Exponential Speedup in $\mathcal{UL}$ Subsumption Checking Relative to General TBoxes for the Constructive Semantics

Michael Mendler and Stephan Scheele

Informatics Theory Group
University of Bamberg, Germany
{michael.mendler,stephan.scheele}@uni-bamberg.de

Abstract. The complexity of the subsumption problem in description logics can vary widely with the choice of the syntactic fragment and the semantic interpretation. In this paper we show that the constructive semantics of concept descriptions, which includes the classical descriptive semantics as a special case, offers exponential speed-up in the existential-disjunctive fragment $\mathcal{UL}$ of $\mathcal{ALC}$.

1 Introduction

One of the key reasoning tasks in the application of DL formalisms is the subsumption problem relative to a set of terminological axioms, called a TBox. The complexity of the decision procedure depends on (i) the language fragment used for concept descriptions, (ii) the structure of the TBox, i.e., whether it is acyclic, cyclic or even permits general concept inclusions (GCIs) and finally (iii) the semantic interpretation of the logical operators. E.g., to deal with cyclic TBoxes the standard extensional (so-called 'descriptive') semantics which considers all fixed points of TBoxes has been made more intensional by restricting to greatest fixed points and least fixed points [17]. Many result have been obtained in exploring the options in all these dimensions, discovering complexities all the way from PTime to NExpTime completeness.

Regarding the full language $\mathcal{ALC}$ [5] it has been shown for the descriptive semantics that subsumption without TBox as well as with acyclic TBoxes is PSpace complete, while it is ExpTime complete for general TBoxes (see [5]). Since many applications do not need all the operators of $\mathcal{ALC}$, restricted languages have been considered for which lower complexity levels can be derived, even down to PTime. There seem to be two main strands among these so-called *sub-boolean* formalisms, the $\mathcal{FL}$-type languages starting from the fragment $\mathcal{FL}_0$ consisting of $\{\forall, \sqcap\}$ and the $\mathcal{EL}$-type languages which build on the operator set $\{\exists, \sqcap\}$. Among these, the latter tend to be much better behaved (e.g., more efficient and less sensitive to extensions) than the former. For instance, it is known that subsumption checking in $\mathcal{FL}_0$ is PTime for empty TBoxes [14], coNP complete for acyclic TBoxes [17], while for cyclic TBoxes it is PSpace complete under the descriptive semantics [13] as well as greatest and least fixed point semantics [1] and it becomes ExpTime complete for general TBoxes (descriptive semantics) [12, 4]. On the other side, subsumption in $\mathcal{EL}$ remains in PTime for cyclic and acyclic TBoxes under all three semantics [3] and even for general descriptive TBoxes [8, 12]. There are several extensions to the $\mathcal{EL}$ language one can make without losing tractability, such as adding $\bot$, $\top$, nominals, concrete domains and more [6]. The addition of disjunction $\sqcup$, which brings back Boolean expressiveness, results in coNP hardness [8] for empty TBoxes, PSpace for acyclic and ExpTime completeness for cyclic and general TBoxes [4]. See [11] for an overview on the $\mathcal{EL}$ family.

Presumably it is fair to assume that applications of logic-based knowledge engineering, in particular those involving mass data or real-time interaction, will tend to sacrifice expressiveness for higher performance of reasoning services. The search for tractable fragments below $\mathcal{ALC}$ is not only expedient from a practical perspective, it still appears there is quite some playground left to be explored. On the one hand, the existing fragments $\mathcal{FL}$ and $\mathcal{EL}$ represent only two of the four corners of the Aristotelian classification square: $\mathcal{FL}_0$ with $\{\forall, \sqcap\}$ permits us to make general statements of the form *"all S are P"* while $\mathcal{EL}$ with $\{\exists, \sqcap\}$ corresponds[1] to *"some S are P"*. Less attention has been given to the so-called contraries *"no S is P"* and *"not all S are P"* which correspond to fragments $\{\forall, \sqcup\}$ and $\{\exists, \sqcup\}$. Are these also useful as a basis in specific applications and if so what are their complexities? On the other hand, there is the semantics issue: The standard descriptive semantics which follows a classical Tarskian model-theory is not the only reasonable way of interpreting concept description languages. There is the Scottian least fixed or greatest fixed point view for cyclic TBoxes introduced by Nebel [18] or the automata-theoretic interpretation of Baader [1]. Also, the concept algebras introduced by Dionne et al. [19, 10] provide alternative ways of giving intensional semantics to concept descriptions and TBoxes. Depending on application and language fragment some of these may be more appropriate than the classical descriptive semantics. The semantics issue, too, leaves room for further systematical investigations.

The aim of this paper is to show that the choice of semantics can play a decisive role for the complexity of the subsumption problem in certain fragments of $\mathcal{ALC}$. Specifically, we will show that for the $\{\exists, \sqcup\}$ fragment, here named $\mathcal{UL}$, there is an exponential gap between the classical descriptive semantics which gives ExpTime-complete subsumption checking while for the *constructive semantics* [16] the problem lies in PTime. Like Dionne et al.'s concept algebras, the constructive interpretation is intensional but is based on a modal extension of Heyting algebras which are more general and also cover full $\mathcal{ALC}$. They admit an elementary presentation using Kripke models which enrich the traditional descriptive interpretation naturally by a reflexive and transitive *refinement* relation for concept abstraction and refinement. The idea behind the constructive

[1] To see this consider conjunction $\sqcap$ as representing generic affirmative statements with $\top$ as nullary case and disjunction $\sqcup$ as a generic refutative statement with $\bot$ as the degenerated case. Of course, the refutation about $\sqcup$ consists in giving choices, thus avoiding commitment.

semantics [16] is that concepts must not be taken as absolute references for clearly delineated sets of object instances but as intensional descriptions that are subject to context and negotiation. This is important for situations in which knowledge is incomplete or evolving such as business auditing, where the data populating the concepts are produced in continuous streams which are potentially infinite. In this case role filling may involve actions that brings data into existence only at auditing time. There, the open world assumption of classical descriptions is too weak, since it is not dynamic, while the constructive semantics is robust. In [16] a number of examples are given to illustrate the issue. Showing that it turns the classically intractable problem for $\mathcal{UL}$ tractable further justifies the interest in the constructive semantics.

