for logics that lack uniform interpolation (eg., the extension of modal logic with the global modality).

2. Quasi-models and type elimination sequences

The interpolation method we will now describe is one that was successfully applied to the guarded-fragment and the guarded-negation fragment in [2]. The construction used there (using mosaics) is quite involved, and the aim, here, is to illustrate the method in the simpler setting of the basic modal language.

In this section, we review quasi-models (a.k.a. type models) and type elimination sequences as a technique for deciding validity (or, equivalently, satisfiability). This technique can be traced back to [16] where it was used to show that the satisfiability problem for Propositional Dynamic Logic (PDL) is in ExpTime. In the next section, we then show how to extract interpolants from type elimination sequences.

A quasi-model is nothing more than a consistent collection of "types", where a type is simply a subset $X\subseteq Y$, where Y is some finite set of relevant formulas (for instance, Y may be the set of all subformulas of a given formula whose validity we are trying to determine).

Our presentation is specifically tailored to the case where the input formula (for which we want to test validity) is an implication between a pair of modal formulas. This will facilitate the interpolant-construction later. In addition, some of our definitions (e.g., the definition of SUBF below) are tailored in anticipation of the specific use case of constructing Lyndon interpolants.

Definition 1 (NNF, SUBF, and LITERALS).

- *A formula is in* negation normal form (NNF) *if it is generated by*

  $$\chi ::= p \mid \neg p \mid \top \mid \bot \mid \chi_1\wedge\chi_2 \mid \chi_1\vee\chi_2 \mid \Diamond\chi \mid \Box\chi$$

  We denote by $\textsc{nnf}(\chi)$ *the result of bringing a formula* χ *into NNF.*

- *If χ is a formula in NNF, we denote by* $\textsc{subf}(\chi)$ *the following set:*

  $$\begin{array}{lcl}\textsc{subf}(\alpha) & = & \{\alpha\} \text{ for all } \alpha \text{ of the form} \\ & & \quad p,\ \neg p,\ \top \text{ or } \bot \\ \textsc{subf}(\chi_1\wedge\chi_2) & = & \textsc{subf}(\chi_1)\cup\textsc{subf}(\chi_2)\cup\{\chi_1\wedge\chi_2\} \\ \textsc{subf}(\chi_1\vee\chi_2) & = & \textsc{subf}(\chi_1)\cup\textsc{subf}(\chi_2)\cup\{\chi_1\vee\chi_2\} \\ \textsc{subf}(\Diamond\chi) & = & \textsc{subf}(\chi)\cup\{\Diamond\chi\} \\ \textsc{subf}(\Box\chi) & = & \textsc{subf}(\chi)\cup\{\Box\chi\}\end{array}$$

  Note that $\textsc{subf}(\neg p)$ *does* not *include* p.

- *If χ is a formula in NNF, we denote by* $\textsc{literals}(\chi)$ *the set of all formulas in* $\textsc{subf}(\chi)$ *of the form p or $\neg p$. For example,* $\textsc{literals}(\Box p\wedge\Diamond\neg q)=\{p,\neg q\}$.

Fix a modal implication $\phi\rightarrow\psi$. The next few definitions are all relative to the given choice of the modal formulas ϕ and ψ.

Definition 2 (Types; Overlap-Consistency).

1. *By a* locally-consistent *subset of* $\textsc{subf}(\chi)$ *we mean a subset* $X\subseteq\textsc{subf}(\chi)$ *such that (i) whenever* $\chi_1\wedge\chi_2\in X$*, then* $\chi_1,\chi_2\in X$*; (ii) whenever* $\chi_1\vee\chi_2\in X$*, then at least one of* χ_1,χ_2 *belongs to X; (iii)* $\bot\notin X$*; and (iv) it is not the case that* $p,\neg p\in X$ *for some proposition letter p.*
2. *An* L-type *is a locally-consistent subset of* $\textsc{subf}(\textsc{nnf}(\phi))$*, and an* R-type *is a locally-consistent subset of* $\textsc{subf}(\textsc{nnf}(\neg\psi))$*. A combined* type *(or just* type *for short) is a pair* $\tau=(\tau_L,\tau_R)$ *where* τ_L *is an L-type and* τ_R *is an R-type.*

