On P/NP Dichotomies for $\mathcal{EL}$ Subsumption under Relational Constraints

A. Kurucz[1], F. Wolter[2], and M. Zakharyaschev[3]

[1] Department of Informatics
King's College London, U.K.
agi.kurucz@kcl.ac.uk
[2] Department of Computer Science
University of Liverpool, U.K.
frank@csc.liv.ac.uk
[3] Department of Computer Science and Information Systems,
Birkbeck College London, U.K.
michael@dcs.bbk.ac.uk

Abstract. We consider the problem of characterising relational constraints under which TBox reasoning in $\mathcal{EL}$ is tractable. We obtain P vs. coNP-hardness dichotomies for tabular constraints and constraints imposed on a single reflexive role.

1 Introduction

In recent years, the problem of describing role boxes (aka relational constraints) under which reasoning is within a given complexity class has become an important research topic in description logic (DL). For example, the development of $\mathcal{SROIQ}$ from $\mathcal{SHIQ}$ has mainly been driven by the desire to allow for more expressive relational constraints for which reasoning is still decidable and tableau decision procedures can be developed. As a result, in $\mathcal{SROIQ}$ one can express, among others, role inclusions of the form $r \circ s \sqsubseteq r$ and $s \circ r \sqsubseteq r$, reflexivity, transitivity and symmetry of roles [6, 7].

For $\mathcal{EL}$, underlying the *OWL 2 EL* profile of the *OWL 2* Web Ontology Language, the complexity of reasoning under relational constraints was investigated in [1, 2, 9]. For example, the subsumption problem for general TBoxes in $\mathcal{EL}$ is tractable for any finite set of constraints of the form

$$r_1(x_1, x_2) \wedge \cdots \wedge r_n(x_n, x_{n+1}) \to r_{n+1}(x_1, x_{n+1}) \tag{1}$$

(the order of the variables is essential). On the other hand, subsumption becomes ExpTime-complete in the presence of symmetry or functionality constraints [2].

The aim of this paper is to take a fresh look at how relational constraints influence the complexity of DL reasoning: rather than putting forward a new class of role boxes for which reasoning is decidable or within a certain complexity class, we attempt to *classify* relational constraints according to whether they lead to decidable or undecidable reasoning problems, or to reasoning within a given complexity bound. The ultimate aim of this approach is to obtain a complete map of how relational constraints determine the complexity of reasoning for most important DLs. Apart from its theoretical interest, such a map can also be used for the selection of role boxes with acceptable computational properties in future standardisation efforts.

In this paper, which extends [8], we take first steps in this program by starting to map out the border between tractability and intractability of TBox reasoning in $\mathcal{EL}$ under arbitrary relational constraints. One of the fundamental questions (left unanswered in this paper) is the following

Dichotomy Question: Is it the case that for any relational constraint, TBox reasoning in $\mathcal{EL}$ is either in P or coNP-hard?

By Ladner's Theorem, unless P = coNP, there exist problems that are coNP-intermediate (neither in P nor coNP-hard). The existence of relational constraints for which TBox reasoning in $\mathcal{EL}$ is coNP-intermediate would indicate that a general and complete map of the boundary between tractable and intractable is extremely hard to obtain. In contrast, a positive answer would probably come with an informative description of the tractable constraints.

Our initial findings indicate that informative dichotomy results on P versus coNP-hardness can indeed be obtained. For example, we show that

(d1) there are only *four universal* constraints on a single reflexive role r under which $\mathcal{EL}$ TBox reasoning is in P: (1) r is arbitrary, (2) the domain of r is a singleton, (3) r is transitive, (4) r is an equivalence relation. All other universal constraints are either invisible to $\mathcal{EL}$ TBox reasoning or lead to coNP-hard $\mathcal{EL}$ subsumption.

Here, by 'invisibility' we understand the following. It is well known that many relational constraints do not influence—or are invisible to—TBox reasoning: for example, for $\mathcal{EL}$ (and even $\mathcal{ALC}$), TBox reasoning over irreflexive relations coincides with TBox reasoning over arbitrary relations, and similarly for the class of finite and tree-like relational structures. In fact, one can use dichotomy **(d1)** to show that there are uncountably many 'visible' universal relational constraints on a single reflexive role for which $\mathcal{EL}$ subsumption is coNP-hard, but only four 'visible' universal constraints for which $\mathcal{EL}$ subsumption is in P.

Another dichotomy we prove in this paper is as follows:

(d2) Consider an arbitrary relational constraint (over a finite number of roles) such that the size of the domain of all interpretations satisfying this constraint is bounded by some natural number $n > 0$. Then $\mathcal{EL}$ subsumption over the interpretations satisfying the constraint is in P if all roles in those interpretations are functional. Otherwise $\mathcal{EL}$ subsumption is coNP-complete.

Currently, not much is known about dichotomies for more expressive languages. We note, however, recent work on an NP vs. PSpace dichotomy for satisfiability of classical modal formulas over frame classes definable by Horn sentences [5].

The paper is structured as follows. In Section 2, we define the extension $\mathcal{EL}_\bot$ of $\mathcal{EL}$ with the concept $\bot$ and all the model-theoretic notions we need. We prove our results for $\mathcal{EL}_\bot$ rather than $\mathcal{EL}$ and show, by a straightforward reduction in Section 6, that they hold for $\mathcal{EL}$ as well. In Section 3, we consider the relation between tractability and convexity (the disjunction property) and prove two general sufficient conditions for non-tractability. Then, in Sections 4 and 5, we prove the dichotomies **(d1)** and **(d2)** mentioned above.