2 Constructive Semantics for $\mathcal{ALC}$

Concept descriptions are built from *role names* N_R and *concept names* N_C as follows

$$C, D ::= A \mid \top \mid \bot \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid C \sqsubseteq D \mid \exists R.C \mid \forall R.C,$$

where $A \in N_C$ and $R \in N_R$. This syntax is more general than standard $\mathcal{ALC}$ in that it includes subsumption $\sqsubseteq$ as a concept–forming operator. The fact that $\sqsubseteq$ can be nested will be irrelevant for the purposes of this paper. Still, we present the constructive semantics for the full system:

Definition 1 ([16]). *A* constructive interpretation *of $\mathcal{ALC}$ is a structure $\mathcal{I} = (\Delta^{\mathcal{I}}, \preceq^{\mathcal{I}}, \bot^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of*

- *a non-empty set $\Delta^{\mathcal{I}}$ of* entities*;*
- *a* refinement *pre-ordering $\preceq^{\mathcal{I}}$ on $\Delta^{\mathcal{I}}$ (reflexive and transitive);*
- *a subset $\bot^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ of* fallible *entities closed under refinement and role filling, i.e., $x \in \bot^{\mathcal{I}}$ and $x \preceq^{\mathcal{I}} y$ or $xR^{\mathcal{I}}y$ implies $y \in \bot^{\mathcal{I}}$ for all $R \in N_R$; also $x \in \bot^{\mathcal{I}}$ means that for all $R \in N_R$ there exists $y \in \Delta^{\mathcal{I}}$ such that $xR^{\mathcal{I}}y$;*
- *finally an interpretation function $\cdot^{\mathcal{I}}$ mapping each role name $R \in N_R$ to a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ and each concept name $A \in N_C$ to a set $\bot^{\mathcal{I}} \subseteq A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ which is closed under refinement, i.e., $x \in A^{\mathcal{I}}$ and $x \preceq^{\mathcal{I}} y$ implies $y \in A^{\mathcal{I}}$.*

$\mathcal{I}$ is lifted from constant $\bot$ and concept names A to arbitrary concepts as seen below, where $\Delta^{\mathcal{I}}_c =_{df} \Delta^{\mathcal{I}} \setminus \bot^{\mathcal{I}}$ is the set of non-fallible *elements in $\mathcal{I}$:*

$$\begin{aligned}
\top^{\mathcal{I}} &=_{df} \Delta^{\mathcal{I}} \\
(\neg C)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}_c.\ x \preceq^{\mathcal{I}} y \Rightarrow y \notin C^{\mathcal{I}}\} \\
(C \sqcap D)^{\mathcal{I}} &=_{df} C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} &=_{df} C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(C \sqsubseteq D)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}_c.\ (x \preceq^{\mathcal{I}} y \ \&\ y \in C^{\mathcal{I}}) \Rightarrow y \in D^{\mathcal{I}}\} \\
(\exists R.C)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}_c.\ x \preceq^{\mathcal{I}} y \Rightarrow \exists z \in \Delta^{\mathcal{I}}.\ (y,z) \in R^{\mathcal{I}} \ \&\ z \in C^{\mathcal{I}}\} \\
(\forall R.C)^{\mathcal{I}} &=_{df} \{x \mid \forall y \in \Delta^{\mathcal{I}}_c.\ x \preceq^{\mathcal{I}} y \Rightarrow \forall z \in \Delta^{\mathcal{I}}.\ (y,z) \in R^{\mathcal{I}} \Rightarrow z \in C^{\mathcal{I}}\}.
\end{aligned}$$

Entities in $\Delta^{\mathcal{I}}$ are partial descriptions representing incomplete information about individuals. Read $a \preceq b$ as saying that b *"is a refinement of"* a. The difference to the classical descriptive semantics is the refinement dimension $\preceq^{\mathcal{I}}$ and the universal quantification $\forall y \in \Delta_c^{\mathcal{I}}.\, x \preceq^{\mathcal{I}} y \Rightarrow \ldots$ in the clauses of Definition 1.

If $\preceq^{\mathcal{I}}$ trivialises to the identity relation, i.e., each entity refines only itself, and if $\bot^{\mathcal{I}} = \emptyset$, then the constructive semantics coincides with the classical descriptive semantics of $\mathcal{ALC}$. Therefore, the constructive semantics includes the classical one.

Let $c\mathcal{ALC}$ be $\mathcal{ALC}$ under the constructive interpretation. Then $c\mathcal{ALC}$ is related to the constructive modal logic CK (Constructive K) [20, 7, 15] as $\mathcal{ALC}$ is related to the classical modal system K [9]. In $c\mathcal{ALC}$ the classical principles of the *Excluded Middle* $C \sqcup \neg C \equiv \top$, *Double Negation* $\neg\neg C \equiv C$, the dualities $\exists R.C \equiv \neg\forall R.\neg C$, $\forall R.C \equiv \neg\exists R.\neg C$ and *Disjunctive Distribution* $\exists R.(C \sqcup D) \equiv \exists R.C \sqcup \exists R.D$ are no longer tautologies but non-trivial TBox statements (cf. [16]). In particular, consider Disjunctive Distribution: Imagine a dynamic data model so that role filling for R is a process which involves accessing a remote data base. Then role filling is a computation which interacts with the environment to access/generate the data. In this case the disjunctive decision $C \sqcup D$ can only be made once the (R-filler) context to access the data has been established. This means that the data model may well guarantee $\exists R.(C \sqcup D)$, yet it would be rather strong to assume it also satisfies $\exists R.C \sqcup \exists R.D$ where the decision is outside the scope of R and thus available before the R-filler is even requested.