3. *A type $\tau = (\tau_L, \tau_R)$ is* overlap-consistent *if there does not exist a proposition letter p such that $p \in \tau_L$ and $\neg p \in \tau_R$ or vice versa.*

Definition 3 (Viable successor). *Let τ, τ' be D-types, for $D \in \{L, R\}$. We say that τ' is a viable successor for τ (notation: $\tau \Rightarrow \tau'$) if for every formula of the form $\Box\chi$ belonging to τ, χ belongs to τ'. This definition extends naturally to combined types: we say that $\tau \Rightarrow \tau'$ holds if $\tau_L \Rightarrow \tau'_L$ and $\tau_R \Rightarrow \tau'_R$.*

Definition 4 (Quasi-model). *A* quasi-model *is a set X of (combined) types, such that*

1. *Every $\tau \in X$ is overlap-consistent.*
2. *For all $\tau \in X$ and $\Diamond\chi \in \tau_D$ (with $D \in \{L, R\}$), there is $\tau' \in X$ such that $\tau \Rightarrow \tau'$ and $\chi \in \tau'_D$.*

This completes the definition of quasi-models. The next theorem states that quasi-models can be used to decide validity of $\phi \to \psi$.

Theorem 1. (Soundness and completeness of quasi-models) *The modal formula $\phi \wedge \neg\psi$ is satisfiable if and only if there is a quasi-model X with a type $\tau = (\tau_L, \tau_R) \in X$ such that $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$.*

Proof. (sketch) In one direction, suppose $M, w \models \phi \wedge \neg\psi$. For every world v of M, we define its type $\tau(v)$ to be $(\tau_L(v), \tau_R(v))$, where $\tau_L(v)$ is the set of formulas from $\text{SUBF}(\text{NNF}(\phi))$ true at v, and $\tau_R(v)$ is the set of formulas from $\text{SUBF}(\text{NNF}(\neg\psi))$ true at v. Let $X = \{\tau(v) \mid v \text{ a world of } M\}$. Then it is easy to show that X is a quasi-model.

Conversely, suppose X is a quasi-model with a type $\tau = (\tau_L, \tau_R) \in X$ such that $\phi \in \tau_L$ and $\dot\neg\psi \in \tau_R$. We construct a model as follows: for each $\tau' \in X$ we create a world $w_{\tau'}$. The accessibility relation connects a pair of worlds $(w_{\tau'}, w_{\tau''})$ whenever $\tau' \Rightarrow \tau''$. Finally, a proposition letter p is set to true at $w_{\tau'}$ whenever $p \in \tau'_L \cup \tau'_R$. It is easy to show (by induction) that the resulting model M_X satisfies the following truth lemma:

$$M_X, w_{\tau'} \models \chi \text{ for all } \chi \in \tau'_L \cup \tau'_R$$

In particular, it follows that $M_X, w_\tau \models \phi \wedge \neg\psi$. □

Clearly, a quasi-model is a finite object. More precisely, since a type is a polynomial-sized object, a quasi-model is an object of singly exponential size (as a function of the size of ϕ and ψ). Therefore, Theorem 1 provides us with a decision procedure for testing satisfiability of $\phi \wedge \neg\psi$. The immediate upper bound we get from this theorem is NExpTime (by non-deterministically guessing the quasi-model). In fact, we can do a little better.

Definition 5 (Type elimination sequence). *A* type elimination sequence *is a sequence $X_0, \ldots, X_n$ where*

1. *X_0 is the set of all types,*
2. *each X_{i+1} is obtained from X_i by removing a type $\tau \in X_i$ that fails to satisfy condition 1 or 2 from the definition of quasi-models.*
3. *X_n is a quasi-model.*

It is easy to see that a type elimination sequence always exists (note that X_0 is finite and that X_n may be empty).