2 Preliminaries

Fix two disjoint countably infinite sets NC of *concept names* and NR of *role names*. We use arbitrary concept names in NC for constructing concepts, but may restrict the set of available role names to some $\mathsf{R} \subseteq \mathsf{NR}$. Throughout this paper, we work with $\mathcal{EL}$ extended with the concept $\bot$, denoting the empty set. Thus, for $\mathsf{R} \subseteq \mathsf{NR}$, the $\mathcal{EL}_\bot$-*concepts* C *over* R are defined inductively as follows:

$$C \quad ::= \quad \top \quad | \quad \bot \quad | \quad A \quad | \quad C_1 \sqcap C_2 \quad | \quad \exists r.C,$$

where $A \in \mathsf{NC}$, $r \in \mathsf{R}$ and C, C_1, C_2 range over $\mathcal{EL}_\bot$-concepts over R. An R-*TBox* is a finite set of *concept inclusions* (CIs) $C \sqsubseteq D$, where C and D are $\mathcal{EL}_\bot$-concepts over R. An R-*interpretation* is of the form $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I} \neq \emptyset$ and $\cdot^\mathcal{I}$ is an *interpretation function* for concept names and role names in R. Complex concepts over R are interpreted in $\mathcal{I}$ as usual. If $C^\mathcal{I} \subseteq D^\mathcal{I}$, we say that $\mathcal{I}$ *satisfies* $C \sqsubseteq D$ and write $\mathcal{I} \models C \sqsubseteq D$. $\mathcal{I}$ is a *model* of an R-TBox $\mathcal{T}$, $\mathcal{I} \models \mathcal{T}$ in symbols, if it satisfies all the CIs in $\mathcal{T}$.

We now define what we understand by *relational constraints* on interpretations. An R-*frame* is a structure $\mathfrak{F} = (\Delta^\mathfrak{F}, \cdot^\mathfrak{F})$ where $\Delta^\mathfrak{F} \neq \emptyset$ and $\cdot^\mathfrak{F}$ is a map associating with each $r \in \mathsf{R}$ a relation $r^\mathfrak{F} \subseteq \Delta^\mathfrak{F} \times \Delta^\mathfrak{F}$. We say that an R-interpretation $\mathcal{I}$ is *based on* an R-frame $\mathfrak{F}$ if $\Delta^\mathcal{I} = \Delta^\mathfrak{F}$ and $r^\mathcal{I} = r^\mathfrak{F}$ for all $r \in \mathsf{R}$. An R-*constraint* is any class $\mathcal{K}$ of R-frames closed under isomorphic copies. For example, a constraint for $\mathsf{R} = \{r_1, r_2, r_3\}$ can consist of all R-frames $\mathfrak{F} = (\Delta^\mathfrak{F}, \cdot^\mathfrak{F})$ with arbitrary $r_1^\mathfrak{F}$, transitive $r_2^\mathfrak{F}$ and functional $r_3^\mathfrak{F}$. An interpretation $\mathcal{I}$ *satisfies* an R-constraint $\mathcal{K}$ if $\mathcal{I}$ is based on some $\mathfrak{F} \in \mathcal{K}$.

The *subsumption problem for an* R-*constraint* $\mathcal{K}$ is to decide, given an R-TBox $\mathcal{T}$ and two concepts C, D over R, whether $\mathcal{I} \models C \sqsubseteq D$ for every model $\mathcal{I}$ of $\mathcal{T}$ based on an R-frame in $\mathcal{K}$, in which case we write $\mathcal{T} \models_\mathcal{K} C \sqsubseteq D$. For singleton $\mathcal{K} = \{\mathfrak{F}\}$, we sometimes write $\mathcal{T} \models_\mathfrak{F} C \sqsubseteq D$.

Example 1. In the extension $\mathcal{EL}_\bot^+$ of $\mathcal{EL}_\bot$ [1], along with a TBox one can define an *RBox* containing inclusions of the form $r_1 \circ \cdots \circ r_n \sqsubseteq r_{n+1}$, where $r_1, \ldots, r_{n+1}$ are role names. Reasoning with RBoxes $\mathcal{R}$ is clearly captured by the frame condition $\mathcal{K}_\mathcal{R}$ containing all NR-frames $\mathfrak{F}$ such that

$$\mathfrak{F} \models \forall x_1 \ldots \forall x_{n+1} \ \big(r_1(x_1, x_2) \land \cdots \land r_n(x_n, x_{n+1}) \to r_{n+1}(x_1, x_{n+1})\big)$$

for all $r_1 \circ \cdots \circ r_n \sqsubseteq r_{n+1}$ in $\mathcal{R}$. According to [1, 9], the subsumption problem for any such $\mathcal{K}_\mathcal{R}$ is decidable in P. On the other hand, the subsumption problem for the class of symmetric frames is ExpTime-complete [2].

We say that R-constraints $\mathcal{K}_1$ and $\mathcal{K}_2$ are *TBox-equivalent* (in $\mathcal{EL}_\bot$) if we have $\mathcal{T} \models_{\mathcal{K}_1} C \sqsubseteq D$ iff $\mathcal{T} \models_{\mathcal{K}_2} C \sqsubseteq D$, for all R-TBoxes $\mathcal{T}$ and $\mathcal{EL}_\bot$-concepts C, D over R. For example, as is well-known, the class of all frames is TBox equivalent to the class of all irreflexive frames, and to the class of all finite frames. For an R-constraint $\mathcal{K}$, we denote by $\mathrm{FrTh}\mathcal{K}$ the union of all those R-constraints that are TBox equivalent in $\mathcal{EL}_\bot$ to $\mathcal{K}$. $\mathrm{FrTh}\mathcal{K}$ and $\mathcal{K}$ are TBox equivalent in $\mathcal{EL}_\bot$, and $\mathrm{FrTh}\mathcal{K}$ is the largest class that is TBox equivalent in $\mathcal{EL}_\bot$ to $\mathcal{K}$.

An R-constraint $\mathcal{K}$ is *TBox-definable* (in $\mathcal{EL}_\bot$) if there exists a set Γ of pairs $(\mathcal{T}, C \sqsubseteq D)$, where $\mathcal{T}$ is an R-TBox and C, D are $\mathcal{EL}_\bot$-concepts over R, such that $\mathcal{K} = \{\mathfrak{F} \mid \mathcal{T} \models_{\mathfrak{F}} C \sqsubseteq D, \text{ for all } (\mathcal{T}, C \sqsubseteq D) \in \Gamma\}$. Thus, $\mathcal{K}$ is TBox-definable iff $\mathcal{K} = \mathrm{FrTh}\mathcal{K}$, and any class of TBox-equivalent constraints contains exactly one TBox-definable class. In a similar way we can define TBox-definable classes of R-constraints for $\mathcal{EL}$ and more expressive DLs, say $\mathcal{ALC}$.

