Reasoning in OWL 2 EL with Hierarchical Concrete Domains (Extended Version)

Francesco Kriegel[1,2]

[1]Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI)
francesco.kriegel@tu-dresden.de

Abstract. The $\mathcal{EL}$ family of description logics facilitates efficient polynomial-time reasoning and has been standardized as the profile OWL 2 EL of the Web Ontology Language. $\mathcal{EL}$ can represent and reason not only with symbolic knowledge but also with concrete knowledge expressed by numbers, strings, and other concrete datatypes. Such concrete domains must be convex to avoid introducing disjunctions "through the backdoor." However, the hitherto existing concrete domains provide only limited utility. In order to overcome this issue, we introduce a novel form of concrete domains based on semi-lattices. They are convex by design and can thus be integrated into Horn-DLs such as $\mathcal{EL}$. Moreover, they allow for FBoxes to express dependencies between concrete features. We describe four instantiations concerned with real intervals, 2D-polygons, regular languages, and graphs.

1 Introduction

Concrete domains can be integrated in description logics (DLs) in order to refer to concrete knowledge expressed by numbers, strings, and other concrete datatypes [8]. They have mainly been investigated with DLs that are not Horn, such as $\mathcal{ALC}$ and its extensions, regarding decidability and complexity [15, 18, 20, 41, 42, 43], reasoning procedures [25, 26, 42, 43, 44, 48], an algebraic characterization [13, 49], and their expressive power [4, 7].

For computationally tractable description logics, such as the $\mathcal{EL}$ family, other conditions on the concrete domains than above must be imposed. On the one hand, it must not be possible to introduce disjunction through the concrete domain into the logical domain so that the DL part retains its Horn character. On the other hand, reasoning in the concrete domain itself should be tractable. Both is guaranteed for p-admissible concrete domains [5]. Concrete domains have also been integrated with DL-Lite [3].

The hitherto existing p-admissible concrete domains for $\mathcal{EL}$ provide only limited utility. Using the concrete domain $\mathcal{D}_{\mathbb{Q},\mathsf{diff}}$ [5], we could express with the concept inclusions $(\mathsf{sys} \geq 140) \sqsubseteq \mathsf{Hypertension}$ and $(\mathsf{dia} \geq 90) \sqsubseteq \mathsf{Hypertension}$ that a systolic blood pressure of 140 or higher indicates hypertension, as does a diastolic blood pressure of at least 90. Since the opposite relations $\leq$ are not available to ensure convexity, neither non-elevated blood pressure (dia. $<$ 120 and sys. $<$ 70) nor elevated blood pressure (dia. between 120 and 140, and sys. between 70 and 90) are expressible. Mixed inequalities $<$, $\leq$, $>$, and $\geq$ may be used under certain limitations which of them may occur in left-hand sides and, respectively, in right-hand sides of concept inclusions [45]. While this retains convexity of the concrete domain, reasoning is then rather impaired since the usual completion procedure is only complete for consistency and classification, but not for subsumption.

An algebraic characterization of p-admissible concrete domains has put forth a further concrete domain $\mathcal{D}_{\mathbb{Q},\mathsf{lin}}$, which supports linear combinations of numerical features [12, 14]. For instance, the concept inclusion $\top \sqsubseteq (\mathsf{sys} - \mathsf{dia} - \mathsf{pp} = 0)$ expresses that the pulse pressure is the difference between the systolic and the diastolic blood pressure. In the medical domain, the combined expressivity of $\mathcal{D}_{\mathbb{Q},\mathsf{diff}}$ and $\mathcal{D}_{\mathbb{Q},\mathsf{lin}}$ would be useful since then with the concept inclusion $\mathsf{ICUPatient} \sqcap (\mathsf{pp} > 50) \sqsubseteq \mathsf{NeedsAttention}$ it could be expressed that intensive-care patients with a pulse pressure exceeding 50 need attention — but this combination is not convex anymore [2].

We introduce a novel form of concrete domains based on semi-lattices. A semi-lattice $(L, \leq, \wedge)$ consists of a set L, a partial order $\leq$, and a binary meet operation $\wedge$. The elements of L are taken as concrete values, and $\leq$ is understood as an "information order," i.e. $p \leq q$ means that p is more specific than q, like a subsumption order between concepts. The meet operation $\wedge$ is used to combine two values p and q to their meet value $p \wedge q$, which is the most general value that is more specific than both p and q. For instance, real intervals form a semi-lattice with subset inclusion $\subseteq$ as partial order and intersection $\cap$ as meet operation. With that, the statement $\mathsf{NonElevatedBP} \equiv (\mathsf{sys} \subseteq [0, 120)) \sqcap (\mathsf{dia} \subseteq [0, 70))$ defines non-elevated blood pressure.

Our new *hierarchical concrete domains* are convex by design, simply because a general value of a feature (such as $\mathsf{sys} \subseteq [0, 120)$) does not imply the disjunction of all more specific feature values (such as $\mathsf{sys} \subseteq [0, 0]$, $\mathsf{sys} \subseteq [1, 1]$, $\ldots$, $\mathsf{sys} \subseteq [119, 119]$). Atomic feature values are supported nonetheless when these are available as atoms in the semi-lattice. For instance, specific numerical values can be represented by singleton intervals.

In addition, we introduce *FBoxes* consisting of *feature inclusions* that describe dependencies between features as well as aggregations of features. For instance, through the feature inclusion $\mathsf{pp} \subseteq \mathsf{sys} - \mathsf{dia}$ we can obtain an interval value of the pulse pressure given intervals of the systolic and the diastolic blood pressure. With the concept inclusion $\mathsf{ICUPatient} \sqcap (\mathsf{pp} \subseteq (50, \infty)) \sqsubseteq \mathsf{NeedsAttention}$ we can now express that intensive-care patients having a pulse pressure above 50 need attention and, unlike in the combination of $\mathcal{D}_{\mathbb{Q},\mathsf{diff}}$ and $\mathcal{D}_{\mathbb{Q},\mathsf{lin}}$, computationally reason with that in polynomial time.

We provide four instantiations of hierarchical concrete domains based on real intervals, 2D-polygons, regular languages, and graphs. The former two are not only convex, but indeed p-admissible, i.e. equipping a DL from the $\mathcal{EL}$ family with them facilitates polynomial-time reasoning. In particular, we can employ linear programming for reasoning in the interval domain when the FBox is affine. The regular-language domain is also convex (again, by design) but requires exponential time for reasoning. However, this only affects the concrete-domain reasoning itself so that reasoning in the logical $\mathcal{EL}$ part still runs in polynomial time. This holds similarly for the graph domain.

Of practical relevance is that our hierarchical concrete domains can be seamlessly integrated into the completion procedure and the ELK reasoner [5, 6, 35]. We demonstrate this for the case where nominals must be used safely, i.e. nominals must not occur in conjunctions and right-hand sides of concept inclusions must not be single nominals. We conjecture that full support for nominals can be achieved in the same way as without concrete domains [34].

2 Preliminaries

We work with the description logic $\mathcal{EL}^{++}[\mathcal{D}]$ (OWL 2 EL) where $\mathcal{D}$ is a P-admissible concrete domain (as defined below). Consider a set $\mathbf{C}$ of *atomic concepts*, a set $\mathbf{R}$ of *roles*, a set $\mathbf{I}$ of *individuals*, a set $\mathbf{F}$ of *features*, and a set $\mathbf{P}$ of *predicates* where each $P \in \mathbf{P}$ has an arity $\mathsf{ar}(P) \in \mathbb{N}$. There are two special concepts $\bot$ and $\top$ with fixed meaning. A *constraint* is of the form $\exists f_1, \ldots, f_k. P$ where P is a k-ary predicate and $f_1, \ldots, f_k$ are features. We may also denote it by $\exists \boldsymbol{f}. P$ where $\boldsymbol{f} \coloneqq (f_1, \ldots, f_k)$ is a feature vector with the same arity as P. *Compound concepts* are built by

$$C ::= \bot \mid \top \mid \{i\} \mid A \mid \exists \boldsymbol{f}. P \mid C \sqcap C \mid \exists r. C$$

where A ranges over all atomic concepts, r over all roles, i over all individuals, and $\exists \boldsymbol{f}. P$ over all constraints. A *knowledge base (KB)* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$ concerning concepts C and D, *role inclusions (RIs)* $R \sqsubseteq s$ involving *role chains* generated by $R ::= \varepsilon \mid R_1$, $R_1 ::= r \mid R_1 \circ R_1$ and roles s, and *range inclusions* $\mathsf{Ran}(r) \sqsubseteq C$ referring to roles r and concepts C — but every $\mathcal{EL}^{++}[\mathcal{D}]$ KB must satisfy an additional condition as explained in Section 4.

As syntactic sugar, we have *concept assertions* $\{i\} \sqsubseteq C$ (also written $i : C$), *role assertions* $\{i\} \sqsubseteq \exists r. \{j\}$ (also written $(i, j) : r$), *domain inclusions* $\exists r. \top \sqsubseteq C$ (also written $\mathsf{Dom}(r) \sqsubseteq C$), and *role exclusions* $\exists r_1. \ldots \exists r_n. \top \sqsubseteq \bot$ (also written $r_1 \circ \cdots \circ r_n \sqsubseteq \bot$). Statements $C \sqsubseteq \bot$ are also called *concept exclusions*. Each KB $\mathcal{K}$ can be subdivided into an *ABox* $\mathcal{A}$ consisting of all concept and role assertions, an *RBox* $\mathcal{R}$ consisting of all role inclusions and exclusions, and a *TBox* $\mathcal{T}$ consisting of the remaining statements. The TBox together with the RBox is also called an *ontology* $\mathcal{O}$. Other authors do not use the denotation "knowledge base" and call it "ontology" instead, i.e. they also consider the assertions as part of the ontology.

The semantics are defined through the fixed concrete domain $\mathcal{D}$ and all interpretations $\mathcal{I}$. The *concrete domain* $\mathcal{D} \coloneqq (\mathsf{Dom}(\mathcal{D}), \cdot^{\mathcal{D}})$ consists of a set $\mathsf{Dom}(\mathcal{D})$ of *values* and an interpretation function $\cdot^{\mathcal{D}}$ that sends each predicate $P \in \mathbf{P}$ to a relation over $\mathsf{Dom}(\mathcal{D})$ with arity $\mathsf{ar}(P)$, i.e. $P^{\mathcal{D}} \subseteq \mathsf{Dom}(\mathcal{D})^{\mathsf{ar}(P)}$.

An interpretation $\mathcal{I} \coloneqq (\mathsf{Dom}(\mathcal{I}), \cdot^{\mathcal{I}})$ consists of a non-empty set $\mathsf{Dom}(\mathcal{I})$, called *domain*, and an interpretation function $\cdot^{\mathcal{I}}$ that maps each atomic concept $A \in \mathbf{C}$ to a subset $A^{\mathcal{I}}$ of $\mathsf{Dom}(\mathcal{I})$, each role $r \in \mathbf{R}$ to a binary relation $r^{\mathcal{I}}$ over

$\mathsf{Dom}(\mathcal{I})$, each individual $i \in \mathbf{I}$ to an element $i^{\mathcal{I}}$ of $\mathsf{Dom}(\mathcal{I})$, and each feature $f \in \mathbf{F}$ to a partial function $f^{\mathcal{I}}$ from $\mathsf{Dom}(\mathcal{I})$ to $\mathsf{Dom}(\mathcal{D})$. The interpretation function $\cdot^{\mathcal{I}}$ is extended to compound concepts as follows: $\bot^{\mathcal{I}} := \emptyset$, $\top^{\mathcal{I}} := \mathsf{Dom}(\mathcal{I})$, $\{i\}^{\mathcal{I}} := \{i^{\mathcal{I}}\}$, $(\exists f.\, P)^{\mathcal{I}} := \{\, x \mid x \in \mathsf{Dom}(f^{\mathcal{I}}) \text{ and } f^{\mathcal{I}}(x) \in P^{\mathcal{D}} \,\}$ where $(f_1, \ldots, f_k)^{\mathcal{I}}(x)$ is defined if all $f_j^{\mathcal{I}}(x)$ are defined and then $(f_1, \ldots, f_k)^{\mathcal{I}}(x) := (f_1^{\mathcal{I}}(x), \ldots, f_k^{\mathcal{I}}(x))$, $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$, and $(\exists r.\, C)^{\mathcal{I}} := \{\, x \mid \text{there is } y \text{ s.t. } (x, y) \in r^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}} \,\}$. Role chains are interpreted by $\varepsilon^{\mathcal{I}} := \{\, (x, x) \mid x \in \mathsf{Dom}(\mathcal{I}) \,\}$ and $(R \circ S)^{\mathcal{I}} := \{\, (x, z) \mid \text{there is } y \text{ s.t. } (x, y) \in R^{\mathcal{I}} \text{ and } (y, z) \in S^{\mathcal{I}} \,\}$, and role ranges are interpreted as $\mathsf{Ran}(r)^{\mathcal{I}} := \{\, y \mid \text{there is } x \text{ s.t. } (x, y) \in r^{\mathcal{I}} \,\}$.

$\mathcal{I}$ *satisfies* a concept/role/range inclusion $X \sqsubseteq Y$, written $\mathcal{I} \models X \sqsubseteq Y$, if $X^{\mathcal{I}} \subseteq Y^{\mathcal{I}}$. If $\mathcal{I}$ satisfies all inclusions in a KB $\mathcal{K}$, then $\mathcal{I}$ is a *model* of $\mathcal{K}$, written $\mathcal{I} \models \mathcal{K}$. If $\mathcal{K}$ has a model, then it is *consistent*, and otherwise *inconsistent*. $\mathcal{K}$ *entails* an inclusion $X \sqsubseteq Y$ if $X \sqsubseteq Y$ is satisfied by all models of $\mathcal{K}$, written $\mathcal{K} \models X \sqsubseteq Y$ or $X \sqsubseteq^{\mathcal{K}} Y$, and we then say that X is *subsumed by* Y w.r.t. $\mathcal{K}$. Furthermore, $\mathcal{K}$ *entails* a KB $\mathcal{L}$ if $\mathcal{K}$ entails all inclusions in $\mathcal{L}$, written $\mathcal{K} \models \mathcal{L}$.

A *constraint inclusion* is of the form $\bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$ where Γ and Δ are finite sets of constraints. $\mathcal{I}$ *satisfies* $\bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$, written $\mathcal{I} \models \bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$, if $\bigcap \{\, \alpha^{\mathcal{I}} \mid \alpha \in \Gamma \,\} \subseteq \bigcup \{\, \beta^{\mathcal{I}} \mid \beta \in \Delta \,\}$. Moreover, $\bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$ is *valid*, written $\mathcal{D} \models \bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$, if it is satisfied in all interpretations. It is easy to see that validity is independent of the concepts, roles, and individuals and that it suffices to consider only one domain element. To this end, a *valuation* is a partial function v from $\mathbf{F}$ to $\mathsf{Dom}(\mathcal{D})$, and it *satisfies* $\exists f.\, P$ if $(v(f_1), \ldots, v(f_k)) \in P^{\mathcal{D}}$. Now, $\bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$ is *valid* iff., for each valuation v, if v satisfies all $\alpha \in \Gamma$, then v satisfies some $\beta \in \Delta$.

We say that $\mathcal{D}$ is P-*admissible* if satisfiability of constraint conjunctions as well as validity of constraint inclusions are decidable in polynomial time and, moreover, $\mathcal{D}$ is *convex*, i.e. for each valid constraint inclusion $\bigsqcap \Gamma \sqsubseteq \bigsqcup \Delta$, there is a constraint $\beta \in \Delta$ such that $\bigsqcap \Gamma \sqsubseteq \beta$ is valid. We can use multiple P-admissible concrete domains by forming their disjoint union, which is P-admissible too.

The following P-admissible concrete domains involving numbers are known in the literature:

1. $\mathcal{D}_{\mathbb{Q},\mathsf{diff}}$ with the constraints $f{=}b$, $f{>}b$, $f{-}g{=}b$ for all features f, g and rational numbers $b \in \mathbb{Q}$ [5]. We write $f{=}b$ instead of $\exists f.\, P_{=b}$ where $(P_{=b})^{\mathcal{D}_{\mathbb{Q},\mathsf{diff}}} := \{b\}$, and $f > b$ instead of $\exists f.\, P_{>b}$ where $(P_{>b})^{\mathcal{D}_{\mathbb{Q},\mathsf{diff}}} := \{\, q \mid q \in \mathbb{Q} \text{ and } q > b \,\}$, and $f - g = b$ instead of $\exists f, g.\, P_{+b}$ where $(P_{+b})^{\mathcal{D}_{\mathbb{Q},\mathsf{diff}}} := \{\, (p, q) \mid p, q \in \mathbb{Q} \text{ and } p + b = q \,\}$. Thus, we obtain $(f = b)^{\mathcal{I}} = \{\, x \mid f^{\mathcal{I}}(x) = b \,\}$, $(f > b)^{\mathcal{I}} = \{\, x \mid f^{\mathcal{I}}(x) > b \,\}$, and $(f - g = b)^{\mathcal{I}} = \{\, x \mid f^{\mathcal{I}}(x) - g^{\mathcal{I}}(x) = b \,\}$.
2. $\mathcal{D}_{\mathbb{Q},\mathsf{lin}}$ provides the constraints $A \cdot f = b$ for all rational matrices $A \in \mathbb{Q}^{m \times n}$, feature vectors $f \in \mathbf{F}^m$, and rational vectors $b \in \mathbb{Q}^n$ of compatible arities [14]. We write $A \cdot f = b$ instead of $\exists f.\, P_{A,b}$ where $(P_{A,b})^{\mathcal{D}_{\mathbb{R},\mathsf{lin}}} := \{\, q \mid q \in \mathbb{Q}^m \text{ and } A \cdot q = b \,\}$, and therefore $(A \cdot f = b)^{\mathcal{I}} = \{\, x \mid A \cdot f^{\mathcal{I}}(x) = b \,\}$. There is a similar concrete domain $\mathcal{D}_{\mathbb{R},\mathsf{lin}}$ based on real numbers.
3. There are 24 numerical concrete domains based on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, and with the constraints $f{<}b$, $f{\leq}b$, $f{=}b$, $f{\geq}b$, $f{>}b$ [45]. However, these constraints may not be used arbitrarily. Instead one uses two subsets $\mathbf{P}_{+}$ and $\mathbf{P}_{-}$ of the predicate set $\mathbf{P} := \{\, P_{<b}, P_{\leq b}, P_{=b}, P_{\geq b}, P_{>b} \mid b \in \mathbb{R} \,\}$ consisting of

positive and, respectively, *negative* predicates.[1] Then, a constraint $\exists f.P$ is *positive* if $P \in \mathbf{P}_+$ and *negative* if $P \in \mathbf{P}_-$. KBs may only contain CIs $C \sqsubseteq D$ for which each constraint in C is negative and every constraint in D is positive. Convexity is now only required w.r.t. constraint inclusions of the form $\alpha_1 \sqcap \cdots \sqcap \alpha_m \sqsubseteq \beta_1 \sqcup \cdots \sqcup \beta_n$ where the α_i are positive constraints and the β_j are negative ones. For instance, with $\mathbb{N}$ we could use all constraints negatively but only $f = b$ positively, or all positively but only $f < b$ and $f \leq b$ negatively, among other choices.
It is straight-forward to generalize this to linear systems or regular expressions instead of numerical comparisons. The downside of all this is, however, that reasoning capabilities of the existing procedures are limited and it is unclear how generalize them. For instance, they are still complete for classification but not for subsumption anymore.

3 Hierarchical Concrete Domains

A *semi-lattice* $\mathbf{L} := (L, \leq, \wedge)$ consists of a set L, a partial order $\leq$ on L, and a binary meet operation $\wedge$ on L, i.e. the following hold for all $p, q, p_1, p_2, p_3 \in L$:

(SL1) $p \leq p$ for each $p \in L$ (reflexive)
(SL2) if $p \leq q$ and $q \leq p$, then $p = q$ (anti-symmetric)
(SL3) if $p_1 \leq p_2$ and $p_2 \leq p_3$, then $p_1 \leq p_3$ (transitive)
(SL4) $p_1 \wedge p_2 \leq p_1$ and $p_1 \wedge p_2 \leq p_2$
(SL5) if $q \leq p_1$ and $q \leq p_2$, then $q \leq p_1 \wedge p_2$.

The strict part $<$ is defined by $p < q$ if $p \leq q$ but $q \not\leq p$, and we then say that p is *more specific than* q. Thus $p \leq q$ iff. $p < q$ or $p = q$, in which case we say that p is *more specific than or equal to* q. And $p \wedge q$ is the *meet* of p and q. It follows from the above conditions that $\wedge$ is associative, commutative, and idempotent. The finitary meet operation $\bigwedge$ is obtained from the binary one by setting $\bigwedge\{p\} := p$, $\bigwedge\{p, q\} := p \wedge q$, and $\bigwedge\{p_1, \ldots, p_n\} := p_1 \wedge \bigwedge\{p_2, \ldots, p_n\}$ whenever $n \geq 3$.

We say that $\mathbf{L}$ is *computable* if L and $\leq$ are decidable and $\wedge$ is computable. If all this is possible in polynomial time, then $\mathbf{L}$ is *polynomial-time computable.* $\mathbf{L}$ is *bounded* if it has a greatest element $\top$, i.e. $p \leq \top$ for every $p \in L$. Then we can also define a nullary meet as $\bigwedge \emptyset := \top$. In order to express impossible combinations of values, it might be convenient to add an artificial smallest element $\bot$ to the semi-lattice, i.e. $\bot \leq p$ for each $p \in L$. We then use $\bot$ to represent contradictory or ill-defined values. More specifically, $p \wedge q = \bot$ if it is impossible to combine the values p and q.

Example 1. A semi-lattice representing grades could have the values $\mathsf{Attended}$, Passed, Failed, 1, 2, 3, 4, 5, 6, 1.0, 1.3, 1.7, 2.0, and so on. Its partial order $\leq$ is defined by $\mathsf{Passed} \leq \mathsf{Attended}$, $\mathsf{Failed} \leq \mathsf{Attended}$, $1 \leq \mathsf{Passed}$, $2 \leq \mathsf{Passed}$, $3 \leq \mathsf{Passed}$, $4 \leq \mathsf{Passed}$, $5 \leq \mathsf{Failed}$, $6 \leq \mathsf{Failed}$, $1.0 \leq 1$, $1.3 \leq 1$, $1.7 \leq 2$, $2.0 \leq 2$,

[1] $\mathbf{P}_+$ and $\mathbf{P}_-$ need not be a partitioning of $\mathbf{P}$, they can overlap, they can be equal, or they can be disjoint, and their union need not be the whole of $\mathbf{P}$.

etc. Here we need to add a smallest element $\bot$ since e.g. the meet of grades 1.0 and 5.0 cannot be reasonably defined.