Theorem 2.

1. *All type elimination sequences (for the given formulas ϕ, ψ) end in the same quasi-model X_n, which can equivalently be characterized as the maximal quasi-model, and as the union of all quasi-models.*
2. *The modal formula $\phi \wedge \neg\psi$ is satisfiable if and only if X_n contains a type τ with $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$.*

Proof. (sketch)
For item 1, it suffices to observe that, if $X_0, \ldots, X_n$ is a type elimination sequence and X is any quasi-model, then, $X \subseteq X_i$ for all $i \le n$. Indeed, this can be shown by a straightforward induction on i.

For item 2, if $\phi \wedge \neg\psi$ is satisfiable, then let X be the quasi-model given by Theorem 1. By item 1, we have that $X \subseteq X_n$, and hence, X_n contains a type τ with $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$. If, on the other hand, $\phi \wedge \neg\psi$ is *not* satisfiable, then it follows from Theorem 1 that there is no quasi-model containing a type τ with $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$. In particular, X_n does not contain such a type. □

This puts the complexity in ExpTime: it suffices to construct an arbitrary type elimination sequence and inspect the final type set of the sequence. Note that the length of the sequence is single exponential because one type gets eliminated each step. Note that this complexity upper bound is still not optimal, because the satisfiability problem for modal logic is in PSpace. However, the quasi-model method is quite generic and can be adapted to various extensions of modal logic that are ExpTime-complete, such as with the global modality (cf. Section 4).

2.1. Excursion: greatest fixed points vs least fixed points

Type elimination sequences can naturally be viewed as greatest fixed point computations. To make this precise, let T be the set of all types, and let $F : 2^T \to 2^T$ be the (monotone) function given by

$$F(X) = \{\tau \in T \mid \begin{array}{l} \tau \text{ is overlap-consistent and for all} \\ \Diamond\chi \in \tau_D \text{ (with } D \in \{L, R\}), \\ \text{there is a } \tau' \in X \\ \text{such that } \tau \Rightarrow \tau' \text{ and } \chi \in \tau'_D\} \end{array}$$

It then follows immediately from Definition 4 and Theorem 2(1) that:

Theorem 3.

1. *Quasi-models are the post-fixed points of F. That is, a set $X \subseteq T$ is a quasi-model if and only if $X \subseteq F(X)$.*
2. *For any type-elimination sequence $X_0, \dots, X_n$, we have that X_n is equal to the greatest fixed point of F.*

This raises the question whether *least fixed points* also have a natural role to play in the story. For this, we need to define two further notions: *well-founded models* and *type introduction sequences* (the natural duals of type elimination sequences). We say that a Kripke model M is *well-founded* if every world w can be assigned an ordinal $\delta(w)$, in such a way that, whenever a world w has a successor v, $\delta(v) < \delta(w)$. If, in addition, all ordinals in question can be picked to be finite, then we say that M has *finite depth.*

Definition 6 (Type introduction sequence). *A* type introduction sequence *is a sequence $X_0, \dots, X_n$, where*

1. $X_0 = \emptyset$
2. *Each X_{i+1} extends X_i with a type in $F(X_i) \setminus X_i$.*
3. $F(X_n) = X_n$.

Note that, since the empty set is a quasi-model, and the set T of all types is finite, a type introduction sequence always exists.

Theorem 4.

1. *All type introduction sequences (for the given formulas ϕ, ψ), end in the same quasi-model X_n, which is the least fixed point of F.*
2. *The modal formula $\phi \wedge \neg\psi$ is satisfiable in a well-founded model if and only if X_n contains a type τ with $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$.*

Proof. (sketch) For item 1, it suffices to observe that if $X_0, \dots, X_n$ is a type introduction sequence, then (i) X_n is a fixed-point of F, and (ii) for all fixed points X of F, we have that $X_i \subseteq X$ for all $i \leq n$. Claim (i) holds by definition; claim (ii) can be shown by induction on n, using the monotonicity of F.