A *universal* R-constraint is a class of R-frames definable by universal first-order sentences in the signature R. Equivalently, by [10], a universal constraint is a first-order definable class of frames closed under taking subframes. The vast majority of frame constraints considered in modal and description logics are universal: transitivity, reflexivity, symmetry, weak linearity, just to mention a few. Typical examples of non-universal (first-order) constraints are the Church-Rosser property and density. As far as universal R-constraints are concerned, $\mathcal{EL}_\bot$ defines the same R-constraints as $\mathcal{ALC}$ (the proof is given in [8]):

Theorem 1. *Let $\mathcal{K}$ be a universal class of R-frames, for some $\mathsf{R} \subseteq \mathsf{NR}$. Then $\mathcal{K}$ is TBox-definable in $\mathcal{EL}_\bot$ iff it is TBox-definable in $\mathcal{ALC}$.*

We conjecture that Theorem 1 can be generalised to arbitrary (not necessarily first-order definable) classes of R-frames closed under subframes. Note that, without the subframe condition, there are classes of frames that are TBox-definable in $\mathcal{ALC}$ but not in $\mathcal{EL}_\bot$. One example is the *Church-Rosser property*

$$\forall x, y_1, y_2 \left(r(x, y_1) \land r(x, y_2) \to \exists z (r(y_1, z) \land r(y_2, z))\right).$$

3 Tractability and Convexity

In this section, we investigate the relationship between convexity (sometimes also called the disjunction property) and tractability. To this end, we need (formally not allowed in $\mathcal{EL}_\bot$) concepts of the form $C \sqcup D$, where C and D are $\mathcal{EL}_\bot$-concepts, which are interpreted in the obvious way by the union of the extensions of the disjuncts C and D. An R-constraint $\mathcal{K}$ is said to be *convex* if, for any R-TBox $\mathcal{T}$ and $\mathcal{EL}_\bot$-concepts F, C, D over R,

(conv) if $\quad \mathcal{T} \models_{\mathcal{K}} F \sqsubseteq C \sqcup D \quad$ then $\quad \mathcal{T} \models_{\mathcal{K}} F \sqsubseteq C \quad$ or $\quad \mathcal{T} \models_{\mathcal{K}} F \sqsubseteq D$.

Although convexity is closely related to tractability, they do not imply each other. It is readily checked that every relational constraint $\mathcal{K}$ defined by Horn sentences is convex. Thus, symmetry and functionality are examples of relational constraints that are convex but non-tractable [2]. The following example shows that tractability of $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ does not imply that $\mathcal{K}$ is convex:

Example 2. Consider the smallest class $\mathcal{K}$ of R-frames, for $\mathsf{R} = \{s, r, r'\}$, which is closed under subframes and contains all two-element irreflexive s-chains such that if $s(x, y)$ then either $r(x, y)$ or $r'(x, y)$. Thus, $\mathcal{K}$ is a universal constraint and $\emptyset \models_{\mathcal{K}} \exists s.\top \sqsubseteq \exists r.\top \sqcup \exists r'.\top$. As, $\emptyset \not\models_{\mathcal{K}} \exists s.\top \sqsubseteq \exists r.\top$ and $\emptyset \not\models_{\mathcal{K}} \exists s.\top \sqsubseteq \exists r'.\top$, $\mathcal{K}$ is not convex. On the other hand, as will be shown in the next section (see Theorem 4), $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is in P.

We now prove two general conditions, based on non-convexity, that imply non-tractability. The proofs of coNP-hardness are by reduction of the following *set splitting problem*, which is known to be NP-complete [4]:

– given a family I of subsets of a finite set S, decide whether there exists a *splitting* of (S, I), i.e., a partition S_1, S_2 of S such that each set $G \in I$ is split by S_1 and S_2 in the sense that it is not the case that $G \subseteq S_i$ for $i \in \{1, 2\}$.

We say that a class $\mathcal{K}$ of R-frames is *concept non-convex* if, for some R-TBox $\mathcal{T}$ and concepts F, C, D over R, we have $\mathcal{T} \models_{\mathcal{K}} F \sqsubseteq C \sqcup D$, and there exist an R-frame $\mathfrak{F} \in \mathrm{FrTh}\mathcal{K}$, a point $x \in \Delta^{\mathfrak{F}}$ and two models $\mathcal{I}_1$ and $\mathcal{I}_2$ of $\mathcal{T}$ based on $\mathfrak{F}$ such that $x \in F^{\mathcal{I}_1} \setminus D^{\mathcal{I}_1}$ and $x \in F^{\mathcal{I}_2} \setminus C^{\mathcal{I}_2}$. Our main tool for proving non-tractability results is the following:

Theorem 2. *If a class $\mathcal{K}$ of R-frames is concept non-convex, then $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is* coNP*-hard.*

Proof. Consider $\mathcal{T}$, F, C and D over R for which $\mathcal{T} \models_{\mathcal{K}} F \sqsubseteq C \sqcup D$, and there exist an R-frame $\mathfrak{F} \in \mathcal{K}$ with $x \in \Delta^{\mathfrak{F}}$ and two models $\mathcal{I}_1$ and $\mathcal{I}_2$ of $\mathcal{T}$ based on $\mathfrak{F}$ such that $x \in F^{\mathcal{I}_1} \setminus D^{\mathcal{I}_1}$ and $x \in F^{\mathcal{I}_2} \setminus C^{\mathcal{I}_2}$. Suppose (S, I) is an instance of the set splitting problem. Denote by $\mathcal{T}_i$, F_i, C_i and D_i, for $i \in S$, the copies of $\mathcal{T}$, F, C and D obtained by replacing every concept name A in them with A_i. Let

$$\mathcal{T}_{S,I} = \bigcup_{i \in S} \mathcal{T}_i \cup \{\bigsqcap_{i \in G} (B \sqcap C_i) \sqsubseteq \bot \mid G \in I\} \cup \{\bigsqcap_{i \in G} (B \sqcap D_i) \sqsubseteq \bot \mid G \in I\},$$