For every KB $\mathcal{K}$ expressed in a decidable DL, the set of all concepts ordered by subsumption $\sqsubseteq^{\mathcal{K}}$ and with conjunction $\sqcap$ as meet operation is a computable, bounded semi-lattice.[2] For each set M, $(\wp(M), \subseteq, \cap, M)$ and $(\wp(M), \supseteq, \cup, \emptyset)$ are bounded semi-lattices. In order to make them computable, it would at least be necessary to restrict them to finite or finitely representable subsets of M. In the following subsections we will introduce several application-relevant semi-lattices based on intervals, polygons, regular languages, and graphs.

Definition 2. *Given a bounded semi-lattice* $\mathbf{L} := (L, \leq, \wedge, \top)$*, the* hierarchical concrete domain $\mathcal{D}_{\mathbf{L}}$ *has values in* $\mathsf{Dom}(\mathcal{D}_{\mathbf{L}}) := L$ *and supports only constraints of the form* $\exists f.\, P_{\leq p}$*, rather written as* $f \leq p$*, involving a feature* f *and a value* p*. The semantics are* $(P_{\leq p})^{\mathcal{D}_{\mathbf{L}}} := \{\, q \mid q \in L \text{ and } q \leq p \,\}$ *and thus* $(f \leq p)^{\mathcal{I}} = \{\, x \mid f^{\mathcal{I}}(x) \leq p \,\}$*. Recall: this means that* f*'s value is* p *or more specific, not smaller. We assume that* $\top$ *stands for an undefined value and thus all valuations are total, i.e.* $v(f) = \top$ *means that* f *has no value under* v*. In order to represent a most general value,* $\mathbf{L}$ *could contain a second-largest element* $\square$*, i.e.* $\square < \top$ *and* $p \leq \square$ *for each* $p \in L \setminus \{\top\}$*. Since* $\bot$ *represents contradictory, ill-defined values, every valuation* v *must not assign* $\bot$ *to any feature* f*, i.e.* $v(f) \neq \bot$*.*

Definition 3. *A* feature inclusion (FI) $f \leq H(g_1, \ldots, g_n)$ *consists of features* $f, g_1, \ldots, g_n$ *and a computable* n*-ary operation* $H\colon L^n \to L$ *that is* monotonic *in the sense that* $H(p_1, \ldots, p_n) \leq H(q_1, \ldots, q_n)$ *whenever* $p_1 \leq q_1, \ldots,$ *and* $p_n \leq q_n$ *(i.e. applying* H *to more specific values yields more specific values). A valuation* v satisfies *this FI if* $v(f) \leq H(v(g_1), \ldots, v(g_n))$*, denoted as* $v \models f \leq H(g_1, \ldots, g_n)$*. An* FBox $\mathcal{F}$ *is a finite set of FIs, and a valuation* v satisfies $\mathcal{F}$*, written* $v \models \mathcal{F}$*, if* v *satisfies every FI in* $\mathcal{F}$*. We call* $\mathcal{F}$ acylic *if the graph* $(\mathbf{F}, \{\, (f, g_1), \ldots, (f, g_n) \mid f \leq H(g_1, \ldots, g_n) \in \mathcal{F} \,\})$ *is, and* cyclic *otherwise.*

The following example illustrates that FIs are "directed specifications" in the sense that values of the right-hand side features $g_1, \ldots, g_n$ yield, through the operation H, an upper bound for the value of the left-hand side feature f. However, this does not work in the other direction unless specified by other FIs.

Example 4. We use three features with interval values over the non-negative integers: sys for the systolic and dia for the diastolic blood pressure, and pp for the pulse pressure, which is the difference between the systolic and the diastolic pressure. The FI $\mathsf{pp} \subseteq \mathsf{sys} - \mathsf{dia}$ allows us to infer a value for pp when values for both sys and dia are given. For instance, under this FI the constraint inclusion $(\mathsf{sys} \subseteq [110, 120]) \sqcap (\mathsf{dia} \subseteq [60, 70]) \sqsubseteq (\mathsf{pp} \subseteq [40, 60])$ is valid. In contrast, the constraint inclusion $(\mathsf{sys} \subseteq [110, 120]) \sqcap (\mathsf{pp} \subseteq [40, 60]) \sqsubseteq (\mathsf{dia} \subseteq [60, 70])$ is not valid w.r.t. the above FI. A countervaluation is v with $v(\mathsf{sys}) = [110, 120]$, $v(\mathsf{dia}) = [0, \infty)$, $v(\mathsf{pp}) = [40, 60]$. This is because $[110, 120] - [0, \infty) = [0, 120]$ and $[40, 60] \subseteq [0, 120]$, i.e. v satisfies the FI, but v does not satisfy the latter constraint inclusion.

[2] More precisely, this holds for the set of all equivalence classes of concepts, i.e. all sets of the form $\{\, D \mid C \sqsubseteq^{\mathcal{K}} D \text{ and } D \sqsubseteq^{\mathcal{K}} C \,\}$ for all concepts C.

Definition 5. *The semantics of the concrete domain $\mathcal{D}_{\mathbf{L}}$ can be restricted w.r.t. an FBox $\mathcal{F}$ by considering only valuations satisfying $\mathcal{F}$. That is, a constraint inclusion $\bigsqcap\Gamma \sqsubseteq \bigsqcup\Delta$ is* valid *in $\mathcal{D}_{\mathbf{L}}$ w.r.t. $\mathcal{F}$, written $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap\Gamma \sqsubseteq \bigsqcup\Delta$, if this inclusion is satisfied in all valuations that satisfy $\mathcal{F}$. Whenever we write "w.r.t. $\mathcal{F}$" in the following, only valuations satisfying $\mathcal{F}$ are considered.*

Using this semantics restricted by an FBox, convexity and P-admissibility are defined as before but the latter additionally takes the FBox $\mathcal{F}$ as part of the input. The underlying semi-lattice $\mathbf{L}$ is taken into account through the computational complexity of its value set L, its partial order $\leq$, and its meet operation $\wedge$.

Definition 6. *$\mathcal{D}_{\mathbf{L}}$ is* admissible *w.r.t. $\mathcal{F}$ if $\mathcal{D}_{\mathbf{L}}$ is convex and satisfiability of constraint conjunctions as well as validity of constraint inclusions are decidable, all w.r.t. $\mathcal{F}$. For a complexity class C, we say that $\mathcal{D}_{\mathbf{L}}$ is* C-admissible *w.r.t. $\mathcal{F}$ if, all w.r.t. $\mathcal{F}$, $\mathcal{D}_{\mathbf{L}}$ is convex and satisfiability of constraint conjunctions as well as validity of constraint inclusions are in C when $\mathcal{F}$ is part of the input.*

Next, we show that a hierarchical concrete domain $\mathcal{D}_{\mathbf{L}}$ is convex w.r.t. $\mathcal{F}$ if the semi-lattice $\mathbf{L}$ is complete or well-founded or if the FBox $\mathcal{F}$ is acyclic. Note that every finite semi-lattice is well-founded, i.e. convexity is guaranteed when a non-acyclic FBox is used with only finitely many values. Convexity is also ensured over non-well-founded semi-lattices when the FBox is empty (since it is acyclic). There might be further conditions that ensure convexity even if $\mathbf{L}$ is neither complete nor well-founded and $\mathcal{F}$ is not acyclic; we leave this for future research.

Definition 7. *Let $\mathbf{L}$ be a bounded semi-lattice and $\mathcal{F}$ be an FBox. Given a finite set Γ of constraints over the concrete domain $\mathcal{D}_{\mathbf{L}}$, a* canonical valuation *of Γ w.r.t. $\mathcal{F}$ is a valuation $v_{\Gamma,\mathcal{F}}$ such that*

1. *$v_{\Gamma,\mathcal{F}} \models \mathcal{F}$ and*
2. *$v_{\Gamma,\mathcal{F}} \models \alpha$ iff. $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap\Gamma \sqsubseteq \alpha$ for each constraint α.*

Moreover, we say that $\mathcal{D}_{\mathbf{L}}$ has canonical valuations w.r.t. $\mathcal{F}$ if such a valuation $v_{\Gamma,\mathcal{F}}$ exists for every finite, w.r.t. $\mathcal{F}$ satisfiable Γ.

Since for each constraint α in Γ, the inclusion $\bigsqcap\Gamma \sqsubseteq \alpha$ is valid, we infer with the second condition that $v_{\Gamma,\mathcal{F}}$ satisfies Γ.

Lemma I. *Let $\mathbf{L}$ be a bounded semi-lattice and $\mathcal{F}$ be an FBox. $\mathcal{D}_{\mathbf{L}}$ is convex w.r.t. $\mathcal{F}$ if it has canonical valuations w.r.t. $\mathcal{F}$.*

Proof. Assume that $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap\Gamma \sqsubseteq \bigsqcup\Delta$. Since $v_{\Gamma,\mathcal{F}} \models \mathcal{F}$ and $v_{\Gamma,\mathcal{F}} \models \bigsqcap\Gamma$, it follows that $v_{\Gamma,\mathcal{F}} \models \bigsqcup\Delta$, i.e. $v_{\Gamma,\mathcal{F}} \models \alpha$ for some $\alpha \in \Delta$. We conclude that $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap\Gamma \sqsubseteq \alpha$. □

A semi-lattice $\mathbf{L}$ is *complete* if every subset $P \subseteq L$ has a meet $\bigwedge P \in L$, i.e. such that $\bigwedge P \leq p$ for each $p \in P$ and, if $q \leq p$ for each $p \in P$, then $q \leq \bigwedge P$. Note that these two conditions generalize (SL4) and (SL5). Every complete semi-lattice is a complete lattice since we can obtain the join operation by $\bigvee P := \bigwedge\{\, q \mid p < q \text{ for each } p \in P \,\}$.

Theorem 8. *For each complete semi-lattice* $\mathbf{L}$ *and for every FBox* $\mathcal{F}$*, the concrete domain* $\mathcal{D}_{\mathbf{L}}$ *has canonical valuations and so is convex w.r.t.* $\mathcal{F}$*.*

Proof. Completeness of $\mathbf{L}$ implies that $\mathbf{L}$ is also a complete lattice. It follows that $L^{\mathbf{F}}$ is a complete lattice as well when equipped with the pointwise lifting of $\leq$, i.e. $v_1 \leq v_2$ iff. $v_1(f) \leq v_2(f)$ for each $f \in \mathbf{F}$.

The FBox $\mathcal{F}$ induces the function $\Phi_{\mathcal{F}} \colon L^{\mathbf{F}} \to L^{\mathbf{F}}$ that sends every assignment $v \colon \mathbf{F} \to L$ to the assignment $\Phi_{\mathcal{F}}(v) \colon \mathbf{F} \to L$ where $\Phi_{\mathcal{F}}(v)(f) \coloneqq v(f) \wedge \bigwedge\{\, H(v(g_1), \ldots, v(g_m)) \mid f \leq H(g_1, \ldots, g_m) \in \mathcal{F} \,\}$.

Since all operations H occurring in $\mathcal{F}$ are monotonic, also $\Phi_{\mathcal{F}}$ is monotonic. To see this, consider two valuations with $v_1 \leq v_2$ (pointwise) and let $f \in \mathbf{F}$ be a feature. Then $v_1(f) \leq v_2(f)$, and $v_1(g_i) \leq v_2(g_i)$ for each FI $f \leq H(g_1, \ldots, g_m) \in \mathcal{F}$ and each $i \in \{1, \ldots, m\}$. Monotonicity of each involved H yields $H(v_1(g_1), \ldots, v_1(g_m)) \leq H(v_2(g_1), \ldots, v_2(g_m))$. Thus, $\Phi_{\mathcal{F}}(v_1)(f) \leq \Phi_{\mathcal{F}}(v_2)(f)$. Since f is arbitrary, we conclude that $\Phi_{\mathcal{F}}(v_1) \leq \Phi_{\mathcal{F}}(v_2)$ (pointwise).

It is easy to see that the fixed points of $\Phi_{\mathcal{F}}$ are exactly the satisfying valuations of $\mathcal{F}$ (ignoring for now that some might map features to $\bot$), i.e. $\Phi_{\mathcal{F}}(v) = v$ iff. $v \models \mathcal{F}$:

v is a fixed point of $\Phi_{\mathcal{F}}$
iff. $v = \Phi_{\mathcal{F}}(v)$
iff. $v(f) = \Phi_{\mathcal{F}}(v)(f)$ for every feature f
iff. $v(f) = v(f) \wedge \bigwedge\{\, H(v(g_1), \ldots, v(g_m)) \mid f \leq H(g_1, \ldots, g_m) \in \mathcal{F} \,\}$ for every feature f
iff. $v(f) \leq \bigwedge\{\, H(v(g_1), \ldots, v(g_m)) \mid f \leq H(g_1, \ldots, g_m) \in \mathcal{F} \,\}$ for every feature f
iff. $v(f) \leq H(v(g_1), \ldots, v(g_m))$ for each FI $f \leq H(g_1, \ldots, g_m) \in \mathcal{F}$
iff. v is a satisfying valuation of $\mathcal{F}$.

Note that $\bigwedge \emptyset = \top$, i.e. the third-last line is trivially satisfied for all features not occurring as left-hand side of a FI in $\mathcal{F}$.

Now, the Knaster-Tarski Theorem [52] yields existence of a greatest fixed point $v_{\Gamma,\mathcal{F}} \colon \mathbf{F} \to L$ among all fixed points of $\Phi_{\mathcal{F}}$ that are pointwise more specific than or equal to $v_\Gamma \colon \mathbf{F} \to L$ where $v_\Gamma(f) \coloneqq \bigwedge\{\, p \mid (f \leq p) \in \Gamma \,\}$ for all f.

Obviously, we have $w \leq v_\Gamma$ iff. w is a satisfying valuation of Γ. If $v_{\Gamma,\mathcal{F}}(f) = \bot$ for some feature f, then we conclude that $w(f) = \bot$ for every valuation w satisfying $\mathcal{F}$ and Γ, i.e. there are no such valuations and thus Γ is unsatisfiable. Otherwise, $v_{\Gamma,\mathcal{F}}$ is a valuation and it remains to verify that $v_{\Gamma,\mathcal{F}}$ is canonical as per Definition 7. Convexity then follows by Lemma I.

1. We have seen above that $\Phi_{\mathcal{F}}(v) = v$ iff. $v \models \mathcal{F}$, and thus $v_{\Gamma,\mathcal{F}}$ satisfies $\mathcal{F}$.
2. $v_{\Gamma,\mathcal{F}}$ satisfies all constraints in Γ since $v_{\Gamma,\mathcal{F}} \leq v_\Gamma$. The if direction is therefore already shown. Regarding the only-if direction, assume $v_{\Gamma,\mathcal{F}} \models (g \leq q)$ and consider a valuation w such that $w \models \mathcal{F}$ and $w \models \bigsqcap \Gamma$. It follows that $v_{\Gamma,\mathcal{F}}(g) \leq q$, $\Phi_{\mathcal{F}}(w) = w$, and $w \leq v_\Gamma$. Since $v_{\Gamma,\mathcal{F}}$ is the greatest fixed point $\leq v_\Gamma$, we have $w \leq v_{\Gamma,\mathcal{F}}$ and thus $w(g) \leq q$.

It follows that Γ is satisfiable iff. $v_{\Gamma,\mathcal{F}}(f) \neq \bot$ for every feature f. □

Theorem 9. *Let* $\mathbf{L}$ *be a computable, bounded semi-lattice and* $\mathcal{F}$ *be an FBox. If* $\mathbf{L}$ *is well-founded or* $\mathcal{F}$ *is acyclic, then the concrete domain* $\mathcal{D}_{\mathbf{L}}$ *has computable canonical valuations and is admissible w.r.t.* $\mathcal{F}$*.*

Proof. Given a finite set Γ of constraints over $\mathcal{D}_{\mathbf{L}}$, we construct a mapping $v_{\Gamma,\mathcal{F}}$ as follows.

- First, we define a mapping $v_0 \colon \mathbf{F} \to L$ by $v_0(f) \coloneqq \bigwedge\{\, p \mid (f \leq p) \in \Gamma \,\}$ for every feature f, and set $i \coloneqq 0$.
- While there is an FI $f \leq H(g_1, \ldots, g_n)$ in $\mathcal{F}$ such that $v_i(f) \not\leq H(v_i(g_1), \ldots, v_i(g_n))$, we initialize the next mapping $v_{i+1} \colon \mathbf{F} \to L$ by $v_i \coloneqq v_{i+1}$ but set $v_{i+1}(f) \coloneqq v_i(f) \wedge H(v_i(g_1), \ldots, v_i(g_n))$, and increase i. Otherwise, we terminate the while-loop and define $v_{\Gamma,\mathcal{F}} \coloneqq v_i$.

Since $\mathbf{L}$ is computable, each single step in the above procedure requires only a finite amount of time. It is easy to see that the while-loop terminates if the semi-lattice $\mathbf{L}$ is well-founded. Now assume that $\mathcal{F}$ is acyclic. We define a "before" relation between FIs by $(f \leq H(g_1, \ldots, g_n))$ "before" $(f' \leq H'(g'_1, \ldots, g'_n))$ if $f \in \{g'_1, \ldots, g'_n\}$. Then let $\prec$ be the transitive reduction (neighborhood relation) of an arbitrary linearization of this "before" relation.[3] During the above while-loop we now go along $\prec$, and thus we are done after polynomially many steps (w.r.t. $\mathcal{F}$).

The returned mapping $v_{\Gamma,\mathcal{F}}$ might assign $\bot$ to features and thus might not be a valuation. We ignore this for the time being.

$v_{\Gamma,\mathcal{F}}$ satisfies $\mathcal{F}$ since it is obtained as the last valuation v_i upon termination of the while-loop, i.e. when v_i satisfies all FIs in $\mathcal{F}$. Moreover, by construction $v_0(f) \leq p$ for each constraint $f \leq p$ in Γ and further $v_0 \geq v_1 \geq v_2 \geq \cdots \geq v_{\Gamma,\mathcal{F}}$, which yields $v_{\Gamma,\mathcal{F}}(f) \leq v_0(f) \leq p$ and thus $v_{\Gamma,\mathcal{F}}$ satisfies Γ.

Next, we show that the above procedure has an invariant: $w \leq v_i$ (pointwise) for each valuation w such that $w \models \mathcal{F}$ and $w \models \bigsqcap \Gamma$. In the end, $w \leq v_{\Gamma,\mathcal{F}}$ (pointwise).

- Since w satisfies Γ, we have $w(f) \leq p$ for every constraint $f \leq p$ in Γ, and thus $w(f) \leq v_0(f)$ for each feature f, i.e. $w \leq v_0$.
- Assume $w \leq v_i$ and let $f \leq H(g_1, \ldots, g_n)$ be the FI not satisfied by v_i and used to obtain v_{i+1}. Since w satisfies $\mathcal{F}$, $w(f) \leq H(w(g_1), \ldots, w(g_n))$. The assumption that $w \leq v_i$ yields that $w(g_1) \leq v_i(g_1)$, $\ldots$, $w(g_n) \leq v_i(g_n)$ and thus $H(w(g_1), \ldots, w(g_n)) \leq H(v_i(g_1), \ldots, v_i(g_n))$ as H is monotonic. The assumption further yields that $w(f) \leq v_i(f)$. It follows that $w(f) \leq v_i(f) \wedge H(v_i(g_1), \ldots, v_i(g_n)) = v_{i+1}(f)$. For every other feature $g \neq f$ we have $w(g) \leq v_i(g) = v_{i+1}(g)$. In the end, $w \leq v_{i+1}$.

Now, if $v_{\Gamma,\mathcal{F}}(f) = \bot$ for some feature f, then we conclude from the above invariant that $w(f) = \bot$ for every valuation w satisfying $\mathcal{F}$ and Γ, i.e. there are

[3] Given a partial order $\leq$, its transitive reduction is $\leq \setminus (\leq \circ \leq)$, i.e. the set of all pairs $(x, y) \in \leq$ such that there is no z with $(x, z) \in \leq$ and $(z, y) \in \leq$. Moreover, a linearization of $\leq$ is a superset that is also a partial order but in which each two elements are comparable, i.e. it contains either (x, y) or (y, x) for each two x, y.

no such valuations and thus Γ is unsatisfiable. Otherwise, $v_{\Gamma,\mathcal{F}}$ is a valuation and it remains to verify that $v_{\Gamma,\mathcal{F}}$ is canonical as per Definition 7. Convexity then follows by Lemma I.

1. We have already seen above that $v_{\Gamma,\mathcal{F}}$ satisfies $\mathcal{F}$.
2. Given a constraint $g \leq q$, we must show that $v_{\Gamma,\mathcal{F}} \models (g \leq q)$ iff. $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap \Gamma \sqsubseteq (g \leq q)$. The if direction holds since $v_{\Gamma,\mathcal{F}} \models \mathcal{F}$ and $v_{\Gamma,\mathcal{F}} \models \bigsqcap \Gamma$. Assume $v_{\Gamma,\mathcal{F}} \models (g \leq q)$ and consider a valuation w such that $w \models \mathcal{F}$ and $w \models \bigsqcap \Gamma$. The former yields $v_{\Gamma,\mathcal{F}}(g) \leq q$ and the latter yields $w \leq v_{\Gamma,\mathcal{F}}$ (pointwise) by the invariant. In particular $w(g) \leq v_{\Gamma,\mathcal{F}}(g)$, and thus $w(g) \leq q$, i.e. $w \models (g \leq q)$ as required.

It follows that Γ is satisfiable iff. $v_{\Gamma,\mathcal{F}}(f) \neq \bot$ for every feature f. Since we obtain $v_{\Gamma,\mathcal{F}}$ in finite time, satisfiability of constraint conjunctions is decidable.

Through Condition 2 in Definition 7 we can decide validity of constraint inclusion $\bigsqcap \Gamma \sqsubseteq \alpha$ where $\alpha := (g \leq q)$. To this end, we first compute $v_{\Gamma,\mathcal{F}}$ by means of the above procedure, then check if $v_{\Gamma,\mathcal{F}}(f) \neq \bot$ for each f (i.e. Γ is satisfiable and $v_{\Gamma,\mathcal{F}}$ is its canonical valuation), and finally check if $v_{\Gamma,\mathcal{F}}(g) \leq q$ (i.e. $v_{\Gamma,\mathcal{F}}$ satisfies α), which can all be done in finite time. $\square$

Now, we want to determine the time requirement for computing a canonical valuation $v_{\Gamma,\mathcal{F}}$, which is measured w.r.t. the constraint set Γ and the FBox $\mathcal{F}$. The semi-lattice $\mathbf{L}$ is only taken into account through the decision and computation procedures for its value set L, partial order $\leq$, and meet operation $\wedge$.