For item 2, first suppose that $M, w \models \phi \wedge \neg\psi$ and M is well-founded. By an ordinal induction on $\delta(v)$, we can show that, for all worlds v of M, the type $\tau(v)$ of v belongs to every fixed point of F. In particular, $\tau(w)$ belongs to X_n. Note that $\text{NNF}(\phi) \in \tau_L(w)$ and $\text{NNF}(\neg\psi) \in \tau_R(w)$. Conversely, it is easy to show by induction on i that, for each type τ in X_i, there is a well-founded model containing a world satisfying all the formulas in τ. (Indeed, since there are only finitely many types to consider, a finite-depth model even suffices.) In particular, if $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$ for some $\tau \in X_n$, then $\phi \wedge \neg\psi$ is satisfiable in a well-founded model. □

Coincidentally, for formulas of the basic modal language, in the absence of frame conditions, it is known that satisfiability coincides with satisfiability in well-founded models (and with satisfiability in finite-depth models). Consequently, if we are interested in testing the satisfiability of $\phi \wedge \neg\psi$, we can freely choose to use type elimination sequences or type introduction sequences. However, in general (e.g., for modal logics of transitive frames) this equivalence no longer holds. Indeed, the modal logic of all transitive frames is K4 while the modal logic of the well-founded transitive frames is GL.

3. Constructing interpolants from type elimination sequences

Recall that a *Lyndon interpolant* for a valid implication $\phi \to \psi$ is a formula θ such that $\phi \to \theta$ and $\theta \to \psi$ are valid, and such that every proposition letter occurring positively (negatively) in θ occurs positively (negatively) in both ϕ and ψ. In particular, a Lyndon interpolant is a Craig interpolant (but not vice versa).

Fix modal formulas ϕ, ψ such that $\phi \to \psi$ is valid. By Theorem 2 there is a type elimination sequence $X_0, \dots, X_n$ such that X_n does not contain any type τ with $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$. For the remainder of this section, we can fix such a sequence.

The core result is:

Theorem 5. *If a type $\tau = (\tau_L, \tau_R)$ gets eliminated in the sequence, then there is a modal formula θ_τ, such that*

- $\models (\bigwedge \tau_L) \to \theta_\tau$ *and*
- $\models \theta_\tau \to \neg(\bigwedge \tau_R)$ *and*
- *Every proposition letter occurring positively (negatively) in θ_τ occurs positively (negatively) in both ϕ and ψ.*

In the remainder of this section, we will show, first of all, that, from this theorem we get, as a corollary, the Lyndon interpolation theorem; secondly, we will prove this theorem itself.

Corollary 1 (Lyndon interpolation). *If $\models \phi \to \psi$, then*

$$\theta = \bigvee_{\substack{\tau_L \text{ an L-type} \\ \text{with } \phi \in \tau_L}} \; \bigwedge_{\substack{\tau_R \text{ an R-type} \\ \text{with } \dot\neg\psi \in \tau_R}} \theta_{(\tau_L, \tau_R)}$$

is a Lyndon interpolant for $\phi \to \psi$.

Proof. If $\phi \wedge \neg\psi$ is unsatisfiable, then the type elimination sequence ends in a quasi-model that does not contain a type τ with $\phi \in \tau_L$ and $\dot\neg\psi \in \tau_R$. Therefore, every such type gets eliminated. This shows that the above formula θ is indeed well defined.

To see that $\models \phi \to \theta$, let $M, w \models \phi$. Let τ_L be the L-type of w, i.e., the set of subformulas of ϕ satisfied at w in M. By construction, $M, w \models \bigwedge \tau_L$. It follows from Theorem 5 that, for all R-types τ_R containing $\dot\neg\psi$, because

(τ_L, τ_R) got eliminated, $M, w \models \theta_{(\tau_L,\tau_R)}$. It follows that $M, w \models \theta$.