where B is a fresh concept name. We show now that there exists a splitting of (S, I) iff $\mathcal{T}_{S,I} \not\models_{\mathcal{K}} \bigsqcap_{i \in S} (B \sqcap F_i) \sqsubseteq \bot$. ($\Rightarrow$) Let S_1, S_2 be a splitting of (S, I). Define an interpretation $\mathcal{I}$ on $\mathfrak{F}$ by taking $A_i^{\mathcal{I}} = A^{\mathcal{I}_1}$ if $i \in S_1$, $A_i^{\mathcal{I}} = A^{\mathcal{I}_2}$ if $i \in S_2$, for all concept names A different from B, and $B^{\mathcal{I}} = \{x\}$. One can readily check that $\mathcal{I} \models \mathcal{T}_{S,I}$ and $\mathcal{I} \not\models \bigsqcap_{i \in S} (B \sqcap F_i) \sqsubseteq \bot$. ($\Leftarrow$) Suppose that $\mathcal{I} \models \mathcal{T}_{S,I}$ and there is $y \in \bigcap_{i \in S} (B^{\mathcal{I}} \cap F_i^{\mathcal{I}})$. We then set $S_1 = \{i \in S \mid y \in C_i^{\mathcal{I}}\}$ and $S_2 = S \setminus S_1$. It is readily checked that S_1, S_2 is a splitting of (S, I).

An R-constraint $\mathcal{K}$ is *closed under disjoint unions* if, for any $\mathfrak{F}_1, \mathfrak{F}_2 \in \mathcal{K}$ with $\Delta^{\mathfrak{F}_1} \cap \Delta^{\mathfrak{F}_2} = \emptyset$, we have $\mathfrak{F}_1 \cup \mathfrak{F}_2 \in \mathrm{FrTh}\mathcal{K}$, where $\Delta^{\mathfrak{F}_1 \cup \mathfrak{F}_2} = \Delta^{\mathfrak{F}_1} \cup \Delta^{\mathfrak{F}_2}$ and $r^{\mathfrak{F}_1 \cup \mathfrak{F}_2} = r^{\mathfrak{F}_1} \cup r^{\mathfrak{F}_2}$. We also say that $\mathcal{K}$ has a *free role* r if, for any $\mathfrak{F} \in \mathcal{K}$ and any $x, y \in \Delta^{\mathfrak{F}}$, the frame obtained by extending $r^{\mathfrak{F}}$ in $\mathfrak{F}$ with the pair (x, y) belongs to $\mathrm{FrTh}\mathcal{K}$. Note that all RBoxes, currently used in DL, correspond to constraints that are closed under disjoint unions and have infinitely many free roles (since typically DLs admit infinitely many role names and have finite RBoxes). The following condition is proved similarly to Theorem 2:

Theorem 3. *Suppose that an* R*-constraint* $\mathcal{K}$ *is closed under disjoint unions and has infinitely many free roles. If* $\mathcal{K}$ *is not convex then* $\mathcal{EL}_\perp$ *subsumption over* $\mathcal{K}$ *is* CONP*-hard.*

4 P/coNP Dichotomy for Tabular Constraints

A class $\mathcal{K}$ of R-frames is called *tabular* if there is $n > 0$ such that $|\Delta^{\mathfrak{F}}| \leq n$ for all $\mathfrak{F} \in \mathcal{K}$. The aim of this section is to characterise the tabular constraints $\mathcal{K}$ over which $\mathcal{EL}_\perp$ subsumption is tractable, that is, there is an algorithm which, given a TBox $\mathcal{T}$ and concepts C, D over R, can decide, in polynomial time, whether $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$. Clearly, $\mathcal{EL}_\perp$ subsumption over any tabular $\mathcal{K}$ belongs to CONP.

The characterisation of tabular constraints we are about to prove dichotomises them into functional and non-functional. A class $\mathcal{K}$ of R-frames is R-*functional* if, for any $\mathfrak{F} \in \mathcal{K}$, $r \in \mathsf{R}$ and $w \in \Delta^{\mathfrak{F}}$, we have $|\{v \in \Delta^{\mathfrak{F}} \mid (w, v) \in r^{\mathfrak{F}}\}| \leq 1$. For R-interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ based on a functional frame $\mathfrak{F}$, we write $\mathcal{I}_1 \leq \mathcal{I}_2$ if $A^{\mathcal{I}_1} \subseteq A^{\mathcal{I}_2}$ for all $A \in \mathsf{NC}$. Clearly, $\leq$ is a partial order.

Lemma 1. *Suppose that* $\mathcal{I}$ *is an interpretation based on a finite* R*-functional frame* $\mathfrak{F}$ *and* $w \in \Delta^{\mathcal{I}}$*. Given any* R*-concept* C*, one can decide in polynomial time in* $|C|$ *whether there exists an* R*-interpretation* $\mathcal{J}$ *such that* $\mathcal{I} \leq \mathcal{J}$ *and* $w \in C^{\mathcal{J}}$*. If such an interpretation exists, then there is a unique minimal (with respect to* $\leq$*)* R*-interpretation* $\mathcal{I}(w, C) \geq \mathcal{I}$ *with* $w \in C^{\mathcal{I}(w,C)}$*; moreover, this minimal interpretation can be constructed in polynomial time in* $|C|$*.*

We are now in a position to prove the main result of this section.