Proposition 10. *Consider a polynomial-time computable, bounded semi-lattice* $\mathbf{L}$ *such that its meet operation returns values of linear size. Further consider an acyclic FBox* $\mathcal{F}$ *in which all occurring operations are polynomial-time computable and return values of linear size. W.r.t.* $\mathcal{F}$*, the concrete domain* $\mathcal{D}_{\mathbf{L}}$ *has polynomial-time computable canonical valuations and is* P*-admissible.*

Proof. We have already seen in the proof of Theorem 9 that the while-loop of the procedure there needs only one iteration per FI in the acyclic FBox $\mathcal{F}$. Since $\mathbf{L}$ is polynomial-time computable and every operation occurring in $\mathcal{F}$ is polynomial-time computable, each single iteration requires only polynomial time w.r.t. its respective input (which is the intermediate assignment v_i and the FBox $\mathcal{F}$). Moreover since the meet operation of $\mathbf{L}$ and each operation in $\mathcal{F}$ return linear-size values, all intermediate assignments v_i have linear size w.r.t. the input (which is the constraint set Γ and the FBox $\mathcal{F}$). It follows that the canonical valuation $v_{\Gamma,\mathcal{F}}$ can be computed in polynomial time. $\square$

The following example shows that Proposition 10 need not hold when the FBox $\mathcal{F}$ contains an operation computable in polynomial time but not returning values of linear size, basically because the size increases can accumulate to an exponential size.

Example II. We consider words over the unary alphabet, say with letter a, partially ordered by equality $=$. The acyclic FBox $\mathcal{F} := \{\, f_{i+1} = H(f_i) \mid$

$i \in \{0, \ldots, n-1\}$ } uses the operation H where $H(w) \coloneqq w \circ w$. Obviously, H is computable in quadratic time and its outputs have quadratic size. Now, for the constraint set $\Gamma \coloneqq \{f_0 = a\}$ we obtain the canonical valuation $v_{\Gamma,\mathcal{F}}$ with $v_{\Gamma,\mathcal{F}}(f_i) = a^{(2^i)}$, which has exponential size and thus cannot be computed in polynomial time.

A further example shows that already the constraint set Γ could enforce a canonical valuation not computable in polynomial time if the meet operation does not return linear-size values.

Example III. Take the semi-lattice consisting of all positive integers and partially ordered by the "is divided by" relation (denoted as $|^{-1}$). Its meet operation yields the least common multiple. Given an increasing enumeration $p_1, p_2, \ldots$ of all primes, the constraint set $\Gamma \coloneqq \{f \,|^{-1}\, p_1, \ldots, f \,|^{-1}\, p_n\}$ has a canonical valuation $v_{\Gamma,\mathcal{F}}$ where $v_{\Gamma,\mathcal{F}}(f) = p_1 \cdot \cdots \cdot p_n$. The size of $v_{\Gamma,\mathcal{F}}(f)$ is exponential in the size of Γ.

Without the assumption that all involved operations yield linear-size results, with similar arguments as for Proposition 10 we obtain exponential complexity.

Proposition 11. *For every polynomial-time computable, bounded semi-lattice* $\mathbf{L}$ *and for every acyclic FBox* $\mathcal{F}$ *in which all occurring operations are polynomial-time computable, the concrete domain* $\mathcal{D}_{\mathbf{L}}$ *has exponential-time computable canonical valuations and is* EXP*-admissible w.r.t.* $\mathcal{F}$*.*

3.1 Intervals

Let N be a non-empty set of real numbers. The semi-lattice $\mathbf{Int}(N)$ consists of all intervals over N, is partially ordered by set inclusion $\subseteq$ and has set intersection $\cap$ as its meet operation. All types of intervals are supported, such as closed intervals $[p, q] \coloneqq \{\, o \mid p \leq o \leq q \,\}$, $[p, +\infty) \coloneqq \{\, o \mid p \leq o \,\}$, $(-\infty, q] \coloneqq \{\, o \mid o \leq q \,\}$, $(-\infty, +\infty) \coloneqq N$, open intervals (p, q), $(p, +\infty)$, $(-\infty, q)$, $(-\infty, +\infty)$ defined with $<$ instead of $\leq$, and also half-open intervals $(p, q]$, $[p, q)$. $\mathbf{Int}(N)$ is already bounded since its greatest element is $N = (-\infty, \infty)$, but we rather identify it with $\Box$ and add an artificial greatest element $\top$. It also has a smallest element $\emptyset = (p, p)$ where $p \in N$ is arbitrary, and we identify this smallest element with the contradictory value $\bot$.

The hierarchical concrete domain $\mathcal{D}_{\mathbf{Int}(N)}$ is called the *interval domain* over N. Since for every number $p \in N$, the singleton $\{p\}$ equals the interval $[p, p]$, we can specify the precise numerical value of a feature with the constraint $f \subseteq \{p\}$, also written $f = p$. Moreover, instead of $f \subseteq [p, q]$ we may also write $p \leq f \leq q$.

Example 12. Through the interval domain over the non-negative 8-bit integers $N \coloneqq \mathbb{N} \cap [0, 2^8 - 1]$ we could express non-elevated blood pressure by $\mathsf{NonElevatedBP} \equiv (\mathsf{sys} \subseteq [0, 120)) \sqcap (\mathsf{dia} \subseteq [0, 70))$, elevated blood pressure by $\mathsf{ElevatedBP} \equiv (\mathsf{sys} \subseteq [120, 140)) \sqcap (\mathsf{dia} \subseteq [70, 90))$, and hypertension by $(\mathsf{sys} \subseteq [140, \infty)) \sqsubseteq \mathsf{Hypertension}$ and $(\mathsf{dia} \subseteq [90, \infty)) \sqsubseteq \mathsf{Hypertension}$. With the above syntactic sugar, the first statement can also be written as $\mathsf{NonElevatedBP} \equiv (0 \leq \mathsf{sys} < 120) \sqcap (0 \leq \mathsf{dia} < 70)$, and similarly for the other two. The concrete values of patient bob can be represented by the assertions $\mathsf{bob} : (\mathsf{sys} = 114)$ and $\mathsf{bob} : (\mathsf{dia} \subseteq [69, 69])$. The KB consisting of all these aforementioned statements entails $\mathsf{bob} : \mathsf{NonElevatedBP}$.

Each binary operation $*$ on N can be lifted to a binary operation on intervals by $[p_1, q_1] * [p_2, q_2] \coloneqq \{\, o_1 * o_2 \mid o_1 \in [p_1, q_1] \text{ and } o_2 \in [p_2, q_2] \,\}$, and similarly for other types of intervals. If $*$ is continuous on a domain containing $[p_1, q_1] \times [p_2, q_2]$, then the resulting set $[p_1, q_1] * [p_2, q_2]$ is also an interval. Moreover, if $*$ is monotonic, then $[p_1, q_1] * [p_2, q_2] = [\mathsf{min}(S), \mathsf{max}(S)]$ where $S \coloneqq \{p_1 * p_2,\ p_1 * q_2,\ q_1 * p_2,\ q_1 * q_2\}$ [28]. For instance, addition $+$, subtraction $-$, and multiplication $\cdot$ are monotonic. We have $[p_1, q_1] + [p_2, q_2] = [p_1 + p_2, q_1 + q_2]$ as well as $[p_1, q_1] - [p_2, q_2] = [p_1, q_1] + [-q_2, -p_2] = [p_1 - q_2, q_1 - p_2]$. Products can be computed without min and max if none of the intervals contains 0 as an interior point. For instance, $[p_1, q_1] \cdot [p_2, q_2] = [p_1 \cdot p_2, q_1 \cdot q_2]$ if all interval bounds are non-negative. Division is technically more involved since one needs to distinguish if the second interval contains 0 or has 0 as an endpoint. We have

- $[p_1, q_1]/[p_2, q_2] = [p_1, q_1] \cdot [1/q_2, 1/p_2]$ if $0 \notin [p_2, q_2]$,
- $[p_1, q_1]/[p_2, 0] = [p_1, q_1] \cdot (-\infty, 1/p_2]$,
- $[p_1, q_1]/[0, q_2] = [p_1, q_1] \cdot [1/q_2, +\infty)$, and
- $[p_1, q_1]/[q_1, q_2] = [p_1, q_1] \cdot ((-\infty, 1/p_2] \cup [1/q_2, +\infty))$ if $0 \in [p_2, q_2]$ but $p_2 \neq 0 \neq q_2$.

In the last case the result is a union of two intervals. In order to support such results, the semi-lattice $\mathbf{Int}(N)$ needs to be replaced by the semi-lattice $\mathbf{UInt}(N)$ consisting of all finite unions of pairwise separated[4] intervals over N. It is also polynomial-time computable, but it is currently unclear w.r.t. which FBoxes $\mathcal{F}$ the concrete domain $\mathcal{D}_{\mathbf{UInt}(N)}$ is P-admissible. Inclusion of such interval unions can be decided in polynomial time since $P_1 \cup \cdots \cup P_m \subseteq Q_1 \cup \cdots \cup Q_n$ iff., for each $i \in \{1, \ldots, m\}$, there is $j \in \{1, \ldots, n\}$ such that $P_i \subseteq Q_j$. Disjunctions cannot be emulated by the use of finite unions of intervals since, for instance, the constraint inclusion $(f \subseteq [0, 1] \cup [2, 3]) \sqsubseteq (f \subseteq [0, 1]) \sqcup (f \subseteq [2, 3])$ is not valid in $\mathcal{D}_{\mathbf{UInt}(N)}$ where $N \coloneqq \mathbb{N} \cap [0, 2^8 - 1]$. For the sake of brevity and clarity we do not go into further details here.

Lemma IV. *For each binary operation $*$ on numbers, the lifted operation $*$ on intervals is monotonic, i.e. can be used in FIs.*

Proof. Consider intervals P, P', Q, Q' such that $P \subseteq Q$ and $P' \subseteq Q'$. We have $P * P' = \{\, p * p' \mid p \in P \text{ and } p' \in P' \,\}$ by definition. The assumption yields that the latter set is contained in $\{\, q * q' \mid q \in Q \text{ and } q' \in Q' \,\}$, which by definition equals $Q * Q'$. That is, $P * P' \subseteq Q * Q'$. □

[4] Two intervals are *separated* if each is disjoint with the other's closure. For instance, $[0, 1)$ and $(1, 2]$ are separated, but $[0, 1]$ and $(1, 2]$ are not.

Example 13. Continuing Example 4, we can additionally consider the two FIs $\mathsf{dia} \subseteq \mathsf{sys} - \mathsf{pp}$ and $\mathsf{sys} \subseteq \mathsf{dia} + \mathsf{pp}$, which allow us to also infer interval values of dia and sys given interval values of the respective other two. Importantly, this does not destroy convexity.

This is in stark contrast to the concrete domain extending $\mathcal{D}_{\mathbb{Q},\mathsf{diff}}$ with constraints $f \geq b$, $f < b$, $f \leq b$, which allows to express interval values as well (in a different way though). There, the constraint inclusion $(\mathsf{sys} - \mathsf{dia} = 40) \sqsubseteq (\mathsf{sys} \leq 120) \sqcup (\mathsf{dia} > 80)$ is valid, violating convexity. Additionally using the expressivity of $\mathcal{D}_{\mathbb{Q},\mathsf{lin}}$, we could express that $\mathsf{pp} = \mathsf{sys} - \mathsf{dia}$ by the CI $\top \sqsubseteq (\mathsf{sys} - \mathsf{dia} - \mathsf{pp} = 0)$ as in Example 3 in [2]. Under this CI, the constraint inclusion $(\mathsf{pp} = 40) \sqsubseteq (\mathsf{sys} \leq 120) \sqcup (\mathsf{dia} > 80)$ would be valid, also violating convexity.

In our interval domain over the non-negative integers and with the cyclic FBox $\{\mathsf{pp} \subseteq \mathsf{sys} - \mathsf{dia}, \mathsf{dia} \subseteq \mathsf{sys} - \mathsf{pp}, \mathsf{sys} \subseteq \mathsf{dia} + \mathsf{pp}\}$, the similar constraint inclusion $(\mathsf{pp} \subseteq [40, 40]) \sqsubseteq (\mathsf{sys} \subseteq [0, 120]) \sqcup (\mathsf{dia} \subseteq (80, \infty))$ is not valid. A countervaluation is v where $v(\mathsf{sys}) = [40, \infty)$, $v(\mathsf{dia}) = [0, \infty)$, $v(\mathsf{pp}) = [40, 40]$. It satisfies the first FI since $[40, \infty) - [0, \infty) = [0, \infty) \supseteq [40, 40]$, the second FI since $[40, \infty) - [40, 40] = [0, \infty) \supseteq [0, \infty)$, and the third FI since $[0, \infty) + [40, 40] = [40, \infty) \supseteq [40, \infty)$.

Recall that the interval semi-lattice $\mathbf{Int}(N)$ is defined for every non-empty set N of real numbers. The set N is partially ordered by the usual ordering $\leq$ and has the meet operation min, i.e. $(N, \leq, \mathsf{min})$ is itself a semi-lattice. It thus makes sense to say that N is complete. The real numbers $\mathbb{R}$, the non-negative real numbers $\mathbb{R}_+$, all closed intervals over $\mathbb{R}$, the integers $\mathbb{Z}$, the natural numbers $\mathbb{N}$, the n-bit integers, the n-bit floating-point numbers, the n-bit fixed-point numbers, and all finite subsets of $\mathbb{R}$ are complete, but the rational numbers $\mathbb{Q}$ is not — for instance, the infimum of $\{ (1+1/n)^{n+1} \mid n \geq 0 \}$ is Euler's number e, an irrational number. It is easy to see that the semi-lattice $\mathbf{Int}(N)$ is complete if the number set N is complete, and so we obtain the below corollary to Theorem 8.

Corollary 14. *If the semi-lattice $(N, \leq, \mathsf{min})$ is complete, then the interval domain $\mathcal{D}_{\mathbf{Int}(N)}$ has canonical valuations and is convex w.r.t. every FBox $\mathcal{F}$.*

Proof. If N is complete, i.e. every subset $P \subseteq N$ has an infimum $\bigwedge P \in N$ and thus also a supremum $\bigvee P \in N$, then the interval semi-lattice $\mathbf{Int}(N)$ is complete as well. We have $\bigcap_{t \in T} \langle_t p_t, q_t \rangle_t = \langle p, q \rangle$ where

- $p := \bigvee_{t \in T} p_t$,
- $q := \bigwedge_{t \in T} q_t$,
- if $p \in \langle_t p_t, q_t \rangle_t$ for each $t \in T$, then $\langle\ := [$, else $\langle\ := ($, and
- if $q \in \langle_t p_t, q_t \rangle_t$ for each $t \in T$, then $\rangle :=]$, else $\rangle :=)$.

In particular, the intersection of closed intervals is a closed interval, but the intersection of open intervals need not be open, e.g. $\bigcap_{n \in \mathbb{N}} (-1/n, 1) = [0, 1)$. The claim now follows from Theorem 8. □

An immediate consequence of Theorem 9 is that the interval domain $\mathcal{D}_{\mathbf{Int}(\mathbb{R})}$ over all real numbers is admissible w.r.t. every acyclic FBox. Moreover, an obvious corollary to Proposition 10 is as follows.

Corollary 15. *W.r.t. each acyclic FBox $\mathcal{F}$ in which all operations are polynomial-time computable and yield linear-size results, the interval domain $\mathcal{D}_{\mathbf{Int}(\mathbb{R})}$ has polynomial-time-computable canonical valuations and is* P*-admissible.*

Next, we employ linear programming to handle affine FBoxes, which might be cyclic. We call an FBox $\mathcal{F}$ *affine* if all operations in FIs in $\mathcal{F}$ are affine, i.e. all FIs are of the form $f \subseteq \sum_{i=1}^{n} P_i \cdot g_i + Q_i$ where the P_i and Q_i are intervals. For instance, the FI $\mathsf{pp} \subseteq \mathsf{sys} - \mathsf{dia}$ is affine, but $\mathsf{bmi} \subseteq \mathsf{bodyMass}/\mathsf{bodyHeight}^2$ is not. Since each affine FI represents two linear inequalities (one for the lower bound of the interval value of f, and another one for the upper bound), we can transform affine FBoxes into linear programs, which can be solved in polynomial time [31]. We thus obtain the following result.

Proposition 16. *Let $\underline{c}, \overline{c} \in \mathbb{R}_+$ be non-negative real numbers such that $\underline{c} \leq \overline{c}$. Restricted to closed intervals only, the interval domain $\mathcal{D}_{\mathbf{Int}([\underline{c},\overline{c}])}$ over the non-negative real numbers between $\underline{c}$ and $\overline{c}$ is* P*-admissible w.r.t. each affine FBox $\mathcal{F}$, i.e. all FIs are of the form $f \subseteq \sum_{i=1}^{n} [\underline{a}_i, \overline{a}_i] \cdot g_i + [\underline{b}, \overline{b}]$.*

Proof. Since $[\underline{c}, \overline{c}]$ is complete, Theorem 8 and Corollary 14 yield that $\mathcal{D}_{\mathbf{Int}([\underline{c},\overline{c}])}$ has canonical valuations and is convex w.r.t. every FBox $\mathcal{F}$. Now fix an affine FBox $\mathcal{F}$ as well as a constraint set Γ. We have seen in the proof of Theorem 8 that $w \subseteq v_{\Gamma,\mathcal{F}}$ for each valuation w satisfying Γ and $\mathcal{F}$, where $v_{\Gamma,\mathcal{F}}$ is the canonical valuation.

It remains to show that we can decide satisfiability of Γ w.r.t. $\mathcal{F}$ in polynomial time and compute the canonical valuation $v_{\Gamma,\mathcal{F}}$ in polynomial time. With similar arguments as at the end of the proof of Theorem 9, it then follows that validity of constraint inclusions w.r.t. $\mathcal{F}$ is decidable in polynomial time.

To this end, we translate Γ and $\mathcal{F}$ into a linear program $\mathsf{LP}(\Gamma, \mathcal{F})$ such that there is a correspondence between the solutions of $\mathsf{LP}(\Gamma, \mathcal{F})$ and the valuations satisfying Γ and $\mathcal{F}$. For each feature f, we introduce two variables $\underline{f}$ and $\overline{f}$ such that $[\underline{f}, \overline{f}]$ represents the interval value of f.

1. First, all these intervals $[\underline{f}, \overline{f}]$ should be non-empty, and to this end we introduce the inequality $\underline{f} \leq \overline{f}$. These intervals should further be subsets of $[\underline{c}, \overline{c}]$, and thus we have the inequalities $\underline{c} \leq \underline{f}$ and $\overline{f} \leq \overline{c}$.
2. Next, consider a constraint $f \subseteq [\underline{p}, \overline{p}]$ in Γ. Replacing the feature with its variables yields $[\underline{f}, \overline{f}] \subseteq [\underline{p}, \overline{p}]$, and so we obtain the inequalities $\underline{p} \leq \underline{f}$ and $\overline{f} \leq \overline{p}$.
3. Last, consider a FI $f \subseteq \sum_{i=1}^{n} [\underline{a}_i, \overline{a}_i] \cdot g_i + [\underline{b}, \overline{b}]$ in $\mathcal{F}$. Since no negative numbers are involved, the product of each coefficient interval $[\underline{a}_i, \overline{a}_i]$ and the interval value of the feature g_i can be computed without the non-linear functions min and max. Replacing the features with their variables yields $[\underline{f}, \overline{f}] \subseteq \sum_{i=1}^{n} [\underline{a}_i, \overline{a}_i] \cdot [\underline{g}_i, \overline{g}_i] + [\underline{b}, \overline{b}]$, and thus $[\underline{f}, \overline{f}] \subseteq [\sum_{i=1}^{n} \underline{a}_i \cdot \underline{g}_i + \underline{b}, \sum_{i=1}^{n} \overline{a}_i \cdot \overline{g}_i + \overline{b}]$. We therefore obtain the inequalities $\sum_{i=1}^{n} \underline{a}_i \cdot \underline{g}_i + \underline{b} \leq \underline{f}$ and $\overline{f} \leq \sum_{i=1}^{n} \overline{a}_i \cdot \overline{g}_i + \overline{b}$. For the standard form we need to bring the linear combination of the variables to the left of $\leq$ and the number to the right.

$\mathsf{LP}(\Gamma, \mathcal{F})$ is the standard form and consists of the following inequalities:

$$\begin{array}{rl}
\underline{f} - \overline{f} \leq 0 & \text{for each feature } f \text{ occurring in } \Gamma \text{ or } \mathcal{F} \\
-\underline{f} \leq -\underline{c} & \text{for each feature } f \text{ occurring in } \Gamma \text{ or } \mathcal{F} \\
\overline{f} \leq \overline{c} & \text{for each feature } f \text{ occurring in } \Gamma \text{ or } \mathcal{F} \\
\underline{p} - \underline{f} \leq 0 & \text{for each constraint } f \subseteq [\underline{p}, \overline{p}] \text{ in } \Gamma \\
\overline{f} - \overline{p} \leq 0 & \text{for each constraint } f \subseteq [\underline{p}, \overline{p}] \text{ in } \Gamma \\
\sum_{i=1}^{n} \underline{a}_i \cdot \underline{g}_i - \underline{f} \leq -\underline{b} & \text{for each FI } f \subseteq \sum_{i=1}^{n} [\underline{a}_i, \overline{a}_i] \cdot g_i + [\underline{b}, \overline{b}] \text{ in } \mathcal{F} \\
\overline{f} - \sum_{i=1}^{n} \overline{a}_i \cdot \overline{g}_i \leq \overline{b} & \text{for each FI } f \subseteq \sum_{i=1}^{n} [\underline{a}_i, \overline{a}_i] \cdot g_i + [\underline{b}, \overline{b}] \text{ in } \mathcal{F} \\
\underline{f} \geq 0 & \text{for each feature } f \text{ occurring in } \Gamma \text{ or } \mathcal{F} \\
\overline{f} \geq 0 & \text{for each feature } f \text{ occurring in } \Gamma \text{ or } \mathcal{F}
\end{array}$$

A solution is an assignment of all variables $\underline{f}$ and $\overline{f}$ with numbers in $\mathbb{R}_+$. By definition of $\mathsf{LP}(\Gamma, \mathcal{F})$, the following statements hold:

- From each valuation v satisfying Γ and $\mathcal{F}$, we obtain a solution of $\mathsf{LP}(\Gamma, \mathcal{F})$ by mapping $\underline{f}$ to the lower bound of the interval value $v(f)$ and likewise mapping $\overline{f}$ to the upper bound of $v(f)$.
- From every solution s of $\mathsf{LP}(\Gamma, \mathcal{F})$, we obtain a valuation v that satisfies Γ and $\mathcal{F}$ by defining $v(f) \coloneqq [s(\underline{f}), s(\overline{f})]$.