To see that $\models \theta \to \psi$, by contraposition, let $M, w \not\models \psi$. Let τ_R be the R-type of w, i.e., the set of subformulas of ψ satisfied at w in M. By construction, $M, w \models \bigwedge \tau_R$. It follows by Theorem 5 that $M, w \models \theta_{(\tau_L,\tau_R)}$ for all τ_L containing ϕ. Therefore, $M, w \not\models \theta$. $\square$

It remains only to prove Theorem 5.

Proof. (of Theorem 5)

The proof is by induction on the stage at which the type gets eliminated. If a type τ is eliminated, this is for one of two reasons:

1. τ is not overlap-consistent. If $p \in \tau_L$ and $\neg p \in \tau_R$, we pick $\theta_\tau = p$. Note that $p \in \textsc{literals}(\phi)$ and $\neg p \in \textsc{literals}(\textsc{nnf}(\neg\psi))$. Since polarity of proposition letter occurrences is preserved when bringing formulas into normal form, and polarity of proposition letter occurrences is inverted when negating a formula, it follows that p occurs positively in both ϕ and ψ. The case where $\neg p \in \tau_L$ and $p \in \tau_R$ is similar, except that we choose $\theta_\tau = \neg p$. By analogous reasoning, in this case, p must occur negatively in both ϕ and ψ.

2. For some $D \in \{L, R\}$, τ_D contains a formula of the form $\Diamond\chi$, and every type $\tau' = (\tau'_L, \tau'_R)$ satisfying $\tau \Rightarrow \tau'$ and $\chi \in \tau'_D$ has already been eliminated earlier on in the elimination sequence.

If $D = L$, then take

$$\theta_\tau \quad = \quad \Diamond \bigvee_{\substack{\tau'_L \text{ an L-type with} \\ \chi \in \tau'_L \text{ and } \tau_L \Rightarrow \tau'_L}} \bigwedge_{\substack{\tau'_R \text{ an R-type with} \\ \tau_R \Rightarrow \tau'_R}} \theta_{(\tau'_L,\tau'_R)}$$

We need to show that (i) $\models \bigwedge \tau_L \to \theta_\tau$ and (ii) $\models \theta_\tau \to \neg(\bigwedge \tau_R)$.

For (i), let $M, w \models \bigwedge \tau_L$. Then $M, w \models \Diamond\chi$. Let v be a witnessing successor, i.e., $(w, v) \in R^M$ and $M, v \models \chi$. Let τ'_L be the L-type of v. By induction hypothesis, we have that $\bigwedge \tau'_L \models \theta_{(\tau'_L,\tau'_R)}$ for all τ'_R such that (τ'_L, τ'_R) got eliminated. It follows, by the induction hypothesis, that $M, v \models \bigwedge_{\tau'_R \text{ an R-type with } \tau_R \Rightarrow \tau'_R} \theta_{(\tau'_L,\tau'_R)}$. Hence, $M, w \models \theta_\tau$.

For (ii), by contraposition, let $M, w \models \bigwedge \tau_R$. We need to show that $M, w \not\models \theta_\tau$. Let v be any successor of w, and let τ'_R be the R-type of v. Observe that $\tau_R \Rightarrow \tau'_R$. It follows by the induction hypothesis that $M, v \not\models \theta_{(\tau'_L,\tau'_R)}$, for all τ'_L such that (τ'_L, τ'_R) got eliminated. Hence, $M, w \not\models \theta_\tau$.