Theorem 4. *Let* $\mathcal{K}$ *be a tabular class of* R*-frames for a finite* $\mathsf{R} \subseteq \mathsf{NR}$*. If* $\mathcal{K}$ *is functional then* $\mathcal{EL}_\perp$ *subsumption over* $\mathcal{K}$ *is in* P*. Otherwise,* $\mathcal{EL}_\perp$ *subsumption over* $\mathcal{K}$ *is* CONP*-complete.*

Proof. Assume first that $\mathcal{K}$ is functional and we are given a TBox $\mathcal{T}$ and a CI $C' \sqsubseteq D'$ over R. Our polynomial time algorithm checking whether $\mathcal{T} \models_{\mathcal{K}} C' \sqsubseteq D'$ runs as follows. Let $\mathfrak{F}_1, \ldots, \mathfrak{F}_n$ be a list of all frames in $\mathcal{K}$ (up to isomorphism). For each $\mathfrak{F}_i$ and each $w \in \mathfrak{F}_i$, we do the following:

1. Let $\mathcal{I}$ be the R-interpretation based on $\mathfrak{F}_i$ with $A^{\mathcal{I}} = \emptyset$ for all $A \in \mathsf{NC}$.
2. Compute $\mathcal{I} := \mathcal{I}(w, C')$ if it exists (cf. Lemma 1). If it does not exist, return 'yes' and stop.
3. Apply the following rule exhaustively: for $C \sqsubseteq D \in \mathcal{T}$ and $v \in \Delta^{\mathcal{I}}$, if $v \in C^{\mathcal{I}}$ and $\mathcal{I}(v, D)$ does not exist, return 'yes' and stop; otherwise, if $\mathcal{I}(v, D) \neq \mathcal{I}$, set $\mathcal{I} = \mathcal{I}(v, D)$.
4. If $w \in (D')^{\mathcal{I}}$, return 'yes.' Otherwise, return 'no.'

It is easy to see that $\mathcal{T} \models_{\mathcal{K}} C' \sqsubseteq D'$ iff the output is 'yes' for all $\mathfrak{F}_i$ and $w \in \Delta^{\mathfrak{F}_i}$.

Suppose $\mathcal{K}$ is not R-functional. Then there exists $\mathfrak{F} \in \mathcal{K}$ with $w \in \Delta^{\mathfrak{F}}$ such that $|\{v \mid (w, v) \in r^{\mathfrak{F}}\}| \geq 2$. Let m be the maximal number for which there exist $r \in \mathsf{R}$, $\mathfrak{F} \in \mathcal{K}$ and $w \in \Delta^{\mathfrak{F}}$ with $|\{v \mid (w, v) \in r^{\mathfrak{F}}\}| = m$. Fix such r, $\mathfrak{F}$ and w. We prove CONP-hardness of $\mathcal{EL}_\perp$ subsumption over $\mathcal{K}$ using Theorem 2. To show that $\mathcal{K}$ is concept non-convex, consider the $\{r\}$-TBox $\mathcal{T}$ with the following CIs:

– $A \sqsubseteq \exists r.B_i$, for $1 \leq i \leq m$;
– $B_i \sqcap B_j \sqsubseteq \bot$, for $1 \leq i < j \leq m$;
– $A \sqsubseteq \exists r.B$
– $B_i \sqsubseteq E$, for $2 \leq i \leq m$.

Clearly, $\mathcal{T} \models_{\mathcal{K}} A \sqsubseteq \exists r.(B \sqcap B_1) \sqcup \exists r.(B \sqcap E)$. Consider next the interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ over $\mathfrak{F}$ where $w_1, \dots, w_m$ are the $r^{\mathfrak{F}}$-successors of w in $\mathfrak{F}$ and

– $A^{\mathcal{I}_i} = \{w\}$ and $B^{\mathcal{I}_i} = \{w_i\}$, for $i = 1, 2$;
– $B_i^{\mathcal{I}_i} = \{w_j\}$, for $i = 1, 2$ and $1 \leq j \leq m$;
– $E^{\mathcal{I}_i} = \{w_2, \dots, w_m\}$, for $i = 1, 2$.

Then we have $\mathcal{I}_i \models \mathcal{T}$, $w \in A^{\mathcal{I}_1} \setminus (\exists r.(B \sqcap E))^{\mathcal{I}_1}$ and $w \in A^{\mathcal{I}_2} \setminus (\exists r.(B \sqcap B_1))^{\mathcal{I}_2}$. By Theorem 2, $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is coNP-hard. And as we have mentioned above, $\mathcal{EL}_\bot$ subsumption for tabular constraints is in coNP.

The above proof of coNP-hardness goes through for many other constraints:

Theorem 5. *Let $\mathcal{K}$ be a class of R-frames such that there are $r \in \mathsf{R}$ and $n \geq 2$ for which (i) no point in frames from $\mathcal{K}$ has $> n$ r-successors, and (ii) at least one point in a frame from $\mathcal{K}$ has ≥ 2 r-successors. Then $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is coNP-hard.*

5 P/coNP-hardness Dichotomy for Universal Reflexive Constraints

In this section, we assume that $\mathsf{R} = \{r\}$ and consider *universal* classes of R-frames $\mathfrak{F}$ with *reflexive* $r^{\mathfrak{F}}$.

Theorem 6. *Let $\mathcal{K}$ be a universal constraint for a single reflexive relation. If $\mathcal{K}$ is not TBox equivalent to any of the following classes:*

(sin) *the class of all singleton frames,*
(tra) *the class of all transitive frames,*
(equ) *the class of all equivalence relations,*
(all) *the class of all frames,*
(sym) *the class of all symmetric frames,*

then $\mathcal{K}$ is concept non-convex, and so $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is coNP-hard. $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is also coNP-hard if $\mathcal{K}$ is TBox equivalent to **(sym)**. *However, if $\mathcal{K}$ is TBox equivalent to one of* **(sin)**, **(tra)**, **(equ)** *or* **(all)**, *then $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is in* P.

Note that there are uncountably many distinct universal TBox definable classes of frames with a single reflexive relation (see [8], where this is proved for quasi-orders). Thus, only four out of uncountably many possible constraints lead to tractable TBox reasoning; for all the rest, $\mathcal{EL}_\bot$ subsumption is coNP-hard.