It follows that Γ is satisfiable w.r.t. $\mathcal{F}$ iff. $\mathsf{LP}(\Gamma, \mathcal{F})$ is solvable.

It remains to specify the objective function of $\mathsf{LP}(\Gamma, \mathcal{F})$. Recall that there is a canonical valuation $v_{\Gamma,\mathcal{F}}$ such that $w \subseteq v_{\Gamma,\mathcal{F}}$ for each valuation w satisfying Γ and $\mathcal{F}$. Translated to solutions of $\mathsf{LP}(\Gamma, \mathcal{F})$, there is a solution $s_{\Gamma,\mathcal{F}}$ that corresponds to $v_{\Gamma,\mathcal{F}}$ and such that, for every solution t, we have $[t(\underline{f}), t(\overline{f})] \subseteq [s_{\Gamma,\mathcal{F}}(\underline{f}), s_{\Gamma,\mathcal{F}}(\overline{f})]$ for all features f. In order to compute $s_{\Gamma,\mathcal{F}}$ with $\mathsf{LP}(\Gamma, \mathcal{F})$, we would thus need to maximize all interval lengths $\overline{f} - \underline{f}$ as objective functions. Since these are all non-negative, it is enough to maximize the sum of all these lengths, which yields the single objective function $\sum_{f \in \mathbf{F}(\Gamma,\mathcal{F})} (\overline{f} - \underline{f})$, where $\mathbf{F}(\Gamma, \mathcal{F})$ is the set of all features occurring in Γ or $\mathcal{F}$. We can therefore use an ordinary LP solver — in particular with an interior-point method from linear programming [31] we can decide in polynomial time if $\mathsf{LP}(\Gamma, \mathcal{F})$ is solvable and, if so, we can further compute in polynomial time the maximal solution $s_{\Gamma,\mathcal{F}}$. □

It remains an open problem, whether the interval domains $\mathcal{D}_{\mathbf{Int}([\underline{c},\overline{c}])}$ remain P-admissible w.r.t. affine FBoxes when all interval types would be considered. We conjecture that the interval bounds can be computed using the same linear program, but determining the correct interval types (closed or open at the lower bound, closed or open at the upper bound) could possibly lead to a combinatorial explosion. It is further unclear whether, without the bounding interval $[\underline{c}, \overline{c}]$, the interval domain $\mathcal{D}_{\mathbf{Int}(\mathbb{R}_+)}$ would still be P-admissible w.r.t. affine FBoxes. The canonical valuation could then send features to intervals with upper bound $+\infty$, in which case the polytope described by the inequations would be unbounded. This requires an LP-solver with support for unbounded solution polytopes.

We can also handle affine FBoxes together with negative numbers, but then need to restrict the coefficient intervals to singletons — as otherwise the non-linear functions min and max would be required to compute a product $[\underline{a}_i, \overline{a}_i] \cdot g_i$, i.e. the system of inequalities would not be linear anymore and could therefore not be solved by linear-programming methods.

Proposition 17. *Let $\underline{c}, \overline{c} \in \mathbb{R}$ be real numbers such that $\underline{c} \leq \overline{c}$. Restricted to closed intervals, the interval domain $\mathcal{D}_{\mathbf{Int}([\underline{c},\overline{c}])}$ over the real numbers in $[\underline{c}, \overline{c}]$ is* P*-admissible w.r.t. each affine FBox $\mathcal{F}$ involving only singleton coefficients, i.e. all FIs are of the form $f \subseteq \sum_{i=1}^{n} \{a_i\} \cdot g_i + [\underline{b}, \overline{b}]$.*

Proof. The proof is similar to Proposition 16, except the following. In Step 3 in the definition of $\mathsf{LP}(\Gamma, \mathcal{F})$, the product of each singleton coefficient $\{a_i\}$ and the interval value of the feature g_i can be computed without the non-linear functions min and max. We have $\{a_i\} \cdot [\underline{g}_i, \overline{g}_i] = [a_i \cdot \underline{g}_i, a_i \cdot \overline{g}_i]$. Thus in $\mathsf{LP}(\Gamma, \mathcal{F})$ we replace every occurrence of $\underline{a}_i \cdot \underline{g}_i$ by $a_i \cdot \underline{g}_i$ and each occurrence of $\overline{a}_i \cdot \overline{g}_i$ by $a_i \cdot \overline{g}_i$.

Since $\mathbb{R}$ contains negative numbers but linear programs in standard form yield non-negative solutions only, we would need to introduce slack variables $\underline{f}^+, \underline{f}^-, \overline{f}^+, \overline{f}^-$ for all features f occurring in Γ or $\mathcal{F}$, and then replace each occurrence of $\underline{f}$ by $\underline{f}^+ - \underline{f}^-$ and likewise $\overline{f}$ by $\overline{f}^+ - \overline{f}^-$ except in the last two inequalities of $\mathsf{LP}(\Gamma, \mathcal{F})$: these are rather replaced by $\underline{f}^+ \geq 0$, $\underline{f}^- \geq 0$, $\overline{f}^+ \geq 0$, $\overline{f}^- \geq 0$. In the end, we again maximize interval lengths by means of the single objective function $\sum_{f \in \mathbf{F}(\Gamma, \mathcal{F})} ((\overline{f}^+ - \overline{f}^-) - (\underline{f}^+ - \underline{f}^-))$. □

Linear programming becomes NP-hard when restricted to integers only [33]. Unless $\mathsf{P} = \mathsf{NP}$, the integer interval domains $\mathcal{D}_{\mathbf{Int}(\mathbb{Z})}$, $\mathcal{D}_{\mathbf{Int}(\mathbb{N})}$, and $\mathcal{D}_{\mathbf{Int}(\{0,1\})}$ are thus not P-admissible w.r.t. affine FBoxes. Integer interval domains are rather suitable for integration into Horn logics that do not allow for polynomial-time reasoning, such as $\mathcal{ELI}$, Horn-$\mathcal{ALC}$, Horn-$\mathcal{SROIQ}$, and existential rules.

Example 18. Example 3 in [2] shows that the combination of the concrete domains $\mathcal{D}_{\mathbb{Q},\mathsf{diff}}$ and $\mathcal{D}_{\mathbb{Q},\mathsf{lin}}$ is not enough to express that intensive-care patients need attention if their pulse pressure is larger than 50 or their current heart rate exceeds their maximal heart rate. Moreover, this combination is not even convex.

With our interval domain these statements can be expressed through the affine FIs $\mathsf{pp} \subseteq \mathsf{sys} - \mathsf{dia}$, and $\mathsf{maxHR} \subseteq 220 - \mathsf{age}$, and $\mathsf{exceedHR} \subseteq \mathsf{hr} - \mathsf{maxHR}$, as well as the CIs $\mathsf{ICUPatient} \sqsubseteq (\mathsf{hr} \subseteq \square) \sqcap (\mathsf{sys} \subseteq \square) \sqcap (\mathsf{dia} \subseteq \square)$, and $\mathsf{ICUPatient} \sqcap (\mathsf{pp} \subseteq (50, \infty)) \sqsubseteq \mathsf{NeedsAttention}$, and $\mathsf{ICUPatient} \sqcap (\mathsf{exceedHR} \subseteq (0, \infty)) \sqsubseteq \mathsf{NeedsAttention}$.

3.2 2D-Polygons

A *2D-polygon* is a finite sequence of successively connected finite line segments in the real plane $\mathbb{R}^2$ such that the end vertex of the last segment equals the start vertex of the first. These line segments form a simple closed curve in $\mathbb{R}^2$, and by the Jordan Curve Theorem (Jordan, 1887) each 2D-polygon has an *interior region* (bounded by the curve) and an *exterior region*. In the following we identify each 2D-polygon with the subset of $\mathbb{R}^2$ consisting of its boundary and the interior region. 2D-polygons are thoroughly studied in Computational Geometry and frequently used in geographic information systems (GIS).

Deciding the set of all 2D-polygons is trivial if they are represented as finite sequences of vertex coordinates in $\mathbb{R}^2$. Clipping algorithms allow for checking in polynomial time if a polygon is a subset of another (i.e. polygon containment without moving or scaling operations). All Boolean operations (union, intersection, difference, xor) can moreover be computed by clipping algorithms in polynomial time, but the results can be of quadratic size and might consist of unions of disjoint 2D-polygons [23, 46, 54]. In order to obtain a semi-lattice, which must be closed under its meet operation, it would therefore be necessary to take the set of all finite unions of separated 2D-polygons: we denote it by $\mathbf{UGon}(\mathbb{R}^2)$, its partial order is containment $\subseteq$, and its meet is intersection $\cap$. According to the above references, $\mathbf{UGon}(\mathbb{R}^2)$ is polynomial-time computable (w.r.t. arithmetic complexity). The hierarchical concrete domain $\mathcal{D}_{\mathbf{UGon}(\mathbb{R}^2)}$ is called *polygon domain* over $\mathbb{R}^2$. A corollary to Proposition 11 is as follows.

Corollary 19. *W.r.t. arithmetic complexity, the polygon domain $\mathcal{D}_{\mathbf{UGon}(\mathbb{R}^2)}$ has exponential-time computable canonical valuations and is* EXP*-admissible w.r.t. each acyclic FBox $\mathcal{F}$ in which all operations are polynomial-time computable.*

To the best of the author's knowledge, it is unclear whether the intersection of n polygons might reach an exponential size. If this worst case would not be possible and, moreover, all operations in $\mathcal{F}$ yield linear-size results, then $\mathcal{D}_{\mathbf{UGon}(\mathbb{R}^2)}$ would even be P-admissible w.r.t. $\mathcal{F}$ (w.r.t. arithmetic complexity).

Example 20. Locations can be represented as polygons in the real plane $\mathbb{R}^2$. For instance, we have "Nöthnitzer Straße 46, 01187 Dresden" $\subseteq$ "01187 Dresden" $\subseteq$ "Dresden" $\subseteq$ "Saxony" $\subseteq$ "Germany" $\subseteq$ "Europe" $\subseteq$ "Earth".

The situation is computationally easier with *convex* 2D-polygons, which contain all line segments between each two of their points. One can think of convex 2D-polygons as two-dimensional generalizations of closed intervals. Both in linear time, we can decide the subset relation $\subseteq$ and compute the intersection operation $\cap$ for convex 2D-polygons [47, 50, 53]. However, deciding the set of all convex 2D-polygons is not trivial anymore but needs linear time [50]. We denote the semi-lattice of all convex 2D-polygons by $\mathbf{CGon}(\mathbb{R}^2)$, and it is linear-time computable (w.r.t. arithmetic complexity). The hierarchical concrete domain $\mathcal{D}_{\mathbf{CGon}(\mathbb{R}^2)}$ is called *convex-polygon domain* over $\mathbb{R}^2$. Obviously, convex polygons are not closed under Boolean operations other than intersection and these can thus not be used in FBoxes. Suitable monotonic operations besides intersection are translation, rotation, and scaling, and these can be computed in linear time as well. Below is a corollary to Proposition 10.

Corollary 21. *W.r.t. each acyclic FBox $\mathcal{F}$ in which all occurring operations are polynomial-time computable and yield linear-size results, the convex-polygon domain $\mathcal{D}_{\mathbf{CGon}(\mathbb{R}^2)}$ has polynomial-time computable canonical valuations and is* P*-admissible (w.r.t. arithmetic complexity).*

Contrary to $\mathbf{Int}(\mathbb{R})$, neither $\mathbf{UGon}(\mathbb{R}^2)$ nor $\mathbf{CGon}(\mathbb{R}^2)$ are complete. The reason is that the unit circle can be obtained as the intersection of regular polygons (for each $n \in \mathbb{N}$ with $n \geq 3$, take a smallest regular n-sided polygon that encloses the unit circle). The polygon semi-lattices are also not well-founded, and thus we cannot obtain corollaries to Theorems 8 and 9 w.r.t. cyclic FBoxes.

3.3 Regular Languages

Given a finite alphabet Σ, the semi-lattice $\mathbf{Reg}(\Sigma)$ consists of all regular languages over Σ, is partially ordered by set inclusion $\subseteq$, and its meet operation is set intersection $\cap$. It is not complete since regular languages are not closed under arbitrary intersections (only under finite ones). More specifically, $L = \bigcap\{\, \Sigma^* \setminus \{w\} \mid w \notin L \,\}$ for each language L, and thus for two symbols $a, b \in \Sigma$ the non-regular language $\{\, a^n b^n \mid n \in \mathbb{N} \,\}$ is an intersection of regular languages. Thus, convexity does not follow from Theorem 8.

In order to obtain a computable semi-lattice, we need to work with finite representations of regular languages. With regular expressions, binary intersections of regular languages can have exponential size even over a binary alphabet [24], i.e. the meet would not be computable in polynomial time. It is no alternative to instead use one-unambiguous/deterministic regular expressions since they cannot describe all regular languages and are not even closed under intersection, even though their inclusion problem is in polynomial time [19, 30, 40].

Using finite automata as representations is preferred, on the one hand since to compute the meet/intersection of two regular languages we can compute the product of the respective finite automata in polynomial time [32]. On the other hand, a language inclusion $L_1 \subseteq L_2$ holds iff. the language equivalence $L_1 \cap L_2 = L_2$ holds, and thus it suffices to check if the product of both finite automata is equivalent to the second automaton. For deterministic automata this is possible in polynomial time [16, 29], but otherwise needs polynomial space [51].

The semi-lattice $\mathbf{DFA}(\Sigma)$ consists of all deterministic finite automata over Σ, is partially ordered by automata inclusion $\preceq$ where $\mathfrak{A} \preceq \mathfrak{B}$ if $L(\mathfrak{A}) \subseteq L(\mathfrak{B})$, and its meet operation is the product $\times$, which satisfies $L(\mathfrak{A} \times \mathfrak{B}) = L(\mathfrak{A}) \cap L(\mathfrak{B})$. It is thus polynomial-time computable. Furthermore, $\mathbf{FA}(\Sigma)$ comprises all finite automata and is polynomial-space computable. Since finite automata and deterministic ones have equal power in the sense that they both describe all regular languages, both semi-lattices can serve as representations of $\mathbf{Reg}(\Sigma)$.

The hierarchical concrete domains $\mathcal{D}_{\mathbf{DFA}(\Sigma)}$ and $\mathcal{D}_{\mathbf{FA}(\Sigma)}$ are called the *regular-language domains* over Σ. Since single words are regular languages, precise string values are supported: we may write $(f = w)$ instead of $(f \preceq \mathfrak{A})$ when $L(\mathfrak{A}) = \{w\}$. Further note that $\square$ is the automaton that accepts every string, $\bot$ accepts no string at all, and $\top$ is an artificial greatest element.

Example 22. Let Σ be an alphabet containing all Latin letters, e.g. The Unicode Standard. We use a feature $\mathsf{hasTitle}$ to represent the title string of a research paper. Further take a DFA $\mathfrak{A}$ such that $L(\mathfrak{A}) = \Sigma^* \circ \{\text{description logic}\} \circ \Sigma^*$. With that, the CI $\mathsf{ScientificArticle} \sqcap (\mathsf{hasTitle} \preceq \mathfrak{A}) \sqsubseteq \mathsf{DLPaper}$ expresses that the concept of all DL papers subsumes the concept of all scientific articles with a title containing "description logic" as substring.

Even without an FBox, the regular-language domains $\mathcal{D}_{\mathbf{DFA}(\Sigma)}$ and $\mathcal{D}_{\mathbf{FA}(\Sigma)}$ are in general not P-admissible. In a nutshell, meets need not be of linear size, and thus accumulating all upper bounds of the same feature could yield an exponentially large automaton. More specifically, if a constraint set Γ contains several constraints $f \leq \mathfrak{A}$ for the same feature f, then computing the value $v_{\Gamma,\mathcal{F}}(f)$ boils down to computing the intersection of all these automata $\mathfrak{A}$. Since emptiness of intersections of finite automata is PSpace-hard [36] and graph reachability is NL-complete, $v_{\Gamma,\mathcal{F}}(f)$ cannot be computed in polynomial time, unless P = PSpace. We obtain, however, the following corollary to Proposition 11.

Corollary 23. *W.r.t. each acyclic FBox $\mathcal{F}$ in which all occurring operations are polynomial-time computable, the regular-language domain $\mathcal{D}_{\mathbf{DFA}(\Sigma)}$ has exponential-time computable canonical valuations and is* EXP*-admissible.*

The DFA operations corresponding to the language operations union $\cup$, intersection $\cap$, and complement $^-$ are polynomial-time computable. $\mathcal{D}_{\mathbf{DFA}(\Sigma)}$ is thus EXP-admissible w.r.t. each acyclic FBox involving these operations only. In contrast, concatenation $\circ$, Kleene-star *, mirror/reversal $^{\leftarrow}$, left-quotients $\backslash$, and right-quotients $/$ on DFAs are exponential-time computable but not polynomial-time computable [55]. However on FAs, all operations but complement are polynomial-time computable, and union, concatenation, Kleene-star, and mirror/reversal even with linear-size results. $\mathcal{D}_{\mathbf{FA}(\Sigma)}$ is EXPSpace-admissible w.r.t. acyclic FBoxes using these polynomial-time operations.

It is worth mentioning that, if we have at most one inclusion (i.e. constraint or FI) per feature, then in the procedure in the proof of Theorem 9 neither the automata product operation nor the automata inclusion relation needs to be used, and so we have the following corollary.

Corollary 24. *Let $\mathcal{F}$ be an acyclic FBox in which all occurring operations are polynomial-time computable and return values of linear size. Further let Γ be a constraint set. If $\mathcal{F} \cup \Gamma$ contains, for each feature f, at most one inclusion with f on the left, then the canonical valuation of Γ w.r.t. $\mathcal{F}$ can be computed in polynomial time.*

Example 25. Assume the features givenName, familyName, and name are used to represent persons' names. Then for instance, the concept $\mathsf{Male} \sqcap (\mathsf{givenName} \preceq \mathfrak{A})$ where $L(\mathfrak{A}) = \{\mathrm{F}\} \circ \Sigma^*$ describes all males whose given name starts with 'F'.

Moreover, the FI $\mathsf{name} \preceq \mathsf{givenName} \circ \{\_\} \circ \mathsf{familyName}$ allows to infer a regular language value of name when values of givenName and familyName are available (i.e. both are not $\top$). If the latter two are precise values (languages consisting of a single word), then also name gets a precise value through the FI. Note that '_' stands for a white space. The FI $\mathsf{shortName} \preceq \mathsf{initial}(\mathsf{givenName}) \circ \{.\_\} \circ \mathsf{familyName}$ generates a shortened form of a name that only contains the initial of the given name followed by a dot, where the function initial is defined by $L(\mathsf{initial}(\mathfrak{A})) \coloneqq \{\, s \mid s \in \Sigma \text{ and there is } w \in \Sigma^* \text{ such that } s \circ w \in L(\mathfrak{A}) \,\}$.

The semi-lattices $\mathbf{Reg}(\Sigma)$, $\mathbf{DFA}(\Sigma)$, and $\mathbf{FA}(\Sigma)$ are not well-founded since, already over the unary alphabet $\{a\}$, the regular languages $L_i \coloneqq \{ a^j \mid i \leq j \}$ where $i \in \mathbb{N}$ form an infinite descending chain $L_0 \supset L_1 \supset L_2 \supset \cdots$. These semi-lattices are also not complete (see above). W.r.t. cyclic FBoxes, we can thus not conclude convexity by Theorems 8 and 9.

For a restricted class of FBoxes, however, we obtain systems of language inclusions known to be solvable in exponential time [11]. An n-ary operation H on $\mathbf{DFA}(\Sigma)$ is *left-linear* if $H(\mathfrak{X}_1, \ldots, \mathfrak{X}_n) = \mathfrak{X}_1 \circ \mathfrak{A}_1 \cup \cdots \cup \mathfrak{X}_n \circ \mathfrak{A}_n \cup \mathfrak{B}$ and *right-linear* if $H(\mathfrak{X}_1, \ldots, \mathfrak{X}_n) = \mathfrak{A}_1 \circ \mathfrak{X}_1 \cup \cdots \cup \mathfrak{A}_n \circ \mathfrak{X}_n \cup \mathfrak{B}$, where $\mathfrak{A}_1, \ldots, \mathfrak{A}_n, \mathfrak{B}$ are DFAs. An FBox $\mathcal{F}$ is *linear* if the operations in its FIs are either all left-linear or all right-linear.

Proposition 26. *The regular-language domain $\mathcal{D}_{\mathbf{DFA}(\Sigma)}$ has exponential-time computable canonical valuations and is* EXP*-admissible w.r.t. each linear FBox.*

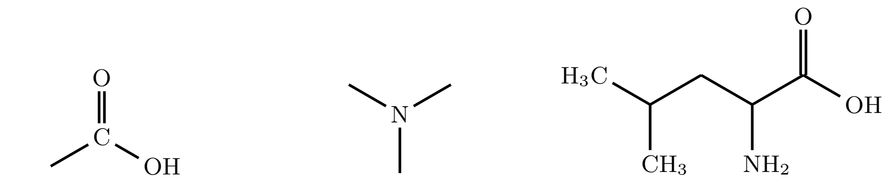
Proof. Fix a left-linear FBox $\mathcal{F}$ and a constraint set Γ. The union $\mathcal{F} \cup \Gamma$ is a system of left-linear inclusions. Now, we can translate between inclusions and equations since $X \subseteq Y$ iff. $X \cup Y = Y$. Let $(\mathcal{F} \cup \Gamma)^=$ be the so obtained system of left-linear equations. Its satisfiability can be decided in exponential time and, more importantly, it has a largest solution, which consists of regular languages, and a representation by DFAs is computable in exponential time [11]. It is easy to see that there is a one-to-one correspondence between solutions of $(\mathcal{F} \cup \Gamma)^=$ and valuations satisfying $\mathcal{F}$ and Γ. It remains to verify that the largest solution yields the canonical valuation as per Definition 7.

1. Each solution of $(\mathcal{F} \cup \Gamma)^=$ satisfies $\mathcal{F}$, and thus also the largest.
2. Each solution satisfies Γ, and thus also the largest, which yields the if direction. For the only-if direction, let $g \preceq \mathfrak{B}$ be satisfied in the largest solution of $(\mathcal{F} \cup \Gamma)^=$, and consider a valuation satisfying $\mathcal{F}$ and Γ, which is another solution of $(\mathcal{F} \cup \Gamma)^=$. The latter is thus contained in the largest solution, and thus it also satisfies $g \preceq \mathfrak{B}$.