This concludes the argument for $D = L$. If $D = R$, the argument is analogous, except that we reason dually, and we now take

$$\theta_\tau \quad = \quad \Box \bigvee_{\substack{\tau'_L \text{ an L-type with} \\ \tau_L \Rightarrow \tau'_L}} \bigwedge_{\substack{\tau'_R \text{ an R-type with} \\ \chi \in \tau'_R \text{ and } \tau_R \Rightarrow \tau'_R}} \theta_{(\tau'_L,\tau'_R)}$$

We omit the details, which are straightforward. $\square$

3.1. Complexity and size bounds

By a careful inspection of the above procedure it can be shown that this yields a method for constructing interpolants in exponential time, provided that the interpolants are represented succinctly using a DAG-style representation of formulas. If formulas are required to be represented as trees (without subformula sharing) then there is an extra exponential blowup and, in this case, the procedure runs in doubly-exponential time and yields doubly-exponentially large interpolants. (By comparison, the interpolation method using ∇-normal form described in Section 1 also runs in exponential time and produces single-exponential length interpolants, even without requiring a DAG representation.)

4. Discussion

As discussed before, there are many different methods for constructing interpolants for modal logic. Each has its own advantages. The method described here generalizes to guarded fragments [2] and yields an effective procedure for interpolant construction that is optimal in terms of formula length and running time, in that context.

4.1. Extensions of the basic modal language

It seems likely that the above method can be adapted to extensions of the basic modal language (besides guarded fragments). In particular, it is natural to ask this question for extensions of the basic modal language that admit filtration.

As an example, consider the extension of the basic modal language with the global modality. In this case, (i) the definition of $\Rightarrow$ can be adapted so that both types agree on all global statements they make, and (ii) the definition of a quasi-model can be suitably extended with a clause for the global modality. Note that unlike in the basic modal case, a quasi-model can now no longer be associated with a single Kripke model. Rather, the types in a quasi-model can be naturally partitioned in terms of whether they agree on all global statements they make; each part of this partition naturally gives rise to a different Kripke model. Nevertheless, Theorem 1 and Theorem 2 still holds true. The interpolant construction naturally extends as well. We omit the details here.

Also, unlike the model-theoretic approach we briefly described in Section 1, the techniques presented here are not inherently restricted to fragments of first-order logic. They may also work for some non-first-order-definable extensions of the modal language.

> *Can the type-elimination-sequence method be used to show interpolation for graded modal logic? And, the extension of the basic modal language with a $\Diamond^\infty$ modality ("there are infinitely many successors satisfying. . . ")? And, monadic first-order logic with the infinity quantifier [4]?*

4.2. Modal logics of restricted frame classes

Besides language extensions, it is also natural to ask whether the interpolation method via quasi-models can be extended to modal logics of restricted classes. A natural example is K4, the modal logic of the class of transitive frames.

> *Can the type-elimination-sequence method be used to show interpolation for K4? And (in the light of the discussion in Section 2.1), can interpolation be shown for the modal logic GL of well-founded transitive frames, by means of an induction on type-introduction-sequences?*

Note that the logics in question are known to have Craig interpolation. These questions are therefore more intended to generate insight in the generality of the method.

> *Can one prove a general interpolation result (and ExpTime complexity upper bound) for a larger class of modal logics of frame classes that admit filtration?*

4.3. Type elimination sequences vs proofs

By Theorem 2, a modal implication $\phi \to \psi$ is valid if and only if there is a type elimination sequence leading to a quasi-model that does not contain any type τ with $\text{NNF}(\phi) \in \tau_L$ and $\text{NNF}(\neg\psi) \in \tau_R$. At the same time, by the soundness and completeness of the axiomatic system **K**, this holds if and only if there is an axiomatic proof for $\phi \to \psi$. And similarly for other proof systems such as sequent calculi.

> *Can a proof (in some modal proof system) be extracted from a type elimination sequence?*

4.4. Other refinements of Craig Interpolation

As discussed in Section 1, modal logic admits a strong form of interpolation called *uniform interpolation*, where the interpolant does not depend on the consequent if the implication (only on its signature). This property is quite brittle. For instance, uniform interpolation is known to fail for the extension of modal logic with the global modality, as well as for the modal logic K4 of transitive frames (although it holds for the modal μ-calculus).