Before we can embark on the main technical developments we need to introduce some pieces of notation. First, it will be convenient (to save indices) to use a semantic validity relation $\models$ as follows: Write $\mathcal{I}; x \models C$ to abbreviate $x \in C^{\mathcal{I}}$ in which case we say that entity x *satisfies* concept C in the interpretation $\mathcal{I}$. Further, $\mathcal{I}$ is a *model* of C, written $\mathcal{I} \models C$ iff $\forall x \in \Delta^{\mathcal{I}}.\mathcal{I}; x \models C$. All uses of $\models$ are extended to sets Φ of concepts, $\mathcal{I}; x \models \Phi$, $\mathcal{I} \models \Phi$ in the usual universal fashion. Let Θ be a set of TBox axioms, and Γ a set of concept descriptions. We write $\Theta; \Gamma \models C$ if for all models $\mathcal{I}$ of Θ it is the case that every entity $x \in \Delta^{\mathcal{I}}$ which satisfies all axioms in Γ must also satisfy concept C. Formally, $\forall \mathcal{I}.\forall x \in \Delta^{\mathcal{I}}.\, (\mathcal{I} \models \Theta \;\&\; \mathcal{I}; x \models \Gamma) \Rightarrow \mathcal{I}; x \models C$. The reasoning task which we are interested in here is subsumption in the presence of TBox axioms: The statement $\Theta; \{C\} \models D$ expresses that concept C is *subsumed* by concept D in the presence of terminology Θ, i.e., for all models $\mathcal{I} \models \Theta$ we have $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$; Unlike in the classical semantics of $\mathcal{ALC}$, we cannot reduce subsumption to non-satisfiability since $\Theta; \{C, \neg D\} \models \bot$ is not the same as $\Theta; \{C\} \models D$. For convenience, the latter will be written $\Theta; C \models D$ saving the braces. Note that subsumption is defined relative to all *constructive* models. If we wish to restrict to the *classical* models we write $\Theta; C \models_{cl} D$, meaning that for all classical models $\mathcal{I} \models \Theta$ we have $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

In the following we will consider TBoxes Θ which consist entirely of subsumptions $C \sqsubseteq D$ such that both C, D are $\sqsubseteq$-free. These are usually called *general* TBoxes. We will investigate subsumption checking relative to general TBoxes for the existential-disjunctive fragment of $\mathcal{ALC}$, called $\mathcal{UL}$. This is the class of concept descriptions formed as

$$C, D ::= \top \mid \bot \mid A \mid C \sqcup D \mid \exists R.C$$

with $A \in N_C$ and $R \in N_R$. Notice that in this fragment we can have unconditional negative and positive axioms C and $\neg C$ included in TBoxes, viz. as subsumptions $C \sqsubseteq \bot$ and $\top \sqsubseteq C$. The PTime result will hold for the *guarded* fragment of $\mathcal{UL}$. A concept description C is *(existentially) guarded* if it is not a disjunction, or equivalently, if all disjunctions appearing in C are guarded behind existential quantifiers. A general $\mathcal{UL}$ TBox Θ is called *(existentially) guarded* if in every $C \sqsubseteq D \in \Theta$ the conclusion D is guarded.

3 Guarded $\mathcal{UL}_0$ is ExpTime hard for Classical Descriptive Semantics

We reduce subsumption in the language $\mathcal{FL}_0$ consisting of operators $\{\forall, \sqcap\}$ to subsumption in $\mathcal{UL}$ under classical descriptive semantics. In fact, we will reduce subsumption in $\mathcal{FL}_0$ to the fragment $\mathcal{UL}_0$, which is $\mathcal{UL}$ without the constants $\top, \bot$, and relative to guarded $\mathcal{UL}_0$ TBoxes. Let Θ be a general TBox in $\mathcal{FL}_0$ and C, D two $\mathcal{FL}_0$ concept descriptions. For simplicity we assume that N_C contains exactly the concept names appearing in Θ or C, D. The first step in our reduction dualises the problem under the classical descriptive semantics, replacing $\sqcap \mapsto \sqcup$ and $\forall \mapsto \exists$ and swapping left and right-hand sides of subsumptions. Formally, define the dual $d(C)$ of a $\mathcal{FL}_0$ concept description recursively as follows: $d(A) =_{df} A$, $d(C \sqcap D) =_{df} d(C) \sqcup d(D)$, $d(\forall R.C) =_{df} \exists R.d(C)$. Further, the dual of the TBox is $d(\Theta) =_{df} \{d(D) \sqsubseteq d(C) \mid C \sqsubseteq D \in \Theta\}$. Then, one shows $\Theta; C \models_{cl} D \iff d(\Theta); d(D) \models_{cl} d(C)$. Obviously, the dualised $\mathcal{FL}_0$ concepts $d(D), d(C)$ are now $\mathcal{UL}_0$ expressions and $d(\Theta)$ is a general TBox of $\mathcal{UL}_0$. $d(\Theta)$ need not be existentially guarded, though. However, this can be fixed.We introduce a new role τ to guard the TBox inclusions making sure that the effect of τ can be compensated for by choosing the right classical interpretations. Given an $\mathcal{UL}_0$ TBox Θ we define its *expansion*