Last, right-linear systems (from right-linear FBoxes) can be treated by their mirrors/reversals, which are left-linear [11]. Their largest solutions must be mirrored again to obtain the canonical valuations. □

When the coefficient languages are finite, then satisfiability of systems of linear inclusions or equations follows from a more general work on set constraints [1]. It further seems to be possible to add support for left-quotients in left-linear systems and for right-quotients in right-linear systems, at least for finite prefix and, respectively, suffix languages [21, 22]. Recall that the left-quotient of L_1 w.r.t. prefix L_2 is $L_2 \backslash L_1 \coloneqq \{ v \mid u \circ v \in L_1 \text{ for some } u \in L_2 \}$, and its right-quotient w.r.t. suffix L_2 is $L_1 / L_2 \coloneqq \{ v \mid v \circ w \in L_1 \text{ for some } w \in L_2 \}$. As a further side note, systems of linear language inclusions have a largest solution even if only the coefficient languages on the right-hand sides are regular, and this largest solution is regular and effectively computable [39].

If precise values (single words) are sufficient for the application, we could also use the semi-lattice $(\Sigma^* \cup \{\bot, \top\}, \leq, \wedge)$ where $\leq$ is the smallest partial order such



(a) $\mathcal{G}_{\text{carboxylic acid group}}$ (b) $\mathcal{G}_{\text{amino group}}$ (c) $\mathcal{G}_{\text{L-leucine}}$

Fig. 1: Three graphs representing chemical compounds

that $\bot < w < \top$ for each $w \in \Sigma^*$. The meet operation $\wedge$ thus satisfies $\top \wedge w = w$, $w \wedge w = w$, and $w \wedge \bot = \bot$ for each $w \in \Sigma^* \cup \{\bot, \top\}$, and $w_1 \wedge w_2 = \bot$ whenever $w_1, w_2 \in \Sigma^*$ with $w_1 \neq w_2$. This semi-lattice is complete and, by Theorem 8, its hierarchical concrete domain is convex w.r.t. every FBox. Since during the computation of a canonical valuation each feature value can be refined at most two times (from $\top$ to some w, and then possibly to $\bot$), this concrete domain is P-admissible w.r.t. each FBox in which all operations are polynomial-time computable. The disadvantage is, however, that string search like in Example 22 is not possible anymore. On the other hand, this suggests that in $\mathcal{D}_{\mathbf{DFA}(\Sigma)}$ and $\mathcal{D}_{\mathbf{FA}(\Sigma)}$ everything involving only precise values is possible in polynomial time.

3.4 Graphs

All finite, labeled graphs constitute a semi-lattice **Graph**, where the partial order $\leq$ is defined by $\mathcal{G} \leq \mathcal{H}$ if there is a homomorphism from $\mathcal{H}$ to $\mathcal{G}$. It is well-known that $\leq$ is NP-complete, but in P for acyclic graphs. The meet of two graphs is their disjoint union and can be computed in linear time, and the greatest element in this semi-lattice is the empty graph. Obviously, **Graph** is neither complete nor well-founded, and so we cannot apply Theorems 8 and 9. It thus remains unclear whether the *graph domain* $\mathcal{D}_{\mathbf{Graph}}$ is convex w.r.t. cyclic FBoxes.

Corollary 27. *The graph domain $\mathcal{D}_{\mathbf{Graph}}$ has computable canonical valuations w.r.t. acylic FBoxes. Moreover, it is NP-admissible w.r.t. every acyclic FBox in which all operations are polynomial-time computable and yield linear-size results, and it is EXP-admissible w.r.t. every acyclic FBox in which all operations are polynomial-time computable.*

Proof. The argumentation is similar to Propositions 10 and 11.

Example 28. Structural formulas of molecules can be represented as labeled graphs. Each node is labeled with the atom it represents, and the edges are labeled with the binding type (e.g. single bond, double bond, etc.). Figure 1 shows three exemplary graphs.[5] Graph (c) represents L-leucine,[6] and we can

[5] Graphs (a) and (b) are molecule parts whereas Graph (c) is a complete molecule, which cannot be a part of another molecule. The lower left node in (a) and all outer nodes in (b) can match any element in a larger molecule, be it partial or complete.

[6] In Graph (c) the skeletal formula is shown, where labels are optional for carbon atoms (C) and the hydrogen atoms (H) attached to them.

integrate it into a KB with the statement $\mathsf{L\text{-}Leucine} \equiv (\mathsf{hasMolecularStructure} \leq \mathcal{G}_{\text{L-leucine}})$. Moreover, the statement $\mathsf{AminoAcid} \equiv (\mathsf{hasMolecularStructure} \leq \mathcal{G}_{\text{carboxylic acid group}}) \sqcap (\mathsf{hasMolecularStructure} \leq \mathcal{G}_{\text{amino group}})$ expresses that amino acids are organic compounds that contain both amino and carboxylic acid functional groups. If $\mathcal{K}$ is the KB consisting of the aforementioned statements, then $\mathcal{K} \models \mathsf{L\text{-}Leucine} \sqsubseteq \mathsf{AminoAcid}$ since $\mathcal{G}_{\text{L-leucine}} \leq \mathcal{G}_{\text{carboxylic acid group}} \wedge \mathcal{G}_{\text{amino group}}$.

4 Reasoning in $\mathcal{EL}^{++}$ with Hierarchical Concrete Domains

Like other convex concrete domains, a hierarchical concrete domain $\mathcal{D}_{\mathbf{L}}$ can be integrated into $\mathcal{EL}^{++}$ but, in addition, every $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ KB may contain finitely many FIs. Of course, a model of such a KB must also satisfy all FIs in it. In order to guarantee that reasoning is decidable, a restriction on the interplay of RIs and range inclusions must be fulfilled by every $\mathcal{EL}^{++}[\mathcal{D}]$ KB [6]. To this end, we define the *range set* of a role r in $\mathcal{K}$ by $\mathsf{Range}(r, \mathcal{K}) := \{\, C \mid$ there is a role s s.t. $\mathcal{R} \models r \sqsubseteq s$ and $\mathsf{Ran}(s) \sqsubseteq C \in \mathcal{K} \,\}$, where $\mathcal{R}$ is the subset of all RIs in $\mathcal{K}$. All such range sets can be computed in polynomial time by first transforming each RI $r_1 \circ \cdots \circ r_n \sqsubseteq s$ into a context-free grammar rule $s \to r_1 \ldots r_n$, see Lemma IV in [10] for details, and then deciding the word problem for this grammar, e.g. with the Cocke–Younger–Kasami algorithm.

Definition 29. *Consider a bounded semi-lattice* $\mathbf{L}$*. An* $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ knowledge base (KB) $\mathcal{K}$ *is a finite set of CIs, RIs, range inclusions, and FIs such that*

1. $\mathsf{Range}(s, \mathcal{K}) \subseteq \mathsf{Range}(r_n, \mathcal{K})$ *for every RI* $r_1 \circ \cdots \circ r_n \sqsubseteq s$ *in* $\mathcal{K}$ *with* $n \geq 2$*,*
2. *and the hierarchical concrete domain* $\mathcal{D}_{\mathbf{L}}$ *is convex w.r.t. all FIs in* $\mathcal{K}$*.*

For a complexity class C *we say that* $\mathcal{D}_{\mathbf{L}}$ *is* C-admissible *w.r.t.* $\mathcal{K}$ *if* $\mathcal{D}_{\mathbf{L}}$ *is* C-*admissible w.r.t. the FBox consisting of all FIs in* $\mathcal{K}$*.*

For Condition 1 range inclusions on s must not imply further concept memberships than already implied by the range inclusions on r_n; otherwise emptiness of intersections of two context-free grammars could be reduced to subsumption [6]. Since $\mathsf{Range}(s, \mathcal{K}) \subseteq \mathsf{Range}(r, \mathcal{K})$ already for each RI $r \sqsubseteq s$ in $\mathcal{K}$, it above suffices to require that $n \geq 2$.

Reasoning in $\mathcal{EL}^{++}[\mathcal{D}]$ can be done by means of a rule-based calculus [5, 6, 35], and a hierarchical concrete domain $\mathcal{D}_{\mathbf{L}}$ can be seamlessly integrated into this calculus. It is only necessary to take the FIs into account, i.e. we replace the rules responsible for the interaction between concrete reasoning and logical reasoning, see Section 4.2 for details. However, we restrict attention to safe nominals, i.e. nominals $\{i\}$ must not occur in conjunctions and each right-hand side of a concept or range inclusion must not be a single nominal $\{i\}$. Full support for nominals in $\mathcal{EL}^{++}[\mathcal{D}]$ is technically quite involved and makes reasoning more expensive: the degree of the polynomial describing the worst-case reasoning time would then be larger by 1 [34]. We conjecture that the same works in $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$.

Range inclusions are not natively supported by the rule-based calculus, but they must rather be eliminated [6]. This transformation was originally described for KBs in normal form only, but can now be done without prior transformation to normal form, see Section 4.1 for details.

Assume that $\mathcal{K}$ is an $\mathcal{EL}^{++}[\mathcal{D}_\mathbf{L}]$ KB with safe nominals only. Without loss of generality we assume in the following that $\mathcal{K}$ contains only CIs of the form $C \sqsubseteq D$ or $\{i\} \sqsubseteq C$, where C and D are built with the following syntax:

$$\begin{aligned} C &::= \bot \mid C_1 \\ C_1 &::= \top \mid C_2 \mid C_2 \sqcap C_2 \mid C_2 \sqcap C_2 \sqcap C_2 \mid \dots \\ C_2 &::= A \mid f \leq p \mid \exists r.\, C_1 \mid \exists r.\, \{i\} \\ R &::= \varepsilon \mid R_1 \\ R_1 &::= r \mid R_1 \circ R_1. \end{aligned}$$

This disallows concepts with $\bot$ as subconcept, since these are equivalent to $\bot$ anyway. It further disallows $\top$ in conjunctions and, likewise, ε in non-empty role chains, since these occurrences of $\top$ or, respectively, ε can be removed without changing the meaning. Moreover, it explicitly allows conjunctions of all arities, so that we do not need to use binary conjunctions and a lot of braces.

A *subconcept* of $\mathcal{K}$ is a concept that occurs as a subexpression in $\mathcal{K}$. More formally, we define the set $\mathsf{Sub}(\mathcal{K})$ of all subconcepts of $\mathcal{K}$ as follows:

- $\mathsf{Sub}(\mathcal{K}) \coloneqq \bigcup \{\, \mathsf{Sub}(C) \cup \mathsf{Sub}(D) \mid C \sqsubseteq D \in \mathcal{K} \,\}$
- $\mathsf{Sub}(\bot) \coloneqq \{\bot\}$
- $\mathsf{Sub}(\top) \coloneqq \{\top\}$
- $\mathsf{Sub}(\{i\}) \coloneqq \{\{i\}\}$
- $\mathsf{Sub}(A) \coloneqq \{A\}$
- $\mathsf{Sub}(f \leq p) \coloneqq \{f \leq p\}$
- $\mathsf{Sub}(C_1 \sqcap \dots \sqcap C_n) \coloneqq \{C_1 \sqcap \dots \sqcap C_n\} \cup \mathsf{Sub}(C_1) \cup \dots \cup \mathsf{Sub}(C_n)$
- $\mathsf{Sub}(\exists r.\, C) \coloneqq \{\exists r.\, C\} \cup \mathsf{Sub}(C)$

4.1 Eliminating Range Inclusions

We first transform $\mathcal{K}$ into a KB $\mathcal{K}^{-\mathsf{Ran}}$ without range inclusions.

1. We copy all statements from $\mathcal{K}$ to $\mathcal{K}^{-\mathsf{Ran}}$ except the range inclusions.
2. For each role r, we choose a fresh atomic concept R_r not occurring in $\mathcal{K}$, and then we add the following CIs to $\mathcal{K}^{-\mathsf{Ran}}$:
   - $R_r \sqsubseteq C$ for each range inclusion $\mathsf{Ran}(r) \sqsubseteq C \in \mathcal{K}$.
   - $R_r \sqsubseteq R_s$ for each two roles r, s such that $\mathcal{R} \models r \sqsubseteq s$.[7]
   - $\top \sqsubseteq R_r$ for each reflexivity statement $\varepsilon \sqsubseteq r \in \mathcal{K}$.
   - $\bigsqcap \mathsf{Range}(r, \mathcal{K}) \sqsubseteq R_r$ for each role r.
3. Last, in every CI in $\mathcal{K}^{-\mathsf{Ran}}$ we recursively replace each existential restriction $\exists r.\, C$ by $\exists r.\, (C \sqcap R_r)$, i.e. we replace each $C \sqsubseteq D$ in $\mathcal{K}^{-\mathsf{Ran}}$ with $\overline{C} \sqsubseteq \overline{D}$ where
   - $\overline{\bot} \coloneqq \bot$
   - $\overline{\top} \coloneqq \top$

[7] Recall that $\mathcal{R}$ consists of all RIs in $\mathcal{K}$.

- $\overline{\{i\}} := \{i\}$ for each individual i
- $\overline{A} := A$ for each atomic concept A
- $\overline{f \leq p} := f \leq p$ for each concrete constraint $f \leq p$[8]
- $\overline{C_1 \sqcap \cdots \sqcap C_n} := \overline{C_1} \sqcap \cdots \sqcap \overline{C_n}$
- $\overline{\exists r.\, C} := \exists r.\, (\overline{C} \sqcap R_r)$

However, we need to be cautious with the existential restrictions $\exists r.\, \{i\}$ since nominals are not allowed in conjunctions (safe nominals). We instead exclude nominals the last case above and additionally define $\overline{\exists r.\, \{i\}} := \exists r.\, \{i\}$. However, whenever such an existential restriction is encountered, we need to find out whether i is an r-successor of some object — if yes, then i is in the range of r and we should add the CI $\{i\} \sqsubseteq R_r$ to $\mathcal{K}^{-\mathsf{Ran}}$ to ensure complete reasoning results.

Instead of checking each time whether i is in the range of r and to keep the reasoning procedure simpler, we rather extend the notion of nominal safety by an additional condition, which is decidable in polynomial time:

- If the KB contains a subconcept $\exists r.\, \{i\}$ and a range inclusion $\mathsf{Ran}(r) \sqsubseteq C$, then $\exists r.\, \{i\}$ must be reachable from $\top$ or a nominal $\{j\}$ in the following sense: there are CIs $C_0 \sqsubseteq D_0, \ldots, C_n \sqsubseteq D_n$ in the KB such that $C_0 = \top$ or $C_0 = \{j\}$ for some $j \in \mathbf{I}$, $\exists r.\, \{i\} \in \mathsf{Sub}(D_n)$, and for each $k \in \{1, \ldots, n\}$, there is a subconcept $E_k \in \mathsf{Sub}(D_{k-1})$ with $E_k \sqsubseteq^{\emptyset} C_k$. This ensures that the individual i is in the range of r, so that it must be an instance of C.

In the end, $\mathcal{K}^{-\mathsf{Ran}}$ can be computed in polynomial time.

Lemma V. *$\mathcal{K}^{-\mathsf{Ran}} \models R_r \sqsubseteq \overline{\bigsqcap \mathsf{Range}(r, \mathcal{K})}$ for each role r.*

Proof. Consider a role r and let $C \in \mathsf{Range}(r, \mathcal{K})$, i.e. there is a role s such that $\mathcal{R} \models r \sqsubseteq s$ and $\mathsf{Ran}(s) \sqsubseteq C \in \mathcal{K}$. Therefore $\mathcal{K}^{-\mathsf{Ran}}$ contains $R_r \sqsubseteq R_s$ and $R_s \sqsubseteq \overline{C}$, and so $\mathcal{K}^{-\mathsf{Ran}}$ entails $R_r \sqsubseteq \overline{C}$. □

Lemma VI. *Each model $\mathcal{I}$ of $\mathcal{K}$ can be extended to a model $\mathcal{J}$ of $\mathcal{K}^{-\mathsf{Ran}}$ such that $\overline{C}^{\mathcal{J}} = C^{\mathcal{I}}$ for each nominal-safe concept C in which the atomic concepts R_r do not occur.*

Proof. Given a model $\mathcal{I}$ of $\mathcal{K}$, we extend it to the interpretation $\mathcal{J}$ by additionally defining $R_r^{\mathcal{J}} := (\bigsqcap \mathsf{Range}(r, \mathcal{K}))^{\mathcal{I}}$. We show by structural induction that $C^{\mathcal{I}} = \overline{C}^{\mathcal{J}}$ for every concept C in which the atomic concepts R_r do not occur. The only interesting cases are concerned with existential restrictions, the other cases are trivial or follow easily from the induction hypothesis.

- Let $x \in (\exists r.\, D)^{\mathcal{I}}$, i.e. there is y with $(x, y) \in r^{\mathcal{I}}$ and $y \in D^{\mathcal{I}}$. The former yields $(x, y) \in r^{\mathcal{J}}$ and, since $\mathcal{I}$ satisfies all range inclusions in $\mathcal{K}$, also $y \in (\bigsqcap \mathsf{Range}(r, \mathcal{K}))^{\mathcal{I}}$, i.e. $y \in R_r^{\mathcal{J}}$. By induction hypothesis the latter yields $y \in \overline{D}^{\mathcal{J}}$, and so $x \in (\exists r.\, (\overline{D} \sqcap R_r))^{\mathcal{J}}$.

[8] This works analogously for concrete constraints $\exists f.\, P$ in general.

- Conversely, assume $x \in (\exists r.(\overline{C} \sqcap R_r))^{\mathcal{J}}$, i.e. there is y with $(x,y) \in r^{\mathcal{J}}$ and $y \in \overline{C}^{\mathcal{J}} \cap R_r^{\mathcal{J}}$. Then $(x,y) \in r^{\mathcal{I}}$ by definition of $\mathcal{J}$ and the induction hypothesis yields $y \in C^{\mathcal{I}}$. Thus, $x \in (\exists r.C)^{\mathcal{I}}$.

Next, we verify that $\mathcal{J}$ satisfies all statements in $\mathcal{K}^{-\mathsf{Ran}}$.

- We first consider a CI $R_r \sqsubseteq \overline{C}$ where $\mathsf{Ran}(r) \sqsubseteq C \in \mathcal{K}$. Assume $y \in R_r^{\mathcal{J}}$, i.e. $y \in (\bigsqcap \mathsf{Range}(r,\mathcal{K}))^{\mathcal{I}}$. Since $C \in \mathsf{Range}(r,\mathcal{K})$, we obtain $y \in C^{\mathcal{I}}$ and thus $y \in \overline{C}^{\mathcal{J}}$.
- Assume $\mathcal{R} \models r \sqsubseteq s$. We need to show that $R_r^{\mathcal{J}} \subseteq R_s^{\mathcal{J}}$. To this end, let $y \in R_r^{\mathcal{J}}$, i.e. $y \in (\bigsqcap \mathsf{Range}(r,\mathcal{K}))^{\mathcal{I}}$. Since $\mathsf{Range}(r,\mathcal{K}) \supseteq \mathsf{Range}(s,\mathcal{K})$, it follows that $y \in (\bigsqcap \mathsf{Range}(s,\mathcal{K}))^{\mathcal{I}}$ and so $y \in R_s^{\mathcal{J}}$.
- Next, we consider a CI $\top \sqsubseteq R_r$, i.e. $\mathcal{K}$ contains $\varepsilon \sqsubseteq r$. For each $x \in \mathsf{Dom}(\mathcal{J})$ we thus have $(x,x) \in r^{\mathcal{I}}$. Since $\mathcal{I}$ satisfies all range inclusions in $\mathcal{K}$, it follows that $x \in (\bigsqcap \mathsf{Range}(r,\mathcal{K}))^{\mathcal{I}}$, and so $x \in R_r^{\mathcal{J}}$.
- Consider the CI $\overline{\bigsqcap \mathsf{Range}(r,\mathcal{K})} \sqsubseteq R_r$ and let $x \in \overline{\bigsqcap \mathsf{Range}(r,\mathcal{K})}^{\mathcal{J}}$. The above yields $x \in (\bigsqcap \mathsf{Range}(r,\mathcal{K}))^{\mathcal{I}}$, i.e. $x \in R_r^{\mathcal{J}}$.
- Now we are concerned with each CI $\overline{C} \sqsubseteq \overline{D}$ where $\mathcal{K}$ contains $C \sqsubseteq D$. Since $\mathcal{I} \models \mathcal{K}$, we have $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. With $\overline{C}^{\mathcal{J}} = C^{\mathcal{I}}$ and $D^{\mathcal{I}} = \overline{D}^{\mathcal{J}}$ it follows that $\overline{C}^{\mathcal{J}} \subseteq \overline{D}^{\mathcal{J}}$.
- Consider a CI $\{i\} \sqsubseteq R_r$ in $\mathcal{K}^{-\mathsf{Ran}}$. By nominal safety, there are CIs $C_0 \sqsubseteq D_0, \ldots, C_n \sqsubseteq D_n$ in $\mathcal{K}$ such that $C_0 = \top$ or $C_0 = \{j\}$ for some $j \in \mathbf{I}$, $\exists r.\{i\} \in \mathsf{Sub}(D_n)$, and for each $k \in \{1, \ldots, n\}$, there is a subconcept $E_k \in \mathsf{Sub}(D_{k-1})$ with $E_k \sqsubseteq^{\emptyset} C_k$. Thus, we have the following:
  - $C_0^{\mathcal{I}} \neq \emptyset$
  - $C_k^{\mathcal{I}} \subseteq D_k^{\mathcal{I}}$ for each $k \in \{0, \ldots, n\}$
  - $D_{k-1}^{\mathcal{I}} \neq \emptyset$ implies $E_k^{\mathcal{I}} \neq \emptyset$ for all $k \in \{1, \ldots, n\}$
  - $E_k^{\mathcal{I}} \subseteq C_k^{\mathcal{I}}$ for each $k \in \{1, \ldots, n\}$
  - $D_n^{\mathcal{I}} \neq \emptyset$ implies $(\exists r.\{i\})^{\mathcal{I}} \neq \emptyset$

  Putting all together yields $(\exists r.\{i\})^{\mathcal{I}} \neq \emptyset$, i.e. there is some $x \in \mathsf{Dom}(\mathcal{I})$ such that $(x, i^{\mathcal{I}}) \in r^{\mathcal{I}}$. Since $\mathcal{I}$ satisfies all range inclusions in $\mathcal{K}$, we have $i^{\mathcal{I}} \in (\bigsqcap \mathsf{Range}(r,\mathcal{K}))^{\mathcal{I}}$. Since $i^{\mathcal{I}} = i^{\mathcal{J}}$, it follows that $i^{\mathcal{J}} \in R_r^{\mathcal{J}}$, as required.
- Last, since every role and every feature has the same extensions in $\mathcal{I}$ and $\mathcal{J}$, both interpretations satisfy the same RIs and FIs. □

Lemma VII. *For each model $\mathcal{J}$ of $\mathcal{K}^{-\mathsf{Ran}}$, there is a model $\mathcal{I}$ of $\mathcal{K}$ such that $\overline{C}^{\mathcal{J}} = C^{\mathcal{I}}$ for every nominal-safe concept C without any occurrence of R_r.*

Proof. Let $\mathcal{J}$ be a model of $\mathcal{K}^{-\mathsf{Ran}}$. From it we obtain the interpretation $\mathcal{I}$ by redefining $r^{\mathcal{I}} := \{ (x,y) \mid (x,y) \in r^{\mathcal{J}} \text{ and } y \in R_r^{\mathcal{J}} \}$ for every role r.