> *Can the type-elimination-sequence based method be adapted to prove uniform interpolation for modal logic?*

It is not clear that this is easy. Perhaps inspiration could be taken from proofs of uniform interpolations based on sequent calculi [15, 1].

The same can be asked for several other, more modest, refinements of the Craig interpolation property, for instance in a multi-modal setting where we can interpolate over modalities, as well as possibly over $\Diamond$'s and $\Box$'s separately. It is more likely that the method can be adapted to prove such interpolation theorems.

References

[1] Bahareh Afshari, Graham E. Leigh, and Guillermo Menéndez Turata. Uniform interpolation from cyclic proofs: The case of modal mu-calculus. In Anupam Das and Sara Negri, editors, *Automated Reasoning with Analytic Tableaux and Related Methods*, pages 335–353, Cham, 2021. Springer International Publishing.

[2] Michael Benedikt, Balder Ten Cate, and Michael Vanden Boom. Effective interpolation and preservation in guarded logics. *ACM Trans. Comput. Logic*, 17(2), dec 2015.

[3] Michael Benedikt, Balder ten Cate, and Efthymia Tsamoura. Generating plans from proofs. *ACM Trans. Database Syst.*, 40(4):22:1–22:45, 2016.

[4] Facundo Carreiro, Alessandro Facchini, Yde Venema, and Fabio Zanasi. Model theory of monadic predicate logic with the infinity quantifier. *Archive for Mathematical Logic*, 2021.

[5] Balder ten Cate. *Model Theory for Extended Modal Languages*. PhD thesis, University of Amsterdam, 2005.

[6] William Craig. Three uses of the herbrand-gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22(3):269–285, 1957.

[7] Giovanna D'Agostino. Uniform interpolation, bisimulation quantifiers, and fixed points. In Balder D. ten Cate and Henk W. Zeevat, editors, *Logic, Language, and Computation*, pages 96–116, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[8] Dov M. Gabbay. Craig's interpolation theorem for modal logics. In Wilfrid Hodges, editor, *Conference in Mathematical Logic — London '70*, pages 111–127, Berlin, Heidelberg, 1972. Springer Berlin Heidelberg.

[9] Eva Hoogland. *Definability and interpolation: model-theoretic investigations*. PhD thesis, University of Amsterdam, 2000.

[10] Jean Christoph Jung, Andrea Mazzullo, and Frank Wolter. More on interpolants and explicit definitions for description logics with nominals and/or role inclusions. In *Proceedings of DL2022*, 2022.

[11] Roger C. Lyndon. An interpolation theorem in the predicate calculus. *Pacific Journal of Mathematics*, 9(1):129 – 142, 1959.

[12] Larisa Maksimova. Amalgamation and interpolation in normal modal logics. *Studia Logica: An International Journal for Symbolic Logic*, 50(3/4):457–471, 1991.

[13] Maarten Marx and Yde Venema. *Multi-dimensional modal logic*, volume 4 of *Applied logic series*. Kluwer, 1997.

[14] Martin Otto. An interpolation theorem. *The Bulletin of Symbolic Logic*, 6(4):447–462, 2000.

[15] Andrew M. Pitts. On an interpretation of second order quantification in first order intuitionistic propositional logic. *The Journal of Symbolic Logic*, 57(1):33–52, 1992.

[16] V. R. Pratt. Models of program logics. In *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)*, pages 115–122, 1979.

[17] Wolfgang Rautenberg. Modal tableau calculi and interpolation. *J. Philos. Log.*, 12(4):403–423, 1983.

[18] Balder ten Cate. Lyndon interpolation for modal logic via type elimination sequences. Technical Report X-2022-01, ILLC, 2022.

[19] Balder ten Cate, Willem Conradie, Maarten Marx, and Yde Venema. Definitorially complete description logics. In Patrick Doherty, John Mylopoulos, and Christopher A. Welty, editors, *Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006*, pages 79–89. AAAI Press, 2006.