$$\begin{aligned} exp(\Theta) =_{df} & \{C \sqsubseteq \exists\tau.D \mid C \sqsubseteq D \in \Theta\} \\ & \cup \{ \exists R.X \sqsubseteq \exists\tau.\exists R.X,\ \exists\tau.\exists R.X \sqsubseteq \exists R.X \ \mid \exists R.X \in \mathrm{Sf}(\Theta)\} \\ & \cup \{A \sqsubseteq \exists\tau.A, \exists\tau.A \sqsubseteq A \mid A \in N_C\}, \end{aligned}$$

where $\mathrm{Sf}(\Theta)$ denotes the set of sub-concepts of Θ. Clearly, all subsumptions in $exp(\Theta)$ are guarded. Because of the additional axioms, the new role τ is semantically "silent" and does not add any information. Observing $\Theta; C \models_{cl} D$ *iff* $exp(\Theta); C \models_{cl} D$, classical subsumption checking for Θ in $\mathcal{FL}_0$ is now reduced to $exp(\Theta)$ in guarded $\mathcal{UL}_0$:

Lemma 1. *Let C, D be $\mathcal{FL}_0$ concept descriptions and Θ a general TBox of $\mathcal{FL}_0$. Then, $\Theta; C \models_{cl} D$ iff $exp(d(\Theta)); d(D) \models_{cl} d(C)$.*

Lemma 1 gives a linear-time reduction of the $\mathcal{FL}_0$ problem to that in $\mathcal{UL}_0$, where the simulating TBox $exp(d(\Theta))$ is existentially guarded. Since subsumption checking in $\mathcal{FL}_0$ for general TBoxes and classical descriptive semantics is ExpTime hard [12, 4] we have the following theorem:

Theorem 1. *$\mathcal{UL}_0$ subsumption checking under the classical semantics and relative to existentially guarded (general) TBoxes is* ExpTime hard.

It is important to realise that the reduction behind Thm. 1 depends crucially on the classical nature of the semantics, the DeMorgan dualities and the existential distribution $\models_{cl} \exists R.(C \sqcup D) \equiv (\exists R.C \sqcup \exists R.D)$. As has been pointed out by Hofmann [12] the presence and absence of such distributive laws can make the difference between ExpTime and PTime complexity. In ExpTime $\mathcal{FL}_0$ the operators $\{\sqcap, \forall\}$ distribute while in the PTime fragment $\{\sqcap, \exists\}$ [2, 12] they do not.

4 Guarded $\mathcal{UL}$ is in PTime for Constructive Descriptive Semantics

In the following we will show that subsumption checking in $\mathcal{UL}$ under the constructive semantics and relative to existentially guarded TBoxes is in PTime. We obtain a PTime decision procedure for $\mathcal{UL}$ by the calculus presented in Fig. 1. The calculus is formulated in the style of Gentzen with left introduction rules $\sqcup L$, $\sqsubseteq L$, $\exists LR$, $\bot L$ and right introduction rules $\sqcup R_1$, $\sqcup R_2$, $\exists LR$, $\top R$ for each logical connective of $\mathcal{UL}$. We will show that the calculus is sound and complete for $\mathcal{UL}$ under the constructive semantics.

Let $\mathcal{I}$ be a constructive interpretation and $a \in \Delta^{\mathcal{I}}$ an entity in $\mathcal{I}$. We say that the pair $(\mathcal{I}, a)$ *satisfies* a sequent $\Theta; C \vdash D$ if $\mathcal{I}$ is a model of Θ, i.e., $\mathcal{I} \models \Theta$, and $\mathcal{I}; a \models C$ as well as $\mathcal{I}; a \not\models D$. A sequent $\Theta; C \vdash D$ is *(constructively) satisfiable*, iff there exists $(\mathcal{I}, a)$ which satisfies the sequent. This is equivalent to non-subsumption $\Theta; C \not\models D$ (see page 4). Completeness of the calculus of Fig. 1 means that if no proof can be found for $\Theta; C \vdash D$ then the sequent is satisfiable. Soundness means that whenever $\Theta; C \vdash D$ is derivable then $\Theta; C \models D$, i.e. C is subsumed by D. While soundness is easy to establish completeness requires a number of auxiliary tools which we introduce next.

Definition 2. *A concept description P is called (constructively) prime in a TBox Θ if for all $D_1, D_2 \in Sf(\Theta) \cup Sf(P)$, whenever $\Theta; P \models D_1 \sqcup D_2$ then $\Theta; P \models D_1$ or $\Theta; P \models D_2$.*

The trouble for the classical semantics is that TBoxes have only very few if any prime concepts. Instead, the TBox must be saturated by adding and manipulating instances in terms of *sets* of concepts. Such sets of concepts are needed to pin down a hypothetical individual sufficiently for going through all the case analyses needed to decide a given subsumption in the context of Θ. On the other hand, constructive TBoxes have more prime elements, i.e., single concept

$$\frac{}{\Theta; C \vdash C}\ Ax \qquad \frac{}{\Theta; \bot \vdash C}\ \bot L \qquad \frac{}{\Theta; C \vdash \top}\ \top R$$

$$\frac{\Theta; E \vdash C}{\Theta; E \vdash C \sqcup D}\ \sqcup R_1 \qquad \frac{\Theta; E \vdash D}{\Theta; E \vdash C \sqcup D}\ \sqcup R_2 \qquad \frac{\Theta; C \vdash F \quad \Theta; D \vdash F}{\Theta; C \sqcup D \vdash F}\ \sqcup L$$

$$\frac{\Theta; E \vdash C \quad \Theta; D \vdash F}{\Theta, C \sqsubseteq D; E \vdash F}\ \sqsubseteq L \qquad \frac{\Theta; C \vdash D}{\Theta; \exists R.C \vdash \exists R.D}\ \exists LR$$

Fig. 1. Gentzen sequent rules for $\mathcal{UL}$.

descriptions which are able to capture a whole set of individuals completely. The striking feature of constructive $\mathcal{ALC}$ is that existential concepts are always constructively prime for the empty TBox.