We first show by induction that $C^{\mathcal{I}} \subseteq \overline{C}^{\mathcal{J}}$ for each concept C not containing any atomic concept R_r. This is obvious for $\bot$, $\top$, nominals, atomic concepts, and constraints. For conjunctions, the claim follows easily by induction hypothesis.

- Assume $C = \exists r.\{i\}$, and let $x \in C^{\mathcal{I}}$, i.e. $(x, i^{\mathcal{I}}) \in r^{\mathcal{I}}$. By definition of $r^{\mathcal{I}}$ and since $i^{\mathcal{I}} = i^{\mathcal{J}}$ we have $(x, i^{\mathcal{J}}) \in r^{\mathcal{J}}$. Thus $x \in \overline{C}^{\mathcal{J}}$ since $\overline{\exists r.\{i\}} = \exists r.\{i\}$.

- It remains to consider $C = \exists r.D$ where D is no nominal. Then $\overline{C} = \exists r.(\overline{D} \sqcap R_r)$. Now let $x \in C^{\mathcal{I}}$, i.e. there is y such that $(x,y) \in r^{\mathcal{I}}$ and $y \in D^{\mathcal{I}}$. By definition of $r^{\mathcal{I}}$ the former yields $(x,y) \in r^{\mathcal{J}}$ and $y \in R_r^{\mathcal{J}}$, and by induction hypothesis the latter yields $y \in \overline{D}^{\mathcal{J}}$. It follows that $x \in \overline{C}^{\mathcal{J}}$.

In the converse direction, we show $\overline{C}^{\mathcal{J}} \subseteq C^{\mathcal{I}}$ by induction. This is obvious for $\bot$, $\top$, nominals, atomic concepts, and constraints. For conjunctions, the claim follows easily by induction hypothesis.

- Consider $C = \exists r.\{i\}$. Then $\overline{C} = C$ and $\mathcal{K}^{-\mathsf{Ran}}$ contains the CI $\{i\} \sqsubseteq R_r$. As a model of $\mathcal{K}^{-\mathsf{Ran}}$, $\mathcal{J}$ satisfies $\{i\} \sqsubseteq R_r$, i.e. $i^{\mathcal{J}} \in R_r^{\mathcal{J}}$. Now, if $x \in \overline{C}^{\mathcal{J}}$, then $(x, i^{\mathcal{J}}) \in r^{\mathcal{J}}$. With $i^{\mathcal{J}} = i^{\mathcal{I}}$ we conclude that $(x, i^{\mathcal{I}}) \in r^{\mathcal{I}}$, i.e. $x \in C^{\mathcal{I}}$.
- Last, let $x \in (\exists r.(\overline{D} \sqcap R_r))^{\mathcal{J}}$ where D is no nominal. Then $(x,y) \in r^{\mathcal{J}}$ for some $y \in \overline{D}^{\mathcal{J}} \cap R_r^{\mathcal{J}}$. The induction hypothesis yields $y \in D^{\mathcal{I}}$, and by definition of $r^{\mathcal{I}}$ we have $(x,y) \in r^{\mathcal{I}}$. So $x \in (\exists r.D)^{\mathcal{I}}$.

It remains to prove that $\mathcal{I}$ satisfies all statements in $\mathcal{K}$.

- First let $C \sqsubseteq D$ be a CI in $\mathcal{K}$. Then $\mathcal{K}^{-\mathsf{Ran}}$ contains $\overline{C} \sqsubseteq \overline{D}$ and thus $\overline{C}^{\mathcal{J}} \subseteq \overline{D}^{\mathcal{J}}$. As shown above, $C^{\mathcal{I}} = \overline{C}^{\mathcal{J}}$ and $\overline{D}^{\mathcal{J}} = D^{\mathcal{I}}$. It follows that $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, i.e. $\mathcal{I}$ satisfies $C \sqsubseteq D$.
- Consider a range inclusion $\mathsf{Ran}(r) \sqsubseteq C$ in $\mathcal{K}$, and let $(x,y) \in r^{\mathcal{I}}$, i.e. $(x,y) \in r^{\mathcal{J}}$ and $y \in R_r^{\mathcal{J}}$. Since $\mathcal{K}^{-\mathsf{Ran}}$ contains the CI $R_r \sqsubseteq \overline{C}$, we have $y \in \overline{C}^{\mathcal{J}}$, and thus $y \in C^{\mathcal{I}}$.
- Now consider a RI $\varepsilon \sqsubseteq r$ in $\mathcal{K}$ and let $x \in \mathsf{Dom}(\mathcal{I})$. Since $\mathcal{J} \models \mathcal{K}^{-\mathsf{Ran}}$ and $\varepsilon \sqsubseteq r \in \mathcal{K}^{-\mathsf{Ran}}$, we have $(x,x) \in r^{\mathcal{J}}$. Since $\top \sqsubseteq R_r \in \mathcal{K}^{-\mathsf{Ran}}$, we also have $x \in R_r^{\mathcal{J}}$. It follows that $(x,x) \in r^{\mathcal{I}}$.
- Next, consider a RI $r \sqsubseteq s$ in $\mathcal{K}$ and assume $(x,y) \in r^{\mathcal{I}}$. Then $(x,y) \in r^{\mathcal{J}}$ and $y \in R_r^{\mathcal{J}}$. Since $\mathcal{J}$ is a model of $\mathcal{K}^{-\mathsf{Ran}}$ and $r \sqsubseteq s$ is also in $\mathcal{K}^{-\mathsf{Ran}}$, we have $(x,y) \in s^{\mathcal{J}}$. Moreover, since $R_r \sqsubseteq R_s \in \mathcal{K}^{-\mathsf{Ran}}$, we infer that $y \in R_s^{\mathcal{J}}$, and thus $(x,y) \in s^{\mathcal{I}}$.
- Further consider a RI $r_1 \circ \cdots \circ r_n \sqsubseteq s$ in $\mathcal{K}$ with $n \geq 2$, and let $(x_0, x_1) \in r_1^{\mathcal{I}}, \ldots, (x_{n-1}, x_n) \in r_n^{\mathcal{I}}$. It follows that $(x_0, x_1) \in r_1^{\mathcal{J}}, \ldots, (x_{n-1}, x_n) \in r_n^{\mathcal{J}}$ and $x_n \in R_{r_n}^{\mathcal{J}}$. Since the RI is also contained in $\mathcal{K}^{-\mathsf{Ran}}$ and thus satisfied by $\mathcal{J}$, we infer $(x_0, x_n) \in s^{\mathcal{J}}$. Since $\mathcal{K}^{-\mathsf{Ran}} \models R_{r_n} \sqsubseteq \overline{\bigsqcap \mathsf{Range}(r_n, \mathcal{K})}$ by Lemma V, it follows that $x_n \in \overline{\bigsqcap \mathsf{Range}(r_n, \mathcal{K})}^{\mathcal{J}}$. Recall from Condition 1 in Definition 29 that $\mathsf{Range}(s, \mathcal{K}) \subseteq \mathsf{Range}(r_n, \mathcal{K})$, and thus $x_n \in \overline{\bigsqcap \mathsf{Range}(s, \mathcal{K})}^{\mathcal{J}}$. Since $\mathcal{K}^{-\mathsf{Ran}}$ contains $\overline{\bigsqcap \mathsf{Range}(s, \mathcal{K})} \sqsubseteq R_s$, we obtain $x_n \in R_s^{\mathcal{J}}$. In the end, $(x_0, x_n) \in s^{\mathcal{I}}$.
- Last, the extensions of every feature in $\mathcal{I}$ and $\mathcal{J}$ are equal, and so $\mathcal{I}$ and $\mathcal{J}$ satisfy the same FIs. $\square$

Regarding an implementation, it is easy to see that we can dispense with each additional atomic concept R_r when $\mathsf{Range}(r, \mathcal{K}) = \emptyset$, but it would have been too tedious to make this distinction in the above proofs.

Proposition VIII. *For each nominal-safe $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ KB $\mathcal{K}$, the following statements hold.*

1. *$\mathcal{K}$ and $\mathcal{K}^{-\mathsf{Ran}}$ are equi-consistent, i.e. $\mathcal{K}$ is consistent iff. $\mathcal{K}^{-\mathsf{Ran}}$ is consistent.*

2. $\mathcal{K}$ and $\mathcal{K}^{-\mathsf{Ran}}$ have the same classification.
3. $\mathcal{K} \models C \sqsubseteq D$ iff. $\mathcal{K}^{-\mathsf{Ran}} \models \overline{C} \sqsubseteq \overline{D}$ for each two nominal-safe concepts C, D in which the atomic concepts R_r do not occur.

Proof. Lemmas VI and VII yield Statement 1. Statement 2 follows from Statement 3, which we show next.

Assume $\mathcal{K}^{-\mathsf{Ran}} \models \overline{C} \sqsubseteq \overline{D}$ and consider a model $\mathcal{I}$ of $\mathcal{K}$ where $x \in C^{\mathcal{I}}$. According to Lemma VI, we can extend $\mathcal{I}$ to a model $\mathcal{J}$ of $\mathcal{K}^{-\mathsf{Ran}}$. Recall that $C^{\mathcal{I}} = \overline{C}^{\mathcal{J}}$ and so $x \in \overline{C}^{\mathcal{J}}$, which further yields $x \in \overline{D}^{\mathcal{J}}$. Since also $\overline{D}^{\mathcal{J}} = D^{\mathcal{I}}$, we conclude that $x \in D^{\mathcal{I}}$.

Conversely, let $\mathcal{K} \models C \sqsubseteq D$ and further let $\mathcal{J}$ be a model of $\mathcal{K}^{-\mathsf{Ran}}$. By Lemma VII, there is a model $\mathcal{I}$ of $\mathcal{K}$ with $D^{\mathcal{I}} = \overline{D}^{\mathcal{J}}$ and $\overline{C}^{\mathcal{J}} = C^{\mathcal{I}}$. It follows that $\overline{C}^{\mathcal{J}} = C^{\mathcal{I}} \subseteq D^{\mathcal{I}} = \overline{D}^{\mathcal{J}}$, i.e. $\mathcal{J}$ satisfies $\overline{C} \sqsubseteq \overline{D}$. □

4.2 The Completion Procedure

Now, we assume that $\mathcal{K}$ is an $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ KB that does not contain any range inclusions. In the following, we construct the set $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$, called the *saturation* of $\mathcal{K}$ w.r.t. $\mathbf{S}$, by means of rules of the form

$$[\gamma_1, \ldots, \gamma_\ell];\ \alpha_1, \ldots, \alpha_m \rightsquigarrow \beta_1, \ldots, \beta_n.$$

Such a rule is *applicable* if the *side conditions* $\gamma_1, \ldots, \gamma_\ell$ are satisfied and there is an assignment σ of the rule's variables to concepts such that $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$ contains all *premises* $\sigma(\alpha_1), \ldots, \sigma(\alpha_m)$ but not all *conclusions* $\sigma(\beta_1), \ldots, \sigma(\beta_n)$. The rule application then adds all conclusions $\sigma(\beta_1), \ldots, \sigma(\beta_n)$ to $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$. In the beginning, $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$ is initialized as the empty set. Then, all rules are applied until no rule is applicable anymore.

To formulate the side conditions, we assume that $\mathbf{S}$ is a set of concepts that contains $\top$ and $\bot$ as well as all subconcepts of $\mathcal{K}$ and is closed under subconcepts. Unless specified otherwise, we will work in the following with the smallest such set $\mathbf{S}$. The *saturation rules* are as follows, where $\mathcal{F}$ is the FBox consisting of all FIs in $\mathcal{K}$:

R_0: $[C \in \mathbf{S}] \rightsquigarrow C \sqsubseteq C$

$\mathsf{R}_\top$: $[C \in \mathbf{S}] \rightsquigarrow C \sqsubseteq \top$

$\mathsf{R}_\sqcap^-$: $C \sqsubseteq D_1 \sqcap \cdots \sqcap D_n \rightsquigarrow C \sqsubseteq D_1, \ldots, C \sqsubseteq D_n$

$\mathsf{R}_\sqcap^+$: $[D_1 \sqcap \cdots \sqcap D_n \in \mathbf{S},\ n \geq 2];\ C \sqsubseteq D_1, \ldots, C \sqsubseteq D_n \rightsquigarrow C \sqsubseteq D_1 \sqcap \cdots \sqcap D_n$

$\mathsf{R}_\exists$: $[\exists r.E \in \mathbf{S}];\ C \sqsubseteq \exists r.D,\ D \sqsubseteq E \rightsquigarrow C \sqsubseteq \exists r.E$

$\mathsf{R}_{\exists,\bot}$: $C \sqsubseteq \exists r.D,\ D \sqsubseteq \bot \rightsquigarrow C \sqsubseteq \bot$

$\mathsf{R}_\bot$: $[D \in \mathbf{S}];\ C \sqsubseteq \bot \rightsquigarrow C \sqsubseteq D$

$\mathsf{R}_\sqsubseteq$: $[D \sqsubseteq E \in \mathcal{K}];\ C \sqsubseteq D \rightsquigarrow C \sqsubseteq E$

R_ε: $[C \in \mathbf{S},\ \varepsilon \sqsubseteq r \in \mathcal{K}] \rightsquigarrow C \sqsubseteq \exists r.C$

$\mathsf{R}_\circ$: $[r_1 \circ \cdots \circ r_n \sqsubseteq s \in \mathcal{K},\ n \geq 1];\ C_0 \sqsubseteq \exists r_1.C_1, \ldots, C_{n-1} \sqsubseteq \exists r_n.C_n \rightsquigarrow C_0 \sqsubseteq \exists s.C_n$

$\mathsf{R}_\mathcal{D}$: $[\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models (f_1 \leq p_1) \sqcap \cdots \sqcap (f_m \leq p_m) \sqsubseteq (g \leq q),\ (g \leq q) \in \mathbf{S}];\ C \sqsubseteq (f_1 \leq p_1), \ldots, C \sqsubseteq (f_m \leq p_m) \rightsquigarrow C \sqsubseteq (g \leq q)$

$\mathsf{R}_{\mathcal{D},\bot}$: $[(f_1 \leq p_1) \sqcap \cdots \sqcap (f_m \leq p_m)$ unsatisfiable in $\mathcal{D}_{\mathbf{L}}$ w.r.t. $\mathcal{F}]$; $C \sqsubseteq (f_1 \leq p_1)$, $\ldots$, $C \sqsubseteq (f_m \leq p_m) \rightsquigarrow C \sqsubseteq \bot$

Proposition IX. *Consider a bounded semi-lattice* $\mathbf{L}$ *and let* $\mathcal{K}$ *be a nominal-safe* $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ *KB without range inclusions. Further let* $\mathbf{S}$ *be a finite set of concepts with* $\mathsf{Sub}(\mathcal{K}) \subseteq \mathbf{S}$ *and* $\top, \bot \in \mathbf{S}$ *and that is closed under subconcepts.*

1. $\mathcal{K}$ *is consistent iff.* $\top \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ *and* $\{i\} \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ *for each* $\{i\} \in \mathbf{S}$.
2. *If* $\mathcal{K}$ *is consistent, then* $\mathcal{K} \models C \sqsubseteq D$ *iff.* $C \sqsubseteq D \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ *for all concepts* $C, D \in \mathbf{S}$.

Proof. It is easy to verify that each rule applied to CIs entailed by $\mathcal{K}$ produces CIs also entailed by $\mathcal{K}$. By an induction along the applications of the above rules it follows that every CI in $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$ is entailed by $\mathcal{K}$. This yields the if direction of Statement 2. We further conclude that, if $\top \sqsubseteq \bot \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$, then $\mathcal{K}$ entails $\top \sqsubseteq \bot$. Since no interpretation satisfies the latter CI, there are no models of $\mathcal{K}$, i.e. $\mathcal{K}$ is inconsistent. If $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$ contains a CI $\{i\} \sqsubseteq \bot$ with $\{i\} \in \mathbf{S}$, then we can argue similarly. So also the only-if direction of Statement 1 holds.

Regarding the if direction of Statement 1, assume that $\top \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ and $\{i\} \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ for each $\{i\} \in \mathbf{S}$. Then the following interpretation $\mathcal{I}_{\mathcal{K},\mathbf{S}}$, called *canonical model* of $\mathcal{K}$ w.r.t. $\mathbf{S}$, is well-defined.

- $\mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}}) := \{\, x_C \mid C \in \mathbf{S} \text{ and } C \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}, \mathbf{S}) \,\}$
- $i^{\mathcal{I}_{\mathcal{K},\mathbf{S}}} := \begin{cases} x_{\{i\}} & \text{if } \{i\} \in \mathbf{S}, \text{ and} \\ x_\top & \text{otherwise, for each individual } i \end{cases}$
- $A^{\mathcal{I}_{\mathcal{K},\mathbf{S}}} := \{\, x_C \mid x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}}) \text{ and } C \sqsubseteq A \in \mathsf{Sat}(\mathcal{K}, \mathbf{S}) \,\}$ for each atomic concept A
- $r^{\mathcal{I}_{\mathcal{K},\mathbf{S}}} := \{\, (x_C, x_D) \mid x_C, x_D \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}}) \text{ and } C \sqsubseteq \exists r.D \in \mathsf{Sat}(\mathcal{K}, \mathbf{S}) \,\}$ for each role r

It remains to interpret the features. If the concrete domain $\mathcal{D}_{\mathbf{L}}$ has canonical valuations, then we define:

- $f^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}(x_C) := v_{\Gamma_C,\mathcal{F}}(f)$ for each feature f and for each $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$, where $v_{\Gamma_C,\mathcal{F}}$ is the canonical valuation of the constraint set $\Gamma_C := \{\, f \leq p \mid C \sqsubseteq (f \leq p) \in \mathsf{Sat}(\mathcal{K}, \mathbf{S}) \,\}$.

The valuation $v_{\Gamma_C,\mathcal{F}}$ exists since Γ_C is satisfiable — otherwise Rule $\mathsf{R}_{\mathcal{D},\bot}$ would have produced $C \sqsubseteq \bot$, a contradiction to $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$. Further recall that $v_{\Gamma_C,\mathcal{F}} \models (f \leq p)$ iff. $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap \Gamma_C \sqsubseteq (f \leq p)$ and, since the Rule $\mathsf{R}_{\mathcal{D}}$ has been applied exhaustively, the latter holds iff. $C \sqsubseteq (f \leq p) \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$.

Otherwise, we interpret the features similarly to Claim 2 in Lemma 7 in [5]. Consider some $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$, i.e. $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$ does not contain $C \sqsubseteq \bot$. As otherwise Rule $\mathsf{R}_{\mathcal{D},\bot}$ would have produced $C \sqsubseteq \bot$, the conjunction $\bigsqcap \Gamma_C$ where $\Gamma_C := \{\, f \leq p \mid C \sqsubseteq (f \leq p) \in \mathsf{Sat}(\mathcal{K}, \mathbf{S}) \,\}$ is satisfiable in $\mathcal{D}_{\mathbf{L}}$ w.r.t. $\mathcal{F}$ (all FIs in $\mathcal{K}$). Now, if every interpretation/valuation satisfying $\mathcal{F}$ and this conjunction $\bigsqcap \Gamma_C$ also satisfied another constraint in $\Delta_C := \{\, g \leq q \mid C \sqsubseteq (g \leq q) \notin$

$\mathsf{Sat}(\mathcal{K}, \mathbf{S})$ but $(g \leq q) \in \mathbf{S}\,\}$, then the constraint inclusion $\bigsqcap \varGamma_C \sqsubseteq \bigsqcup \varDelta_C$ would be valid in $\mathcal{D}_{\mathbf{L}}$ w.r.t. $\mathcal{F}$. Since $\mathcal{D}_{\mathbf{L}}$ is convex w.r.t. $\mathcal{F}$, some $g \leq q$ in $\varDelta_C$ would be implied by $\bigsqcap \varGamma_C$, but then Rule $\mathsf{R}_{\mathcal{D}}$ would have produced $C \sqsubseteq (g \leq q)$, a contradiction. There is thus a valuation $v_C \colon \mathbf{F} \to \mathsf{Dom}(\mathcal{D}_{\mathbf{L}})$ that satisfies $\mathcal{F}$ and such that, for each constraint $f \leq p$ in $\mathbf{S}$, $C \sqsubseteq (f \leq p) \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ iff. v_C satisfies $f \leq p$. With all these valuations v_C we can now define:

- $f^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}(x_C) := v_C(f)$ for every feature f and for each $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$.

We continue with proving that $x_C \in D^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ iff. $C \sqsubseteq D \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ for each $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$ and for each $D \in \mathbf{S}$. We show this claim by structural induction on D. (This is possible since $\mathbf{S}$ is closed under subconcepts.)