Here we only give a brief sketch of the proof of Theorem 6. Note first that the polynomial upper bound follows from [1, 8]; non-tractability for **(sym)** is shown similarly to Theorem 7 below. To prove the remaining claim, we require

Lemma 2. *Let $\mathcal{K}$ be a universal class of reflexive frames.*

- *If $\mathcal{K}$ is not TBox equivalent to* **(all)***, then there exists a finite reflexive tree $\mathfrak{F}$ such that $\mathfrak{F} \notin \mathrm{FrTh}\mathcal{K}$.*
- *If $\mathcal{K}$ consists of symmetric frames and is not TBox equivalent to* **(sym)***, then there exists a finite reflexive and symmetric tree $\mathfrak{F}$ such that $\mathfrak{F} \notin \mathrm{FrTh}\mathcal{K}$.*
- *If $\mathcal{K}$ consists of transitive frames and is not TBox equivalent to* **(tra)***, then there exists a finite reflexive and transitive tree $\mathfrak{F}$ such that $\mathfrak{F} \notin \mathrm{FrTh}\mathcal{K}$.*

Proof sketch. We prove the first claim; the remaining ones are treated similarly. As $\mathcal{K}$ is not TBox equivalent to **(all)**, there are $\mathcal{T}$, C, D such that $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$ and $\mathcal{T} \not\models_{\mathcal{K}'} C \sqsubseteq D$, where $\mathcal{K}'$ is the class of all frames. By applying standard unravelling to a witness interpretation for $\mathcal{T} \not\models_{\mathcal{K}'} C \sqsubseteq D$, we obtain a (possibly infinite) reflexive tree $\mathfrak{F} \notin \mathrm{FrTh}\mathcal{K}$. If $\mathfrak{F}$ is finite, we are done. Otherwise, using the fact that $\mathcal{K}$ is universal and employing Tarski's finite embedding property [10], we can show that there is a finite subtree $\mathfrak{F}'$ of $\mathfrak{F}$ such that $\mathfrak{F}' \notin \mathrm{FrTh}_T\mathcal{K}$.

Having Lemma 2 at hand, we can now proceed with a case distinction. Suppose $\mathcal{K}$ is a non-empty universal class of reflexive frames that is not TBox equivalent to any of the classes mentioned in Theorem 6. Then, by Lemma 2, there exists a reflexive tree $\mathfrak{F}$ such that $\mathfrak{F} \notin \mathrm{FrTh}\mathcal{K}$, $\mathfrak{F}' \in \mathrm{FrTh}\mathcal{K}$, for any *proper* subframe $\mathfrak{F}'$ of $\mathfrak{F}$, and one of the following conditions holds:

1. $\mathfrak{F}$ is the singleton frame;
2. $\mathfrak{F}$ is the two-element r-chain;
3. $\mathfrak{F}$ contains a point w with least two r-successors, and all r-successors of w are leaves in $\mathfrak{F}$;
4. $\mathfrak{F}$ contains distinct points w, w_1, w_2 such that $(w, w_1) \in r^{\mathfrak{F}}$, $(w_1, w_2) \in r^{\mathfrak{F}}$ and w_2 is a leaf, which is the only r-successor of w_1.

Case 1. This case is actually impossible because it implies that $\mathcal{K}$ is empty (remember that $\mathcal{K}$ is universal, and so closed under subframes).

Case 2. In this case, $\mathcal{K}$ is a class of symmetric frames. Since we assume that $\mathcal{K}$ is not TBox equivalent to **(sym)**, one can apply the second claim of Lemma 2 to obtain a finite reflexive and symmetric tree $\mathfrak{F}$ such that $\mathfrak{F} \notin \mathrm{FrTh}\mathcal{K}$. A case distinction (similar to the one we are currently doing) shows that, since $\mathcal{K}$ is not TBox equivalent to **(sin)**, $\mathcal{K}$ is concept non-convex.

Case 3. Let us remove a proper r-successor of w from $\mathfrak{F}$ and denote by $\mathfrak{H}$ the resulting frame, which belongs to $\mathrm{FrTh}\mathcal{K}$. Let w_1 be one of the remaining successors of w in $\mathfrak{H}$. Denote by $\mathfrak{H}'$ the frame obtained from $\mathfrak{H}$ by adding a fresh r-successor w_2 to w_1, and by w_0 the root of $\mathfrak{H}'$. Two cases are possible now.

Case 3.1: either $\mathfrak{H}' \in \mathrm{FrTh}\mathcal{K}$ or the expansion of $\mathfrak{H}'$ by adding (w, w_2) to $r^{\mathfrak{H}'}$ is in $\mathrm{FrTh}\mathcal{K}$. Take additional concept names A and $\bar{A}$. To show that $\mathcal{K}$ is concept non-convex, we will use $C_1 = \exists r^2.(A' \sqcap \exists r^2.\bar{A}')$ and $C_2 = \exists r^2.(\bar{A}' \sqcap \exists r^2.A')$, where $A' = A_{w_1} \sqcap A$, $\bar{A}' = A_{w_1} \sqcap \bar{A}$ and $\exists r^m.C$ is an abbreviation defined inductively by taking $\exists r^0.C = C$ and $\exists r^{m+1}.C = \exists r.\exists r^m.C$.

In addition, we require a generic way of describing frames using TBoxes. Given an R-frame $\mathfrak{R}$, let A_u be a fresh concept name for every $u \in \Delta^{\mathfrak{R}}$. Let $\mathcal{T}_S(\mathfrak{R})$ be the (possibly infinite) TBox with the following CIs:

– $A_u \sqsubseteq \exists r.A_v$, for $(u,v) \in r^{\mathfrak{R}}$;
– $A_u \sqcap A_v \sqsubseteq \bot$, for $u \neq v$;
– $A_u \sqcap \exists r.A_v \sqsubseteq \bot$, for $(u,v) \notin r^{\mathfrak{R}}$.

One can show that, for any R-frame $\mathfrak{R}$ with root w (from which all other points are reachable via roles) and any R-frame $\mathfrak{F}$, we have $\mathcal{T}_S(\mathfrak{R}) \not\models_{\mathfrak{F}} A_w \sqsubseteq \bot$ iff $\mathfrak{R}$ is a p-morphic image of a subframe of $\mathfrak{F}$.

Returning to Case 3.1., define $\mathcal{T}$ to be the TBox with the following CIs:

$$\mathcal{T}_S(\mathfrak{H}), \quad A_w \sqsubseteq \exists r^2.A', \quad A_w \sqsubseteq \exists r^2.\bar{A}'.$$

Then $\mathcal{T} \models_{\mathcal{K}} A_{w_0} \sqsubseteq \exists r^m.(A_w \sqcap C_1) \sqcup \exists r^m.(A_w \sqcap C_2)$, where m is the distance between w_0, w, but $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \exists r^m.(A_w \sqcap C_1)$, $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \exists r^m.(A_w \sqcap C_2)$.