Proposition 1. *Every existentially guarded $\mathcal{UL}$ concept description C is constructively prime for the empty TBox.*

Proof. By induction on the structure of C and the definition of $\models$ for the constructive semantics (see page 4). □

Proposition 1 is the reason why for existentially guarded $\mathcal{UL}$ the constructive calculus for $\mathcal{ALC}$ [16] can be restricted to the simple form of sequents $\Theta; C \vdash D$ manipulated in Fig. 1 without losing completeness. The following Props. 2 and 3 will be instrumental for proving this.

Proposition 2 (Cut Admissibility). *In the proof system of Fig. 1 the cut rule is admissible, i.e., if $\Theta; C \vdash D$ and $\Theta; D \vdash E$ then $\Theta; C \vdash E$.*

Proof. Given derivations for $\Theta; C \vdash D$ and for $\Theta; D \vdash E$ we show $\Theta; C \vdash E$ by induction on the derivations and the *cut formula* D. □

Proposition 3 (Provable Primes). *Let Θ be an existentially guarded TBox and P, C_1, C_2 concept descriptions. If P is existentially guarded then $\Theta; P \vdash C_1 \sqcup C_2$ implies $\Theta; P \vdash C_1$ or $\Theta; P \vdash C_2$.*

Proof. We proceed by induction on the derivation for $\Theta; P \vdash D_1 \sqcup D_2$. Since P is not a disjunction, the last rule in this derivation cannot be Ax, $\sqcup L$, $\exists LR$ or $\top R$. If the last rule is $\bot L$ or one of $\sqcup R_1$, $\sqcup R_2$ the claim follows immediately. The only interesting case is when the derivation ends in an application of $\sqsubseteq L$, i.e., it looks like

$$\frac{\dfrac{\vdots\ \pi_1}{\Theta; P \vdash E} \qquad \dfrac{\vdots\ \pi_2}{\Theta; F \vdash D_1 \sqcup D_2}}{\Theta, E \sqsubseteq F; P \vdash D_1 \sqcup D_2}\ \sqsubseteq L$$

By assumption on our TBox from which $E \sqsubseteq F$ is drawn, F is guarded. Hence we may invoke the induction hypothesis on the sub-derivation π_2. This gives a derivation π_2' for $\Theta; F \vdash D_i$ for some $i = 1, 2$, from which we get

$$\dfrac{\dfrac{\vdots\, \pi_1}{\Theta; P \vdash E} \quad \dfrac{\vdots\, \pi_2'}{\Theta; F \vdash D_i}}{\Theta, E \sqsubseteq F; P \vdash D_i} \sqsubseteq L$$

as desired. □

The next observation we need to make is that all concept descriptions in a $\mathcal{UL}$ TBox Θ can be decomposed and fully covered by the guarded elements of $\mathrm{Sf}(\Theta)$. Define a relation $C \trianglerighteq P$ iff P is a (maximal) guarded sub-expression of C. If C is guarded then $C \trianglerighteq C$. Otherwise, $C = C_1 \sqcup C_2$ is a disjunction and one of C_i must contain the guarded sub-expression P.

Lemma 2 (Coverings). *Let C be a concept description in $\mathcal{UL}$ and Θ a TBox. Then,*

(i) If $C \trianglerighteq P$ and $\Theta; C \vdash E$ then $\Theta; P \vdash E$
(ii) If $\Theta; P \vdash E$ for all $C \trianglerighteq P$ then $\Theta; C \vdash E$.

Proof. (i) The proof depends on cut admissibility and the fact that if $C \trianglerighteq P$ then $\Theta; P \vdash C$. This can be proved easily by induction on C. (ii) Again, this is by induction on the structure of C. For constants $\top$, $\bot$, concept names and existentials the statement is trivial, since for these $C \trianglerighteq C$. For disjunctions we notice that $C_1 \sqcup C_2 \trianglerighteq P$ iff $C_1 \trianglerighteq P$ or $C_2 \trianglerighteq P$. Hence, if $\Theta; P \vdash E$ for all $C_1 \sqcup C_2 \trianglerighteq P$ we have $\Theta; C_i \vdash E$ for both $i = 1, 2$ by induction hypothesis and thus $\Theta; C_1 \sqcup C_2 \vdash E$ by way of rule $\sqcup L$. □

From now on we will assume that the $\mathcal{UL}$ TBox Θ is existentially guarded. We build a constructive interpretation $\mathcal{I}$ based on guarded concept descriptions by letting $\Delta^{\mathcal{I}}$ consist of the set of pairs (C, Ψ_C) where C is guarded and Ψ_C is a finite map associating with each role $R \in N_R$ a set $\Psi_C(R) \subseteq \mathrm{Sf}(\Theta)$ of concept descriptions such that the formal *(intensional) consistency* requirement $\Theta; C \nvdash \Psi_C$ holds. This consistency condition is an abbreviation for $\Theta; C \nvdash \exists R. \bigsqcup_{X \in \Psi_C(R)} X$ for all $R \in N_R$ such that $\Psi_C(R) \neq \emptyset$. The guarded concept description C in a pair (C, Ψ_C) records the *extensional* information characterising the entity represented by (C, Ψ_C) while the second component Ψ_C encodes additional *intensional* information. Each entry $X \in \Psi_C(R)$ specifies the R-fillers in the sense that every entity consistent with the concepts in $\Psi_C(R)$ will be an R-filler of (C, Ψ_C). In our case we can further restrict the structure to $\Psi_C = \emptyset$ or $\Psi_C = [R \mapsto X]$, i.e. $\Psi_C(R) = \emptyset$ for all $R \in N_R$ or for exactly one $R \in N_R$ we have $\Psi_C(R) = \{X\}$, respectively. Note that the pair $(C, \emptyset)$ is always consistent regardless of C and that $(C, [R \mapsto X])$ is consistent iff $\Theta; C \nvdash \exists R.X$.