- If $D = \top$, then $x_C \in \top^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ by the very definition of semantics and $C \sqsubseteq \top \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by Rule $\mathsf{R}_\top$.
- Let $D = \bot$. Since $x_C \notin \bot^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ by the very definition of semantics, the only-if direction holds. Conversely, if $C \sqsubseteq \bot$ was in $\mathsf{Sat}(\mathcal{K}, \mathbf{S})$, then x_C would not be in $\mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$, a contradiction, and thus the if direction also holds.
- Assume $D = \{i\}$. If $x_C \in \{i\}^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$, then $C = \{i\}$ as well, and thus $C \sqsubseteq \{i\} \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by Rule R_0.
  In the opposite direction, if $C \sqsubseteq \{i\} \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$, then this CI can only have been created by Rule R_0, i.e. $C = \{i\}$ and thus $x_C \in \{i\}^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. To see this, note that Rules $\mathsf{R}_\top$, $\mathsf{R}_\sqcap^+$, $\mathsf{R}_\exists$, $\mathsf{R}_{\exists,\bot}$, R_ε, $\mathsf{R}_\circ$, $\mathsf{R}_{\mathcal{D}}$, and $\mathsf{R}_{\mathcal{D},\bot}$ never produce CIs with nominals as conclusion. Moreover, $C \sqsubseteq \{i\}$ could not have been created by Rule $\mathsf{R}_\sqcap^-$ since $\{i\}$ cannot occur in any conjunction (safe nominals). $C \sqsubseteq \{i\}$ could not have been created by Rule $\mathsf{R}_\bot$ since $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$ requires that $C \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}, \mathbf{S})$. Last, $C \sqsubseteq \{i\}$ could not have been introduced by Rule $\mathsf{R}_\sqsubseteq$ since $\{i\}$ cannot be the conclusion of any CI in $\mathcal{K}$ (safe nominals).
- If $D = A$, then $x_C \in A^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ iff. $C \sqsubseteq A \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by definition of $\mathcal{I}_{\mathcal{K},\mathbf{S}}$.
- In the case where D is a constraint $f \leq p$, the claim follows from the above definition of the feature interpretations $f^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. If this was done with the canonical valuations $v_{\varGamma_C,\mathcal{F}}$, then $x_C \in (f \leq p)^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ iff. $v_{\varGamma_C,\mathcal{F}} \models (f \leq p)$ iff. $C \sqsubseteq (f \leq p) \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$. Otherwise, it similarly holds that $x_C \in (f \leq p)^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ iff. $v_C \models (f \leq p)$ iff. $C \sqsubseteq (f \leq p) \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$.
- For $D = D_1 \sqcap \dots \sqcap D_n$ we have:
  $x_C \in (D_1 \sqcap \dots \sqcap D_n)^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$
  iff. $x_C \in D_1^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}, \ \dots, \ x_C \in D_n^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ by definition of semantics
  iff. $\{C \sqsubseteq D_1, \ \dots, \ C \sqsubseteq D_n\} \subseteq \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by induction hypothesis
  iff. $C \sqsubseteq D_1 \sqcap \dots \sqcap D_n \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by Rules $\mathsf{R}_\sqcap^+$ and $\mathsf{R}_\sqcap^-$
- Last, assume $D = \exists r.E$. Recall that $x_C \in (\exists r.E)^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ iff. there is x_F with $(x_C, x_F) \in r^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ and $x_F \in E^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. The former holds iff. $C \sqsubseteq \exists r.F \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by definition of $\mathcal{I}_{\mathcal{K},\mathbf{S}}$, and the latter implies $F \sqsubseteq E \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by induction hypothesis. Rule $\mathsf{R}_\exists$ ensures that $C \sqsubseteq \exists r.E \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$.
  It remains to show the opposite direction. If $C \sqsubseteq \exists r.E \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$, then we also have $E \sqsubseteq E \in \mathsf{Sat}(\mathcal{K}, \mathbf{S})$ by Rule R_0. The element x_E is in $\mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$ since otherwise x_C would not be in $\mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$ by Rule $\mathsf{R}_{\exists,\bot}$, a contradiction. So $(x_C, x_E) \in r^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$, and $x_E \in E^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ by induction hypothesis. It follows that $x_C \in (\exists r.E)^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$, as required.

Next, we show that $\mathcal{I}_{\mathcal{K},\mathbf{S}}$ is a model of $\mathcal{K}$.

- Consider a CI $D \sqsubseteq E \in \mathcal{K}$ and an element $x_C \in D^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. By the above claim, the latter implies $C \sqsubseteq D \in \mathsf{Sat}(\mathcal{K},\mathbf{S})$, and thus Rule $\mathsf{R}_{\sqsubseteq}$ yields $C \sqsubseteq E \in \mathsf{Sat}(\mathcal{K},\mathbf{S})$. With the above claim we conclude that $x_C \in E^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$.
- Assume a RI $\varepsilon \sqsubseteq r \in \mathcal{K}$ and an element $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$. Then $C \in \mathbf{S}$ and Rule R_ε adds the CI $C \sqsubseteq \exists r.C$ to $\mathsf{Sat}(\mathcal{K},\mathbf{S})$. The definition of $\mathcal{I}_{\mathcal{K},\mathbf{S}}$ ensures that $(x_C, x_C) \in r^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$.
- Take a RI $r_1 \circ \cdots \circ r_n \sqsubseteq s \in \mathcal{K}$ with $n \geq 1$ and a pair $(x_{C_0}, x_{C_n}) \in (r_1 \circ \cdots \circ r_n)^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. Then there are intermediate elements x_{C_i} with $(x_{C_0}, x_{C_1}) \in r_1^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$, $\ldots$, $(x_{C_{n-1}}, x_{C_n}) \in r_n^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. By definition of $\mathcal{I}_{\mathcal{K},\mathbf{S}}$ we have $\{C_0 \sqsubseteq \exists r_1.C_1, \ldots, C_{n-1} \sqsubseteq \exists r_n.C_n\} \subseteq \mathsf{Sat}(\mathcal{K},\mathbf{S})$. Rule $\mathsf{R}_\circ$ yields $C_0 \sqsubseteq \exists s.C_n \in \mathsf{Sat}(\mathcal{K},\mathbf{S})$, i.e. $(x_{C_0}, x_{C_n}) \in s^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$.
- If the feature extensions are defined through the canonical valuations $v_{\Gamma_C,\mathcal{F}}$, $\mathcal{I}_{\mathcal{K},\mathbf{S}}$ satisfies all FIs since all canonical valuations satisfy $\mathcal{F}$ (the FIs in $\mathcal{K}$). Otherwise, the instead used valuations v_C satisfy $\mathcal{F}$ and thus $\mathcal{I}_{\mathcal{K},\mathbf{S}}$ satisfies every FI as well.

Since $\mathcal{I}_{\mathcal{K},\mathbf{S}} \models \mathcal{K}$, we conclude that $\mathcal{K}$ is consistent.

Last, it remains to verify the only-if direction of Statement 2. To this end, assume that $\mathcal{K}$ is consistent and let $\mathcal{K} \models C \sqsubseteq D$ for concepts $C, D \in \mathbf{S}$.

- If $\mathsf{Sat}(\mathcal{K},\mathbf{S})$ contains $C \sqsubseteq \bot$, then the CI $C \sqsubseteq D$ was added by an application of Rule $\mathsf{R}_\bot$ to $\mathsf{Sat}(\mathcal{K},\mathbf{S})$.
- Now let $C \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K},\mathbf{S})$, i.e. $x_C \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K},\mathbf{S}})$. Since $\mathcal{K}$ is consistent, $\mathcal{I}_{\mathcal{K},\mathbf{S}}$ is a model of $\mathcal{K}$ and thus satisfies the CI $C \sqsubseteq D$. Since $C \sqsubseteq C \in \mathsf{Sat}(\mathcal{K},\mathbf{S})$ by Rule R_0, the above claim yields $x_C \in C^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$ and thus $x_C \in D^{\mathcal{I}_{\mathcal{K},\mathbf{S}}}$. Another application of the above claim shows that $C \sqsubseteq D \in \mathsf{Sat}(\mathcal{K},\mathbf{S})$. □

Lemma X. $\mathsf{Sat}(\mathcal{K},\mathbf{S})$ *can be computed in polynomial time.*

Proof. All rules but $\mathsf{R}_\circ$ only yield CIs $C \sqsubseteq D$ in which both concepts C and D are contained in $\mathbf{S}$, i.e. the size of all CIs produced by these rules is at most quadratic in the size of $\mathbf{S}$ and the total number of rule applications is at most quadratic too. The Rule $\mathsf{R}_\circ$ instead produces CIs $C_0 \sqsubseteq \exists s.C_n$ where C_0 and C_n are both in $\mathbf{S}$ but $\exists s.C_n$ need not always be in $\mathbf{S}$. Thus, the overall number of produced CIs in $\mathsf{Sat}(\mathcal{K},\mathbf{S})$ is bounded by $k^2 \cdot \ell$, where k is the number of concepts in $\mathbf{S}$ and ℓ is the number of RIs in $\mathcal{K}$. A single rule application needs only polynomial time. Finally, finding the next applicable rule is possible in polynomial time as follows. One tries the rules in the order given. For Rule $\mathsf{R}_\sqcap^+$, one goes through all conjunctions $D_1 \sqcap \cdots \sqcap D_n \in \mathbf{S}$, which are polynomially many, and for each of them one checks if CIs $C \sqsubseteq D_1, \ldots, C \sqsubseteq D_n$ have already been produced. (Naïvely checking all subsets of already produced CIs would need exponential time instead.) One similarly checks for applicability of Rule $\mathsf{R}_\circ$. For the other rules it is obvious that applicability can be checked in polynomial time. □

By putting Propositions VIII and IX together we obtain the following.

Corollary XI. *Assume that* $\mathbf{L}$ *is a bounded semi-lattice and let* $\mathcal{K}$ *be a nominal-safe* $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ *KB. Further consider a finite set* $\mathbf{S}$ *of concepts in which the atomic concepts* R_r *do not occur, that is closed under subconcepts, and such that* $\top, \bot \in \mathbf{S}$ *and* $\mathsf{Sub}(\mathcal{K}) \subseteq \mathbf{S}$*. Then let* $\overline{\mathbf{S}} := \{\, \overline{C} \mid C \in \mathbf{S} \,\}$*.*

1. $\mathcal{K}$ *is consistent iff.* $\top \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ *and* $\{i\} \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ *for each* $\{i\} \in \mathbf{S}$*.*
2. *If* $\mathcal{K}$ *is consistent, then* $\mathcal{K} \models C \sqsubseteq D$ *iff.* $\overline{C} \sqsubseteq \overline{D} \in \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ *for all concepts* $C, D \in \mathbf{S}$*.*

4.3 Computational Complexity

Next, we determine the computational complexity of the saturation procedure. To this end, we show that each $\mathcal{EL}^{++}[\mathcal{D}]$ KB has at most polynomially many subconcepts, and that the size of $\mathsf{Sub}(\mathcal{K})$ is polynomial in the size of $\mathcal{K}$. The *size* is defined recursively:

- $|\mathcal{K}| := \sum(\, |C \sqsubseteq D| \mid C \sqsubseteq D \in \mathcal{K}\,)$
- $|C \sqsubseteq D| := |C| + |D| + 1$
- $|\bot| := 1$
- $|\top| := 1$
- $|\{i\}| := 1$
- $|A| := 1$
- $|\exists f_1, \ldots, f_k.P| := k + 2$
- $|C_1 \sqcap \cdots \sqcap C_n| := |C_1| + \cdots + |C_n| + (n-1)$
- $|\exists r.C| := |C| + 2$

We show by induction on the structure of C that the size of $\mathsf{Sub}(C)$ is polynomial in the size of C.

- Recall that $\mathsf{Sub}(C) = \{C\}$ if C is $\bot$, $\top$, a nominal $\{i\}$, or a atomic concept A. In these cases the size of $\mathsf{Sub}(C)$ is obviously linear in the size of C.
- Regarding conjunctions. Since $\mathsf{Sub}(C_1 \sqcap \cdots \sqcap C_n) = \{C_1 \sqcap \cdots \sqcap C_n\} \cup \mathsf{Sub}(C_1) \cup \cdots \cup \mathsf{Sub}(C_n)$, the size of $\mathsf{Sub}(C_1 \sqcap \cdots \sqcap C_n)$ is the size of $C_1 \sqcap \cdots \sqcap C_n$ plus the sizes of $\mathsf{Sub}(C_1)$, ..., $\mathsf{Sub}(C_n)$. By induction hypothesis, the size of each $\mathsf{Sub}(C_i)$ is polynomial in the size of C_i. Since the size of each C_i is bounded by the size of $C_1 \sqcap \cdots \sqcap C_n$, it follows that the size of $\mathsf{Sub}(C_1 \sqcap \cdots \sqcap C_n)$ is polynomial in the size of $C_1 \sqcap \cdots \sqcap C_n$.
- For existential restrictions, we have $\mathsf{Sub}(\exists r.C) = \{\exists r.C\} \cup \mathsf{Sub}(C)$. Thus the size of $\mathsf{Sub}(\exists r.C)$ is the size of $\exists r.C$ plus the size of $\mathsf{Sub}(C)$. By induction hypothesis, the latter size if polynomial in the size of C, which is bounded by the size of $\exists r.C$. We conclude that the size of $\mathsf{Sub}(\exists r.C)$ is polynomial in the size of $\exists r.C$.

Finally, since for each CI $C \sqsubseteq D$ in $\mathcal{K}$ the size of C and the size of D are both bounded by the size of $\mathcal{K}$, we conclude that the size of $\mathsf{Sub}(\mathcal{K})$ is polynomial in the size of $\mathcal{K}$.

Theorem 30. *Let* $\mathbf{L}$ *be a bounded semi-lattice. For all nominal-safe* $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ *KBs w.r.t. which the hierarchical concrete domain* $\mathcal{D}_{\mathbf{L}}$ *is* P*-admissible, the following reasoning tasks can be done in polynomial time: consistency, classification, subsumption checking, instance checking, and concept satisfiability.*

Proof. According to Corollary XI, KB consistency and subsumption checking can be done by first computing $\mathcal{K}^{-\mathsf{Ran}}$ and $\overline{\mathbf{S}}$ (both in polynomial time), then computing $\mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ (in polynomial time by Lemma X), and finally looking up whether it contains particular CIs, where for checking a subsumption $C \sqsubseteq D$ the set $\mathbf{S}$ must contain both C and D. Instance checking is a special form of subsumption checking since CAs can be expressed by means of nominals. Obviously also concept satisfiability is a special form of subsumption checking. Finally, $\mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ contains a classification of $\mathcal{K}$. □

Currently the fastest $\mathcal{ELR}^{\bot}$ reasoner is ELK [35], which is a highly optimized, multi-threaded implementation of the polynomial-time saturation algorithm. It can classify SNOMED CT, a large medical ontology with more than 360,000 atomic concepts, in a few seconds on a modern laptop. $\mathcal{ELR}^{\bot}$ is $\mathcal{EL}^{++}[\mathcal{D}]$ without range restrictions, nominals, and concrete domains. It might be useful to extend ELK with support for range restrictions, safe nominals, and hierarchical concrete domains.

In the proof of the above result, we build a canonical model of the input KB iff. it is consistent. Now with the hierarchical concrete domains we can use the canonical valuations for this. The benefit is that the canonical model is complete for all assertions $\{i\} \sqsubseteq C$, before it was only complete for such assertions where C contains no concrete constraints. Our canonical models are thus appropriate for computing optimal repairs [9, 10, 37, 38] of KBs involving concrete domains.

We can also use NP- or EXP-admissible concrete domains in $\mathcal{EL}^{++}$. Reasoning works in the very same way, i.e. the logical reasoning can still be done in polynomial time, but the concrete reasoning is more expensive.

Theorem 31. *Fix a bounded semi-lattice* $\mathbf{L}$*. For all nominal-safe* $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ *KBs w.r.t. which the hierarchical concrete domain* $\mathcal{D}_{\mathbf{L}}$ *is* NP*-admissible, the following reasoning problems are in* NP*: consistency, concept satisfiability, subsumption checking, and instance checking. They are in* EXP *if* $\mathcal{D}_{\mathbf{L}}$ *is* EXP*-admissible. In both cases, the classification can be computed in exponential time.*

4.4 The Canonical Model

Definition XII. *Let* $\mathbf{L}$ *be a bounded semi-lattice such that the hierarchical concrete domain* $\mathcal{D}_{\mathbf{L}}$ *has canonical valuations, and assume that the signature contains only finitely many individuals. Further consider a consistent, nominal-safe* $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ *KB* $\mathcal{K}$ *and define* $\mathbf{S} \coloneqq \{\bot, \top\} \cup \mathsf{Sub}(\mathcal{K}) \cup \{\, \{i\} \mid i \text{ is an individual} \,\}$ *and* $\overline{\mathbf{S}} \coloneqq \{\, \overline{C} \mid C \in \mathbf{S} \,\}$*. The* canonical model $\mathcal{I}_{\mathcal{K}}$ *is obtained from the canonical model* $\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}$ *in the proof of Proposition IX by redefining role extensions as in Lemma VII.*

It follows from Lemma X that the canonical model $\mathcal{I}_{\mathcal{K}}$ can be computed in polynomial time.

We will show that $\mathcal{I}_{\mathcal{K}}$ is *universal w.r.t. nominal-safe assertions*, i.e. $\mathcal{K} \models i{:}C$ iff. $\mathcal{I}_{\mathcal{K}} \models i : C$ for each individual i and for each nominal-safe concept C. The above canonical models are thus suitable for computing optimal repairs of ABoxes w.r.t. static ontologies. More generally, we will show that $\mathcal{K} \models C \sqsubseteq D$ iff. $\mathcal{I}_{\mathcal{K}} \models C \sqsubseteq D$ for each $C \in \mathbf{S}$ and for each nominal-safe concept D. Therefore these canonical models are also appropriate for computing optimal fixed-premise repairs of KBs (where the ontology is not considered static but can be modified).

Definition XIII. *A* nominal-safe simulation *from an interpretation $\mathcal{I}$ to another interpretation $\mathcal{J}$ is a relation $\mathfrak{S} \subseteq \mathsf{Dom}(\mathcal{I}) \times \mathsf{Dom}(\mathcal{J})$ such that*

1. *$(i^{\mathcal{I}}, i^{\mathcal{J}}) \in \mathfrak{S}$ for every individual i*

and the following hold for each pair $(x, y) \in \mathfrak{S}$:

2. *For each atomic concept A, if $x \in A^{\mathcal{I}}$, then $y \in A^{\mathcal{J}}$.*
3. *For every role r, if $(x, x') \in r^{\mathcal{I}}$, then there is y' such that $(x', y') \in \mathfrak{S}$ and $(y, y') \in r^{\mathcal{J}}$.*
4. *For each constraint $f \leq p$, if $x \in (f \leq p)^{\mathcal{I}}$, then $y \in (f \leq p)^{\mathcal{J}}$.*
5. *For every individual i, if $(x, i^{\mathcal{I}}) \in r^{\mathcal{I}}$, then $(y, i^{\mathcal{J}}) \in r^{\mathcal{J}}$.*

Lemma XIV. *If $\mathfrak{S}$ is a nominal-safe simulation from $\mathcal{I}$ to $\mathcal{J}$ with $(x, y) \in \mathfrak{S}$, and C is a nominal-safe concept with $x \in C^{\mathcal{I}}$, then $y \in C^{\mathcal{J}}$.*

Proof. We show the claim by induction on C. The cases where C is $\bot$ or $\top$ are trivial, and those where C is an atomic concept, a constraint, or of the form $\exists r.\{i\}$ follow directly from Definition XIII. When C is a conjunction, then the claim follows easily from the induction hypothesis.

It remains to investigate the case $C = \exists r.D$. To this end, let $x \in (\exists r.D)^{\mathcal{I}}$, i.e. there is x' such that $(x, x') \in r^{\mathcal{I}}$ and $x' \in D^{\mathcal{I}}$. Definition XIII yields some y' such that $(x', y') \in \mathfrak{S}$ and $(y, y') \in r^{\mathcal{J}}$. So we infer that $y' \in D^{\mathcal{J}}$ by induction hypothesis, and thus $y \in (\exists r.D)^{\mathcal{J}}$, as required. □

Lemma XV. *Consider a bounded semi-lattice $\mathbf{L}$ such that $\mathcal{D}_{\mathbf{L}}$ has canonical valuations, and let $\mathcal{K}$ be a consistent nominal-safe $\mathcal{EL}^{++}[\mathcal{D}_{\mathbf{L}}]$ KB.*

1. *A concept $C \in \mathbf{S}$ is satisfiable w.r.t. $\mathcal{K}$ iff. $x_{\overline{C}} \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K}})$.*
2. *$\mathcal{K} \models C \sqsubseteq D$ iff. $x_{\overline{C}} \in D^{\mathcal{I}_{\mathcal{K}}}$ for each $\mathcal{K}$-satisfiable concept $C \in \mathbf{S}$ and for each nominal-safe concept D.*[9]

Proof. We begin with the first claim. Recall that $\mathbf{S} := \{\bot, \top\} \cup \mathsf{Sub}(\mathcal{K}) \cup \{\, \{i\} \mid i \text{ is an individual}\,\}$, and let $C \in \mathbf{S}$.

C is satisfiable w.r.t. $\mathcal{K}$.
iff. $\mathcal{K} \not\models C \sqsubseteq \bot$

[9] D is an arbitrary nominal-safe concept and need not be in $\mathbf{S}$.

iff. $\overline{C} \sqsubseteq \bot \notin \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ (by Corollary XI)
iff. $x_{\overline{C}} \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}})$ (see proof of Proposition IX)
iff. $x_{\overline{C}} \in \mathsf{Dom}(\mathcal{I}_{\mathcal{K}})$ (by Definition XII)

Next, we show the second claim. Let $\mathcal{K} \models C \sqsubseteq D$. Since $\overline{C} \in \overline{\mathbf{S}}$, Rule R_0 adds $\overline{C} \sqsubseteq \overline{C}$ to $\mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$, and thus the claim in the proof of Proposition IX yields $x_{\overline{C}} \in \overline{C}^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}$. Lemma VII yields that $x_{\overline{C}} \in C^{\mathcal{I}_{\mathcal{K}}}$ and that $\mathcal{I}_{\mathcal{K}}$ is a model of $\mathcal{K}$. We therefore conclude that $x_{\overline{C}} \in D^{\mathcal{I}_{\mathcal{K}}}$.

In the converse direction, assume $x_{\overline{C}} \in D^{\mathcal{I}_{\mathcal{K}}}$ and further consider a model $\mathcal{I}$ of $\mathcal{K}$ such that $y \in C^{\mathcal{I}}$. By Lemma VI we obtain from $\mathcal{I}$ a model $\mathcal{J}$ of $\mathcal{K}^{-\mathsf{Ran}}$ such that $C^{\mathcal{I}} = \overline{C}^{\mathcal{J}}$. We will show that the relation $\mathfrak{S} := \{\, (x_{\overline{E}}, y) \mid y \in E^{\mathcal{I}} \,\}$ is a simulation from $\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}$ to $\mathcal{J}$. Then, $y \in C^{\mathcal{I}}$ implies $(x_{\overline{C}}, y) \in \mathfrak{S}$. Furthermore, $x_{\overline{C}} \in D^{\mathcal{I}_{\mathcal{K}}}$ implies $x_{\overline{C}} \in \overline{D}^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}$ by definition of $\mathcal{I}_{\mathcal{K}}$ and Lemma VII, and so $y \in \overline{D}^{\mathcal{J}}$ by Lemma XIV. Finally, Lemma VI yields $y \in D^{\mathcal{I}}$, and we are done.

It remains to verify that $\mathfrak{S}$ is a nominal-safe simulation.