Case 3.2: suppose that Case 3.1 does not hold. Denote by w_0 the root of $\mathfrak{H}$. Take a fresh concept name A and consider the TBox $\mathcal{T}$ with the following CIs:

– $\mathcal{T}_S(\mathfrak{H})$,
– $A \sqcap \exists r.A_v \sqsubseteq \bot$, for all v with $(w,v) \notin r^{\mathfrak{H}}$,
– $A_v \sqcap \exists r.A \sqsubseteq \bot$, for all v with both $(v,w) \notin r^{\mathfrak{H}}$ and $(v,w_1) \notin r^{\mathfrak{H}}$,
– $A \sqcap \exists r.A_{w'} \sqsubseteq \exists r.A_w$, for $(w,w') \in r^{\mathfrak{H}}$, $w' \neq w_1$,
– $A_w \sqsubseteq \exists r.(A \sqcap \exists r.A_{w_1})$,
– if w has an r-predecessor w_p, then $A_{w_p} \sqcap \exists r.A \sqsubseteq \exists r.(A_w \sqcap \exists r.(A \sqcap \exists r.A_w))$.

Then $\mathcal{T} \models_{\mathcal{K}} A_{w_0} \sqsubseteq \exists r^m.(A_w \sqcap \exists r.(A \sqcap \exists r.A_w)) \sqcup \exists r^m.(A_w \sqcap \exists r.(A_{w_1} \sqcap \exists r.A))$, but $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq B$ for either of the disjuncts B in the right-hand side.

Case 4. A case distinction similar to, but much more tedious than the previous ones shows that $\mathcal{K}$ is concept non-convex if the constraint $\mathcal{K}$ is not transitive. The case where $\mathcal{K}$ is a class of transitive frames has been considered in [8], and one can easily modify the proofs given there to show that all universal classes of transitive and reflexive frames, which are not TBox equivalent to **(sin)**, **(equ)** or the class of all transitive and reflexive frames, are concept non-convex.

Typically, in DL applications one role is not enough. Therefore, the question is whether the four universal constraints guaranteeing tractability for a single reflexive relation still ensure tractability if more than one role is considered. This is well known to be the case for transitivity and reflexivity, and this is trivially the case for the singleton frame. Equivalence relations behave not so well:

Theorem 7. *If $\mathcal{K}$ is a constraint consisting of two (or more) equivalence relations, then $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is* NP*-hard. In particular, tractability of $\mathcal{EL}_\bot$ subsumption is not preserved under fusions in the sense of [3].*

Proof sketch.[4] The proof is by reduction of SAT. Let φ be a formula in NNF with the variables $p_1, \dots, p_{2n}$, and let r_1, r_2 be equivalence relations. We use T_k, F_k for the truth-values of the variable p_k, and L_j as a marker for the level j in a 'tree.' We generate a full binary tree of depth $2n+1$, using the CIs

$$L_{2i} \sqsubseteq \exists r_1.(T_{2i+1} \sqcap L_{2i+1}) \sqcap \exists r_1.(F_{2i+1} \sqcap L_{2i+1}), \tag{2}$$

$$L_{2i+1} \sqsubseteq \exists r_2.(T_{2i+2} \sqcap L_{2i+2}) \sqcap \exists r_2.(F_{2i+2} \sqcap L_{2i+2}), \tag{3}$$

[4] Based on an idea suggested by Carsten Lutz.

for $i < n$. Then we propagate the truth-values T_k and F_k to the leaves using

$$L_{2j} \sqcap \exists r_2.(L_{2j-1} \sqcap Q_k) \sqsubseteq Q_k, \quad \text{for } 1 \leq j \leq n,\ 1 \leq k \leq 2j-1, \tag{4}$$
$$L_{2j+1} \sqcap \exists r_1.(L_{2j} \sqcap Q_k) \sqsubseteq Q_k, \quad \text{for } 1 \leq j < n,\ 1 \leq k \leq 2j, \tag{5}$$

for $Q_k = T_k, F_k$. Take a fresh X_ψ, for every subformula ψ of φ, and the CIs

$$X_{p_k} \equiv T_k, \quad X_{\neg p_k} \equiv F_k, \quad X_{\psi_1 \wedge \psi_2} \equiv X_{\psi_1} \sqcap X_{\psi_2}, \tag{6}$$
$$X_{\psi_1} \sqsubseteq X_{\psi_1 \vee \psi_2}, \quad X_{\psi_2} \sqsubseteq X_{\psi_1 \vee \psi_2}. \tag{7}$$

Let $\mathcal{T}$ be the TBox containing all the CIs (2)–(7), and $L_{2n} \sqcap X_\varphi \sqsubseteq \bot$. One can show that $\mathcal{T} \models_{\mathcal{K}} L_0 \sqsubseteq \bot$ iff φ is satisfiable.

6 $\mathcal{EL}$ and $\mathcal{EL}_\bot$

So far, we have considered $\mathcal{EL}_\bot$ rather than $\mathcal{EL}$. The main reason is that $\bot$ makes proofs more transparent. We now show that Theorems 4–7 above hold for $\mathcal{EL}$.

An R-frame $\mathfrak{F}'$ is called a *generated subframe* of an R-frame $\mathfrak{F}$ if it is a subframe of $\mathfrak{F}$ and, for all $u, v \in \Delta^{\mathfrak{F}}$ and $r \in \mathsf{R}$, if $(u, v) \in r^{\mathfrak{F}}$ and $u \in \Delta^{\mathfrak{F}'}$ then $v \in \Delta^{\mathfrak{F}'}$. Given $v \in \Delta^{\mathfrak{F}}$, the *subframe of* $\mathfrak{F}$ *generated by* v is the smallest generated subframe of $\mathfrak{F}$ containing v.