The canonical model $\mathcal{I}$ is given by an interpretation of concept names, filler and refinement relations $R^{\mathcal{I}}$ and $\preceq^{\mathcal{I}}$ on $\mathcal{I}$, respectively, as follows:

$$\begin{aligned}(C, \Psi_C) \in A^{\mathcal{I}} &\iff \Theta; C \vdash A \\ (C, \Psi_C) \in \bot^{\mathcal{I}} &\iff \Theta; C \vdash \bot \\ (C, \Psi_C)\ R^{\mathcal{I}}\ (D, \Psi_D) &\iff \forall X \in \Psi_C(R).\ \Theta; D \nvdash X \\ (C, \Psi_C)\ \preceq^{\mathcal{I}}\ (D, \Psi_D) &\iff \Theta; D \vdash C.\end{aligned}$$

Observe that the constructive semantics distinguishes between two kinds of consistencies, viz. *extensional consistency*, if $(C, \Psi_C) \notin \bot^{\mathcal{I}}$ and *intensional consistency*, if $\Theta; C \nvdash \Psi_C$. Our canonical $\mathcal{I}$ contains only intensionally consistent objects, though they may be extensionally inconsistent. In such an entity $(C, \Psi_C) \in \bot^{\mathcal{I}}$ we must necessarily have $\Psi_C(R) = \emptyset$ for all $R \in N_R$. For if $\Psi_C(R) = \{X\}$ for some R and X then $\Theta; \bot \vdash \exists R.X$ by rule $\bot L$ which by the assumption $\Theta; C \vdash \bot$ and cut admissibility would yield $\Theta; C \vdash \exists R.X$ in contradiction to intensional consistency of (C, Ψ_C). But $\Psi_C(R) = \emptyset$ means the entity (C, Ψ_C) does not exclude any R-fillers, i.e. $(C, \Psi_C)\ R^{\mathcal{I}}\ (D, \Psi_D)$ for every intensionally consistent pair (D, Ψ_D). Notice again that every pair $(C, \emptyset)$ is intensionally consistent and thus an object in $\mathcal{I}$.

Clearly, $\preceq^{\mathcal{I}}$ is reflexive because of rule Ax and transitive because of cut admissibility Prop. 2. It is not antisymmetric, though, due to the slack in the second component Ψ_D which has no influence on $\preceq^{\mathcal{I}}$. In fact, it contains cycles since, e.g., $(C, \Psi_C) \preceq^{\mathcal{I}} (C, \emptyset)$ and also $(C, \emptyset) \preceq^{\mathcal{I}} (C, \Psi_C)$ for all Ψ_C. Truth valuations $\bot^{\mathcal{I}}$ and $A^{\mathcal{I}}$ are closed under $\preceq^{\mathcal{I}}$ due to cut admissibility. Extensionally inconsistent objects $\bot^{\mathcal{I}}$ partake in all $\mathcal{UL}$ concept descriptions, i.e., for all $\mathcal{UL}$ concepts, $\bot^{\mathcal{I}} \subseteq C^{\mathcal{I}}$. Thus, $\mathcal{I}$ is a constructive model according to Def. 1[2].

Lemma 3. *Let Θ be an existentially guarded TBox and E an arbitrary $\mathcal{UL}$ concept description. Then, for every $(C, \Psi_C) \in \mathcal{I}$, we have $\mathcal{I}; (C, \Psi_C) \models E$ iff $\Theta; C \vdash E$.*

Proof. We proceed by induction on E. For concept names and $\bot$ the statement follows directly from the definition of $A^{\mathcal{I}}$. For $\top$ we need to use rule $\top R$. The interesting cases are existentials and disjunction:

Let $(C, \Psi_C) \in \mathcal{I}$ and suppose $\Theta; C \nvdash \exists R.E$, which in particular implies $\Theta; C \nvdash \bot$ by cut admissibility and $\bot L$. Then the pair $(C, [R \mapsto E])$ is intensionally and extensionally consistent. The latter follows from $\Theta; C \nvdash \bot$ and the former from $\Theta; C \nvdash \exists R.E$. Thus, $(C, [R \mapsto E])$ is an entity in $\mathcal{I}$ and also $(C, [R \mapsto E]) \preceq^{\mathcal{I}} (C, [R \mapsto E])$ with $(C, [R \mapsto E]) \in \Delta_c^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus \bot^{\mathcal{I}}$. Consider any R-filler $(C, [R \mapsto E])\ R^{\mathcal{I}}\ (D, \Psi_D)$ which must satisfy $\Theta; D \nvdash E$ by definition. Now we use the induction hypothesis to infer $\mathcal{I}; (D, \Psi_D) \not\models E$ which proves $\mathcal{I}; (C, \Psi_C) \not\models \exists R.E$ according to Def. 1, all in all. Vice versa, suppose that $\Theta; C \vdash \exists R.E$ and let $(C, \Psi_C) \preceq^{\mathcal{I}} (D, \Psi_D) \in \Delta_c^{\mathcal{I}}$ be given arbitrarily, which means $\Theta; D \vdash C$ and by

[2] To be fully compatible also with the constructive semantics of $\neg$, $\forall R$ we should also make sure that all R-fillers of objects in $\bot^{\mathcal{I}}$ are themselves extensionally inconsistent. This is not needed for the existential-disjunctive fragment at hand.