Theorem 8. *Let $\mathcal{K}$ be an R-constraint closed under generated subframes, for a finite R. Then $\mathcal{EL}_\bot$ subsumption over $\mathcal{K}$ is polynomially reducible to $\mathcal{EL}$ subsumption over $\mathcal{K}$, and, for any R-constraint $\mathcal{K}'$ closed under generated subframes, $\mathcal{K}'$ is TBox-equivalent to $\mathcal{K}$ in $\mathcal{EL}_\bot$ iff $\mathcal{K}'$ is TBox-equivalent to $\mathcal{K}$ in $\mathcal{EL}$.*

Proof. Let $\mathcal{T}$ and $C \sqsubseteq D$ in $\mathcal{EL}_\bot$ be given. We may assume that $\bot$ occurs in them only in the form $E \sqsubseteq \bot$, with E being an $\mathcal{EL}$-concept. Let B be a fresh concept name, and let $\mathcal{T}'$ and D' result from $\mathcal{T}$ and D, respectively, by replacing all $\bot$ with B. Set $\mathcal{T}'' = \mathcal{T}' \cup \{\exists r.B \sqsubseteq B \mid r \in \mathsf{R}\} \cup \{B \sqsubseteq D'\}$. We claim that $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$ iff $\mathcal{T}'' \models_{\mathcal{K}} C \sqsubseteq D'$. Clearly, if $\mathcal{T} \not\models_{\mathcal{K}} C \sqsubseteq D$, then $\mathcal{T}'' \not\models_{\mathcal{K}} C \sqsubseteq D'$: for if we have a witness model for $\mathcal{T} \not\models_{\mathcal{K}} C \sqsubseteq D$, then we can interpret B by the empty set to obtain a model of $\mathcal{T}''$ refuting $C \sqsubseteq D'$. Conversely, if $\mathcal{T}'' \not\models_{\mathcal{K}} C \sqsubseteq D'$, take an interpretation $\mathcal{I}$ based on a frame in $\mathcal{K}$ and $v \in \Delta^{\mathcal{I}}$ such that $\mathcal{I} \models \mathcal{T}''$ but $v \in C^{\mathcal{I}} \setminus (D')^{\mathcal{I}}$. Let $\mathfrak{F}$ be the subframe generated by v in the underlying frame of $\mathcal{I}$. Then $\mathfrak{F} \in \mathcal{K}$ and $B^{\mathcal{I}} \cap \Delta^{\mathfrak{F}} = \emptyset$. Hence $\mathcal{T} \not\models_{\mathfrak{F}} C \sqsubseteq D$, as required.

It follows from Theorem 8 that Theorems 6 and 7 hold for $\mathcal{EL}$ in place of $\mathcal{EL}_\bot$. Theorem 4 can be proved for $\mathcal{EL}$ as follows. Let $\mathcal{K}$ be a non-functional tabular constraint. Then the class $\mathcal{K}'$ of subframes of frames from $\mathcal{K}$ is still a non-functional tabular constraint and $\models_{\mathcal{K}'}$ is polynomially reducible to $\models_{\mathcal{K}}$, both for $\mathcal{EL}$ and $\mathcal{EL}_\bot$ (using relativisation). Thus, by Theorem 4 for $\mathcal{EL}_\bot$ and Theorem 8, the $\mathcal{EL}$ subsumption problem for $\mathcal{K}'$ is coNP-hard. Hence it is coNP-hard for $\mathcal{K}$. Theorem 5 can be proved similarly.

7 Open Problems and Conjectures

The main open problem in the area is the dichotomy question formulated in the introduction. If the answer to this question is positive, then the proof will probably require some new techniques and a great number of case distinctions.

We conjecture that a transparent dichotomy, possibly more involved than Theorem 6, can be obtained for arbitrary relational constraints on a single reflexive relation. Of course, an additional problem in this case is how to deal with non first-order constraints. A possible approach can be illustrated by the following result from [8]. Call a constraint *subframe* if it is closed under the formation of subframes. A *Noetherian* partial order is a reflexive and transitive relation without infinite ascending chains. Let $\mathcal{N}$ be the (non-elementary) class of all Noetherian partial orders. It is proved in [8] that $\mathcal{EL}_\bot$ subsumption over a subframe constraint $\mathcal{K} \subseteq \mathcal{N}$ is tractable iff $\mathcal{K}$ is TBox equivalent either to the single element frame or to $\mathcal{N}$.

When moving beyond the 'bounded' constraints of Theorems 4 and 5, it seems to be much harder to obtain general results for relations that can be nonreflexive than for the reflexive ones. For example, in contrast to the reflexive case, $\mathcal{EL}_\bot$ subsumption is now also in P for the constraints $\mathcal{K}_n$ consisting of (irreflexive) trees of depth $\leq n$. Thus, there are infinitely many transitive classes with a single relation for which $\mathcal{EL}_\bot$ subsumption is tractable.

References

1. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. of IJCAI. pp. 364–369 (2005)
2. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope further. In: Proc. OWLED (2008)
3. Baader, F., Lutz, C., Sturm, H., Wolter, F.: Fusions of description logics and abstract description systems. J. Artif. Intell. Res. (JAIR) 16, 1–58 (2002)
4. Garey, M., Johnson, D.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman & Co. (1979)
5. Hemaspaandra, E., Schnoor, H.: On the complexity of elementary modal logics. In: Proc. of STACS, pp. 349–360 (2008)
6. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible $\mathcal{SROIQ}$. In: Proc. KR, pp. 57–67 (2006)
7. Kazakov, Y.: An extension of complex role inclusion axioms in the description logic $\mathcal{SROIQ}$. In: Proc. of IJCAR. pp. 472–486 (2010)
8. Kurucz, A., Wolter, F., Zakharyaschev, M.: Islands of tractability for relational constraints: towards dichotomy results for the description logic $\mathcal{EL}$. In: Advances in Modal Logic, Vol.8. pp. 271–291 (2010)
9. Sofronie-Stokkermans, V.: Locality and subsumption testing in $\mathcal{EL}$ and some of its extensions. In: Advances in Modal Logic, Vol.7. pp. 315–339 (2008)
10. Tarski, A.: Contributions to the theory of models I, II. Indag. Mathematicae 16, 572–588 (1954)