cut admissibility $\Theta; D \vdash \exists R.E$. We claim that there exists a guarded component P of E such that $(P, \emptyset)$ is an R-filler of (D, Ψ_D). If $\Psi_D(R) = \emptyset$ we can take any guarded $E \trianglerighteq P$ and have $(D, \Psi_D)\ R^{\mathcal{I}}\ (P, \emptyset)$ for trivial reasons. Otherwise, if $\Psi_D(R) = \{X\}$ we must find a guarded component P of E such that $\Theta; P \not\vdash X$. We proceed by contradiction. Suppose, $\Theta; P \vdash X$ for every P with $E \trianglerighteq P$. By Lem. 2 (ii) this implies $\Theta; E \vdash X$. From there, an application of $\exists LR$ gives $\Theta; \exists R.E \vdash \exists R.X$ and further cut admissibility $\Theta; D \vdash \exists R.X$ which would contradict consistency of (D, Ψ_D). Thus, as claimed, $(D, \Psi_D)\ R^{\mathcal{I}}\ (P, \emptyset)$ for some guarded component P of E. By Lem. 2 (i) we have $\Theta; P \vdash E$ from which the induction hypothesis tells us that $\mathcal{I}; (P, \emptyset) \models E$ which completes the proof that $\mathcal{I}; (C, \Psi_C) \models \exists R.E$ according to Def. 1.

Now assume $\mathcal{I}; (C, \Psi_C) \models E_1 \sqcup E_2$, i.e., $\mathcal{I}; (C, \Psi_C) \models E_i$ for some $i = 1, 2$. The induction hypothesis obtains $\Theta; C \vdash E_i$ and thus $\Theta; C \vdash E_1 \sqcup E_2$ by virtue of rule $\sqcup R_1$ (for $i = 1$) or $\sqcup R_2$ (for $i = 2$). In the other direction, given $\Theta; C \vdash E_1 \sqcup E_2$, we get $\Theta; C \vdash E_1$ or $\Theta; C \vdash E_2$ based on the fact that C is guarded (Prop. 3). In either case we can use the induction hypothesis to conclude $\mathcal{I}; (C, \Psi_C) \models E_i$ from which $\mathcal{I}; (C, \Psi_C) \models E_1 \sqcup E_2$ follows as required. □

Lemma 3 means that $\mathcal{I}$ is a constructive model of the TBox Θ. To see this let $F \sqsubseteq E \in \Theta$ and $(C, \Psi_C) \preceq^{\mathcal{I}} (D, \Psi_D)$ such that $\mathcal{I}; (D, \Psi_D) \models F$. Therefore, $\Theta; D \vdash F$ because of Lem. 3. From here we get $\Theta; D \vdash E$ by way of rules $\sqsubseteq L$ and Ax, which implies $\mathcal{I}; (D, \Psi_D) \models E$ by Lem. 3. Hence, overall, $\mathcal{I}; (C, \Psi_C) \models F \sqsubseteq E$.

Proposition 4. *The rules in Fig. 1 are sound for constructive subsumption in $\mathcal{UL}$ and complete relative to existentially guarded (general) TBoxes Θ.*

Proof. Soundness, viz. that $\Theta; C \vdash D$ implies $\Theta; C \models D$, is proved by induction on derivations. Regarding completeness, let a subsumption $\Theta; C \models D$ be constructively valid. Consider the canonical model $\mathcal{I}$ constructed above which satisfies $\mathcal{I} \models \Theta$. Take any guarded component $C \trianglerighteq P$ of C. By Lem. 2 (i) and rule Ax we have $\Theta; P \vdash C$. Hence, $\mathcal{I}; (P, \emptyset) \models C$ because of Lem. 3. But then the assumption $\Theta; C \models D$ implies $\mathcal{I}; (P, \emptyset) \models D$ which means $\Theta; P \vdash D$ again by Lem. 3. Since P with $C \trianglerighteq P$ was arbitrary we have shown that $\Theta; C \vdash D$ by virtue of Lem. 2 (ii). □

Theorem 2. *Subsumption checking in $\mathcal{UL}$ under the constructive semantics and relative to existentially guarded (general) TBoxes is in* PTime.

Proof. Let Θ be an existential TBox and C and D concept descriptions in $\mathcal{UL}$. We want to decide the subsumption $C \sqsubseteq D$ relative to Θ. Due to Prop. 4 we can decide $\Theta; C \models D$ by proof search in the calculus of Fig. 1. Because of the sub-formula property each possible node in a derivation tree is determined by a pair of concept descriptions in $\mathrm{Sf}(\Theta \cup \{C, D\})$. Thus the number of possible nodes in a tree is at worst quadratic in the size of Θ. We may systematically tabulate all pairs $(X, Y) \in \mathrm{Sf}(\Theta \cup \{C, D\})^2$ such that $\Theta; X \vdash Y$ using dynamic programming memorisation techniques and thus decide any fixed subsumption in polynomial time. □

5 Conclusion and Future Work

This paper highlights the dramatic effect that the choice of semantic interpretation can have on the complexity of reasoning in DL. Arguably, as long as there is no application domain identifiable for the $\mathcal{UL}$ fragment these results may only have theoretical interest. Perhaps our results can be extended to more expressive DLs that may eventually support practical applications. Note that from the expressiveness point of view the constructive semantics genuinely enriches the classical descriptive interpretation, so that the low complexity of $\mathcal{UL}$ is somewhat unexpected. However, as discussed in [16], the constructive semantics has useful applications which are independent of this complexity result.

We can see some immediate ways to generalise our results. The main results carry over to all TBoxes in which every $C \in \mathrm{Sf}(\Theta)$ has a non-empty *prime cover* $\mathrm{Cov}(C) \subseteq \mathrm{Sf}(\Theta)$ satisfying the conditions of Prop. 3 and Lem. 2. We have exploited this here for guarded TBoxes in the $\mathcal{UL}$ language where every concept has a prime cover. We conjecture that the PTIME result extends directly to $\mathcal{UL}^-$ which is $\mathcal{UL}$ plus limited universal quantification $\forall R.\bot$. For other extensions like (i) permitting negated atoms $\neg A$, (ii) adding conjunctions $\sqcap$ we expect coNP hardness, but hopefully not more since the Boolean reasoning should remain safely contained between the levels of existentials so that the Boolean combinatorics does not spill over quantifiers (exploiting that in constructive logic $\forall\neg$ is not the same as $\neg\exists$, etc.). One of our referees has pointed out that if we remove the guardedness restriction then we could express Disjunctive Distribution in the TBox which suggests that complexity might rise to EXPTIME. Since we apply constructive rather than classical semantics it is not clear whether such Disjunctive Distribution in TBoxes necessarily lead to exponentially large TBoxes/derivations or they can be eliminated in terms of an exponential number of polynomially sized TBoxes. We leave these investigations to future work.

Finally, it is worthwhile to add that the dualising reduction from $\mathcal{FL}_0$ to $\mathcal{UL}_0$ used in Section 3 can also be applied to axiomatic (cyclic or non-cyclic) TBoxes so that several other complexity results for the classical descriptive semantics should follow in a similar fashion. E.g., all the hardness/completeness results for $\mathcal{EL}$ (see, e.g., [11]) should carry over to the *"no S is P"* fragment $\{\forall, \sqcup\}$ under the classical semantics. Similarly, coNP-hardness for $\mathcal{FL}_0$ in acyclic TBoxes should imply coNP-hardness for $\mathcal{UL}_0$, PSPACE-hardness of $\mathcal{FL}_0$ should give PSPACE-hardness of $\mathcal{UL}_0$ in cyclic TBoxes. Of course, this applies to the classical descriptive semantics. For the constructive semantics the situation is open since DeMorgan-style dualisation does not work.

Acknowledgements This work has been funded by the German Research Council (DFG) under `ME 1427/4-1`. The authors would like to thank the anonymous reviewers for their suggestions to improve the presentation of this paper.

References

1. F. Baader. Using automata theory for characterizing the semantics of terminological cycles. *Annals of Math. and Artificial Intelligence*, pages 175–219, 1996.
2. F. Baader. Computing the least common subsumer in the description logic $\mathcal{EL}$ w.r.t. terminological cycles with descriptive semantics. In *Proc. Int'l Conf. on Conceptual Structures (ICCS2003)*, pages 117–130. Springer LNAI 2746, 2003.
3. F. Baader. Terminological cycles in a description logic with existential restrictions. In *Proc. 18th Int'l Joint Conf. on Artificial Intelligence*, pages 325–330. Morgan Kaufmann, 2003.
4. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$-envelope. In *Proc. 19th Joint Conf. on Artificial Intelligence (IJCAI 2005)*, pages 364–369, 2005.
5. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementations and Applications*. Cambridge University Press, 2003.
6. Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope further. In Kendall Clark and Peter F. Patel-Schneider, editors, *In Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.
7. G. Bellin, V. de Paiva, and E. Ritter. Extended Curry-Howard correspondence for a basic constructive modal logic. In *Methods for Modalities II*, November 2001.
8. S. Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and – what else? In R. López de Mantáras and L. Saitta, editors, *Proc. Eur. Conf. on Artificial Intelligence (ECAI-2004)*, pages 298–302, 2004.
9. V. de Paiva. Constructive description logics: what, why and how. In *Context Representation and Reasoning*, Riva del Garda, August 2006.
10. R. Dionne, E. Mays, and F. J. Oles. The equivalence of model-theoretic and structural subsumption in description logics. In *Proc. 13th Int'l Joint Conf. on Artificial Intelligence (IJCAI'93)*, pages 710–718, 1993.
11. Ch. Haase and C. Lutz. Complexity of subsumption in the $\mathcal{EL}$ family of description logics: acyclic and cyclic TBoxes. In *Proc. Europ. Conf. on Artificial Intelligence ECAI 2008*, 2008.
12. M. Hofmann. Proof-theoretic approach to description-logic. In *Proc. Logic in Computer Science LICS 2005*, 2005.
13. Y. Kazakov and H. de Nivelle. Subsumption of concepts in $\mathcal{FL}_0$ for (cyclic) terminologies with respect to descriptive semantics is PSPACE complete. In *Proc. Int'l Workshop Description Logics (DL 2003)*, volume 81 of *CEUR Electronic workshop proceedings*, 2003.
14. H. J. Levesque and R. J. Brachman. Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, 3:78–93, 1897.
15. M. Mendler and V. de Paiva. Constructive CK for contexts. In L. Serafini and P. Bouquet, editors, *Context Representation and Reasoning (CRR-2005)*, volume 13 of *CEUR Proceedings*, July 2005. Also presented at the Association for Symbolic Logic Annual Meeting, Stanford University, USA, 22nd March 2005.
16. M. Mendler and S. Scheele. Towards constructive description logics for abstraction and refinement. In *21st Int'l Workshop on Description Logics (DL2008)*. CEUR Workshop proceedings 353, May 2008.
17. B. Nebel. Terminological reasoning is inherently intractable. *Artificial Intelligence*, pages 235–249, 1990.

18. B. Nebel. Terminological cycles: Semantics and computational properties. In J. F. Sowa, editor, *Principles of Semantic Networks*. Morgan Kaufmann, 1991.
19. F. J. Oles R. Dionne, E. Mays. A non-well-founded approach to terminological cycles. In *Proc. 10th Nat'l Conf. of the Americ. Assoc. for Artificial Intelligence (AAAI'92)*, 1992.
20. D. Wijesekera. Constructive modal logic I. *Annals of Pure and Applied Logic*, 50:271–301, 1990.