1. Consider an individual i. It is trivial that $i^{\mathcal{I}} \in \{i\}^{\mathcal{I}}$, and so $(x_{\{i\}}, i^{\mathcal{I}}) \in \mathfrak{S}$. Since $\{i\} \in \overline{\mathbf{S}}$, we have $i^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}} = x_{\{i\}}$. Moreover, $i^{\mathcal{I}} = i^{\mathcal{J}}$ by definition of $\mathcal{J}$. We conclude that $(i^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}, i^{\mathcal{J}}) \in \mathfrak{S}$.

For the other conditions we consider a pair $(x_{\overline{E}}, y) \in \mathfrak{S}$, i.e. $y \in E^{\mathcal{I}}$.

2. Let $x_{\overline{E}} \in A^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}$ for an atomic concept A, i.e. $\overline{E} \sqsubseteq A \in \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$. Proposition IX yields that $\mathcal{K}^{-\mathsf{Ran}} \models \overline{E} \sqsubseteq A$. With $\mathcal{J}$ being a model of $\mathcal{K}^{-\mathsf{Ran}}$ we infer $\overline{E}^{\mathcal{J}} \subseteq A^{\mathcal{J}}$. According to Lemma VI, we have $E^{\mathcal{I}} = \overline{E}^{\mathcal{J}}$, and thus $y \in A^{\mathcal{J}}$.
3. Assume $(x_{\overline{E}}, x_{\overline{F}}) \in r^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}$ for a role r, i.e. $\overline{E} \sqsubseteq \exists r. \overline{F} \in \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$. With Proposition IX we infer $\mathcal{K}^{-\mathsf{Ran}} \models \overline{E} \sqsubseteq \exists r. \overline{F}$ and thus $\overline{E}^{\mathcal{J}} \subseteq (\exists r. \overline{F})^{\mathcal{J}}$. Since $y \in E^{\mathcal{I}}$ and $E^{\mathcal{I}} = \overline{E}^{\mathcal{J}}$ by Lemma VI, there is z with $(y, z) \in r^{\mathcal{J}}$ and $z \in \overline{F}^{\mathcal{J}}$. Since $\overline{F}^{\mathcal{J}} = F^{\mathcal{I}}$ by Lemma VI, the latter implies $(x_{\overline{F}}, z) \in \mathfrak{S}$, and we are done.
4. Consider $x_{\overline{E}} \in (f \leq p)^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}$ for a constraint $f \leq p$. Since $\mathcal{D}_{\mathbf{L}}$ has canonical valuations, we have $f^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}(x_{\overline{E}}) = v_{\Gamma_{\overline{E}}, \mathcal{F}}(f)$, and thus $v_{\Gamma_{\overline{E}}, \mathcal{F}}(f) \leq p$ or rather $v_{\Gamma_{\overline{E}}, \mathcal{F}} \models (f \leq p)$. It follows that $\mathcal{D}_{\mathbf{L}}, \mathcal{F} \models \bigsqcap \Gamma_{\overline{E}} \sqsubseteq (f \leq p)$. Recall that $\Gamma_{\overline{E}} = \{\, g \leq q \mid \overline{E} \sqsubseteq (g \leq q) \in \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}) \,\}$.
   Since $\mathcal{F} \subseteq \mathcal{K}^{-\mathsf{Ran}}$, we have $\mathcal{J} \models \mathcal{F}$. Since $y \in E^{\mathcal{I}}$, we have $y \in \overline{E}^{\mathcal{J}}$. Recall from the proof of Proposition IX that $\mathcal{K}^{-\mathsf{Ran}} \models \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$, i.e. $\mathcal{J} \models \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$. It follows that $y \in (\bigsqcap \Gamma_{\overline{E}})^{\mathcal{J}}$ and thus $y \in (f \leq p)^{\mathcal{J}}$.
5. Last, assume $(x_{\overline{E}}, i^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}) \in r^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}}$ for an individual i. Recall that $\{i\} \in \mathbf{S}$, and therefore $i^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}} = x_{\{i\}}$ and $\overline{E} \sqsubseteq \exists r. \{i\} \in \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$. Since $\mathcal{J}$ is a model of $\mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$ and $E^{\mathcal{I}} = \overline{E}^{\mathcal{J}}$, it follows that $y \in (\exists r. \{i\})^{\mathcal{J}}$, i.e. $(y, i^{\mathcal{J}}) \in r^{\mathcal{J}}$.[10] □

[10] Here we need that $\mathbf{S}$ contains all nominals. Otherwise, when $\{i\} \notin \mathbf{S}$, we would have $i^{\mathcal{I}_{\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}}}} = x_{\top}$ and thus $\overline{E} \sqsubseteq \exists r. \top \in \mathsf{Sat}(\mathcal{K}^{-\mathsf{Ran}}, \overline{\mathbf{S}})$. Thus, we could only infer that $y \in (\exists r. \top)^{\mathcal{J}}$, but not that $(y, i^{\mathcal{J}}) \in r^{\mathcal{J}}$.

Proposition XVI. *$\mathcal{K} \models C \sqsubseteq D$ iff. $\mathcal{I}_\mathcal{K} \models C \sqsubseteq D$ for each $C \in \mathbf{S}$ and for each nominal-safe concept D.*

Proof. Let $\mathcal{K} \models C \sqsubseteq D$ and $x_{\overline{E}} \in C^{\mathcal{I}_\mathcal{K}}$. Then $\mathcal{K} \models E \sqsubseteq C$ by Lemma XV, and thus $\mathcal{K} \models E \sqsubseteq D$. Again by Lemma XV we obtain that $x_{\overline{E}} \in D^{\mathcal{I}_\mathcal{K}}$, as required.

Now let $\mathcal{I}_\mathcal{K} \models C \sqsubseteq D$. Since $\mathcal{K} \models C \sqsubseteq C$, Lemma XV yields $x_{\overline{C}} \in C^{\mathcal{I}_\mathcal{K}}$. It follows that $x_{\overline{C}} \in D^{\mathcal{I}_\mathcal{K}}$, and thus $\mathcal{K} \models C \sqsubseteq D$ by Lemma XV. □

5 Future Prospects

An interesting question for future research is whether non-local feature inclusions $f \leq H(R_1 \circ g_1, \ldots, R_n \circ g_n)$ would lead to undecidability or could be reasoned with, where the R_i are role chains. The operator must then be defined for lists of values, like in the non-local feature inclusion $\mathsf{combinedWealth} \subseteq \sum(\mathsf{hasAccount} \circ \mathsf{balance}) + \sum(\mathsf{holdsAsset} \circ \mathsf{value})$ over the interval domain, which computes the aggregated wealth of a person or company. At first sight, it seems that the undecidability proof for $\mathcal{EL}(\mathcal{D}_{\mathbb{Q}^2,\mathsf{aff}})$ [14] cannot be adapted to this setting. (Mind the braces: $(\mathcal{D})$ instead of $[\mathcal{D}]$ allows for role chains in front of features.) The computation of canonical valuations must then take into account the graph structure induced by the role assertions entailed by the knowledge base.

In order to get rid of the global bounds $\underline{c}$ and $\overline{c}$ in Propositions 16 and 17, linear-program solvers that can work with solution polytopes over the extended reals $\mathbb{R}_+ \cup \{\infty\}$ would be helpful.

It is currently unclear whether the graph domain is admissible w.r.t. cyclic FBoxes. Approaches to solving systems of equations or inequations involving graphs would be necessary to tackle this question.

Since the hierarchical concrete domains are convex by design, they are also appropriate for other Horn logics such as $\mathcal{ELI}$, Horn-$\mathcal{ALC}$, Horn-$\mathcal{SROIQ}$, and existential rules. It would thus be interesting to extend the chase procedure with support for such domains.

Acknowledgements

This work has been supported by Deutsche Forschungsgemeinschaft (DFG) in Project 389792660 (TRR 248: Foundations of Perspicuous Software Systems) and by the Saxon State Ministry for Science, Culture, and Tourism (SMWK) by funding the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI).

References

1. Aiken, A., Kozen, D., Vardi, M.Y., Wimmers, E.L.: The Complexity of Set Constraints. In: Börger, E., Gurevich, Y., Meinke, K. (eds.) Computer Science Logic, 7th Workshop, CSL '93, Swansea, United Kingdom, September 13-17, 1993, Selected Papers. LNCS, vol. 832, pp. 1–17. Springer, Heidelberg (1993). https://doi.org/10.1007/BFB0049320

2. Alrabbaa, C., Baader, F., Borgwardt, S., Koopmann, P., Kovtunova, A.: Combining Proofs for Description Logic and Concrete Domain Reasoning. In: Fensel, A., Ozaki, A., Roman, D., Soylu, A. (eds.) Rules and Reasoning - 7th International Joint Conference, RuleML+RR 2023, Oslo, Norway, September 18-20, 2023, Proceedings. LNCS, vol. 14244, pp. 54–69. Springer, Heidelberg (2023). https://doi.org/10.1007/978-3-031-45072-3_4
3. Baader, F., Borgwardt, S., Lippmann, M.: Query Rewriting for DL-Lite with n-ary Concrete Domains. In: Sierra, C. (ed.) Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pp. 786–792. ijcai.org (2017). https://doi.org/10.24963/IJCAI.2017/109
4. Baader, F., Bortoli, F.D.: The Abstract Expressive Power of First-Order and Description Logics with Concrete Domains. In: Hong, J., Park, J.W. (eds.) Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing, SAC 2024, Avila, Spain, April 8-12, 2024, pp. 754–761. ACM (2024). https://doi.org/10.1145/3605098.3635984
5. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ Envelope. In: Kaelbling, L.P., Saffiotti, A. (eds.) IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005, pp. 364–369. Professional Book Center (2005). http://ijcai.org/Proceedings/05/Papers/0372.pdf
6. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ Envelope Further. In: Clark, K., Patel-Schneider, P.F. (eds.) Proceedings of the Fourth OWLED Workshop on OWL: Experiences and Directions, Washington, DC, USA, 1-2 April 2008. CEUR Workshop Proceedings. CEUR-WS.org (2008). https://ceur-ws.org/Vol-496/owled2008dc%5C_paper%5C_3.pdf
7. Baader, F., De Bortoli, F.: Logics with Concrete Domains: First-Order Properties, Abstract Expressive Power, and (Un)Decidability. SIGAPP Appl. Comput. Rev. **24**(3), 5–17 (2024). https://doi.org/10.1145/3699839.3699840
8. Baader, F., Hanschke, P.: A Scheme for Integrating Concrete Domains into Concept Languages. In: Mylopoulos, J., Reiter, R. (eds.) Proceedings of the 12th International Joint Conference on Artificial Intelligence. Sydney, Australia, August 24-30, 1991, pp. 452–457. Morgan Kaufmann (1991). http://ijcai.org/Proceedings/91-1/Papers/070.pdf
9. Baader, F., Koopmann, P., Kriegel, F., Nuradiansyah, A.: Computing Optimal Repairs of Quantified ABoxes w.r.t. Static $\mathcal{EL}$ TBoxes. In: Platzer, A., Sutcliffe, G. (eds.) Proceedings of the 28th International Conference on Automated Deduction (CADE-28), July 11–16, 2021, Virtual Event, United States. LNCS, vol. 12699, pp. 309–326. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-030-79876-5_18
10. Baader, F., Kriegel, F.: Pushing Optimal ABox Repair from $\mathcal{EL}$ Towards More Expressive Horn-DLs. In: Kern-Isberner, G., Lakemeyer, G., Meyer, T. (eds.) Proceedings of the 19th International Conference on Principles of

Knowledge Representation and Reasoning, KR 2022, Haifa, Israel, July 31 – August 5, 2022, pp. 22–32 (2022). https://doi.org/10.24963/kr.2022/3

11. Baader, F., Küsters, R.: Unification in a Description Logic with Transitive Closure of Roles. In: Nieuwenhuis, R., Voronkov, A. (eds.) Logic for Programming, Artificial Intelligence, and Reasoning, 8th International Conference, LPAR 2001, Havana, Cuba, December 3-7, 2001, Proceedings. LNCS, vol. 2250, pp. 217–232. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45653-8_15
12. Baader, F., Rydval, J.: An Algebraic View on p-Admissible Concrete Domains for Lightweight Description Logics. In: Faber, W., Friedrich, G., Gebser, M., Morak, M. (eds.) Logics in Artificial Intelligence - 17th European Conference, JELIA 2021, Virtual Event, May 17-20, 2021, Proceedings. LNCS, vol. 12678, pp. 194–209. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-030-75775-5_14
13. Baader, F., Rydval, J.: Description Logics with Concrete Domains and General Concept Inclusions Revisited. In: Peltier, N., Sofronie-Stokkermans, V. (eds.) Automated Reasoning - 10th International Joint Conference, IJCAR 2020, Paris, France, July 1-4, 2020, Proceedings, Part I. LNCS, vol. 12166, pp. 413–431. Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-030-51074-9_24
14. Baader, F., Rydval, J.: Using Model Theory to Find Decidable and Tractable Description Logics with Concrete Domains. J. Autom. Reason. **66**(3), 357–407 (2022). https://doi.org/10.1007/S10817-022-09626-2
15. Baader, F., Sattler, U.: Description logics with aggregates and concrete domains. Inf. Syst. **28**(8), 979–1004 (2003). https://doi.org/10.1016/S0306-4379(03)00003-6
16. Bonchi, F., Pous, D.: Checking NFA equivalence with bisimulations up to congruence. In: Giacobazzi, R., Cousot, R. (eds.) The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13, Rome, Italy - January 23 - 25, 2013, pp. 457–468. ACM (2013). see also [27] and [17]. https://doi.org/10.1145/2429069.2429124
17. Bonchi, F., Pous, D.: Hacking nondeterminism with induction and coinduction. Commun. ACM **58**(2), 87–95 (2015). https://doi.org/10.1145/2713167
18. Borgwardt, S., Bortoli, F.D., Koopmann, P.: The Precise Complexity of Reasoning in $\mathcal{ALC}$ with ω-Admissible Concrete Domains. In: Giordano, L., Jung, J.C., Ozaki, A. (eds.) Proceedings of the 37th International Workshop on Description Logics (DL 2024), Bergen, Norway, June 18-21, 2024. CEUR Workshop Proceedings. CEUR-WS.org (2024). https://ceur-ws.org/Vol-3739/paper-1.pdf
19. Brüggemann-Klein, A., Wood, D.: One-Unambiguous Regular Languages. Inf. Comput. **140**(2), 229–253 (1998). https://doi.org/10.1006/INCO.1997.2688
20. Carapelle, C., Turhan, A.: Description Logics Reasoning w.r.t. General TBoxes Is Decidable for Concrete Domains with the EHD-Property. In:

Kaminka, G.A., Fox, M., Bouquet, P., Hüllermeier, E., Dignum, V., Dignum, F., van Harmelen, F. (eds.) ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016). Frontiers in Artificial Intelligence and Applications, pp. 1440–1448. IOS Press (2016). https://doi.org/10.3233/978-1-61499-672-9-1440

21. Charatonik, W., Podelski, A.: Co-definite Set Constraints. In: Nipkow, T. (ed.) Rewriting Techniques and Applications, 9th International Conference, RTA-98, Tsukuba, Japan, March 30 - April 1, 1998, Proceedings. LNCS, vol. 1379, pp. 211–225. Springer, Heidelberg (1998). https://doi.org/10.1007/BFB0052372
22. Devienne, P., Talbot, J., Tison, S.: Solving Classes of Set Constraints with Tree Automata. In: Smolka, G. (ed.) Principles and Practice of Constraint Programming - CP97, Third International Conference, Linz, Austria, October 29 - November 1, 1997, Proceedings. LNCS, vol. 1330, pp. 62–76. Springer, Heidelberg (1997). https://doi.org/10.1007/BFB0017430
23. Greiner, G., Hormann, K.: Efficient Clipping of Arbitrary Polygons. ACM Trans. Graph. **17**(2), 71–83 (1998). https://doi.org/10.1145/274363.274364
24. Gruber, H., Holzer, M.: Finite Automata, Digraph Connectivity, and Regular Expression Size. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part II - Track B: Logic, Semantics, and Theory of Programming & Track C: Security and Cryptography Foundations. LNCS, vol. 5126, pp. 39–50. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70583-3_4
25. Haarslev, V., Lutz, C., Möller, R.: A Description Logic with Concrete Domains and a Role-forming Predicate Operator. J. Log. Comput. **9**(3), 351–384 (1999). https://doi.org/10.1093/LOGCOM/9.3.351
26. Haarslev, V., Möller, R., Wessel, M.: The Description Logic $\mathrm{ALCNH}_{\mathrm{R}+}$ Extended with Concrete Domains: A Practically Motivated Approach. In: Goré, R., Leitsch, A., Nipkow, T. (eds.) Automated Reasoning, First International Joint Conference, IJCAR 2001, Siena, Italy, June 18-23, 2001, Proceedings. LNCS, vol. 2083, pp. 29–44. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45744-5_4
27. Henzinger, T.A., Raskin, J.: The equivalence problem for finite automata: technical perspective. Commun. ACM **58**(2), 86 (2015). https://doi.org/10.1145/2701001
28. Hickey, T.J., Ju, Q., van Emden, M.H.: Interval arithmetic: From principles to implementation. J. ACM **48**(5), 1038–1068 (2001). https://doi.org/10.1145/502102.502106
29. Hopcroft, J.E., Karp, R.M.: A Linear Algorithm for Testing Equivalence of Finite Automata. Tech. rep. TR71-114, Cornell University (1971). https://hdl.handle.net/1813/5958

30. Hovland, D.: The inclusion problem for regular expressions. J. Comput. Syst. Sci. **78**(6), 1795–1813 (2012). `https://doi.org/10.1016/J.JCSS.2011.12.003`
31. Jiang, S., Song, Z., Weinstein, O., Zhang, H.: A faster algorithm for solving general LPs. In: Khuller, S., Williams, V.V. (eds.) STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021, pp. 823–832. ACM (2021). `https://doi.org/10.1145/3406325.3451058`
32. Karakostas, G., Lipton, R.J., Viglas, A.: On the complexity of intersecting finite state automata and NL versus NP. Theor. Comput. Sci. **302**(1-3), 257–274 (2003). `https://doi.org/10.1016/S0304-3975(02)00830-7`
33. Karp, R.M.: Reducibility Among Combinatorial Problems. In: Miller, R.E., Thatcher, J.W. (eds.) Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA. The IBM Research Symposia Series, pp. 85–103. Plenum Press, New York (1972). `https://doi.org/10.1007/978-1-4684-2001-2_9`
34. Kazakov, Y., Krötzsch, M., Simančík, F.: Practical Reasoning with Nominals in the $\mathcal{EL}$ Family of Description Logics. In: Brewka, G., Eiter, T., McIlraith, S.A. (eds.) Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012. AAAI Press (2012). `http://www.aaai.org/ocs/index.php/KR/KR12/paper/view/4540`
35. Kazakov, Y., Krötzsch, M., Simančík, F.: The Incredible ELK - From Polynomial Procedures to Efficient Reasoning with $\mathcal{EL}$ Ontologies. J. Autom. Reason. **53**(1), 1–61 (2014). `https://doi.org/10.1007/S10817-013-9296-3`
36. Kozen, D.: Lower Bounds for Natural Proof Systems. In: 18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977, pp. 254–266. IEEE Computer Society (1977). `https://doi.org/10.1109/SFCS.1977.16`
37. Kriegel, F.: Beyond Optimal: Interactive Identification of Better-than-optimal Repairs. In: Proceedings of the 40th ACM/SIGAPP Symposium On Applied Computing (SAC 2025). Association for Computing Machinery (2025). To appear. `https://doi.org/10.1145/3672608.3707750`
38. Kriegel, F.: Optimal Fixed-Premise Repairs of $\mathcal{EL}$ TBoxes. In: Bergmann, R., Malburg, L., Rodermund, S.C., Timm, I.J. (eds.), pp. 115–130. Springer (2022). `https://doi.org/10.1007/978-3-031-15791-2_11`
39. Kunc, M.: Largest solutions of left-linear language inequalities. In: Ésik, Z., Fülöp, Z. (eds.) Proceedings of the 11th International Conference on Automata and formal languages (AFL 2005), Dobogókő, Hungary, May 17–20, 2005, pp. 178–186 (2005). `http://www.math.muni.cz/~kunc/math/left_linear.ps`

40. Losemann, K., Martens, W., Niewerth, M.: Closure properties and descriptional complexity of deterministic regular expressions. Theor. Comput. Sci. **627**, 54–70 (2016). https://doi.org/10.1016/J.TCS.2016.02.027
41. Lutz, C.: NEXPTIME-complete description logics with concrete domains. ACM Trans. Comput. Log. **5**(4), 669–705 (2004). https://doi.org/10.1145/1024922.1024925
42. Lutz, C.: The complexity of description logics with concrete domains. PhD thesis, RWTH Aachen University, Germany (2002). http://nbn-resolving.org/urn:nbn:de:hbz:82-opus-3032.
43. Lutz, C., Areces, C., Horrocks, I., Sattler, U.: Keys, Nominals, and Concrete Domains. J. Artif. Intell. Res. **23**, 667–726 (2005). https://doi.org/10.1613/JAIR.1542
44. Lutz, C., Miličić, M.: A Tableau Algorithm for Description Logics with Concrete Domains and General TBoxes. J. Autom. Reason. **38**(1-3), 227–259 (2007). https://doi.org/10.1007/S10817-006-9049-7
45. Magka, D., Kazakov, Y., Horrocks, I.: Tractable Extensions of the Description Logic $\mathcal{EL}$ with Numerical Datatypes. J. Autom. Reason. **47**(4), 427–450 (2011). https://doi.org/10.1007/S10817-011-9235-0
46. Martínez, F., Ogáyar, C.J., Jiménez, J., Ruiz, A.J.R.: A simple algorithm for Boolean operations on polygons. Adv. Eng. Softw. **64**, 11–19 (2013). https://doi.org/10.1016/J.ADVENGSOFT.2013.04.004
47. O'Rourke, J., Chien, C., Olson, T., Naddor, D.: A new linear algorithm for intersecting convex polygons. Comput. Graph. Image Process. **19**(4), 384–391 (1982). https://doi.org/10.1016/0146-664X(82)90023-5
48. Pan, J.Z., Horrocks, I.: Reasoning in the SHOQ($\mathrm{D_n}$) Description Logic. In: Horrocks, I., Tessaris, S. (eds.) Proceedings of the 2002 International Workshop on Description Logics (DL2002), Toulouse, France, April 19-21, 2002. CEUR Workshop Proceedings. CEUR-WS.org (2002). https://ceur-ws.org/Vol-53/Pan-Horrocks-shoqdn-2002.ps
49. Rydval, J.: Using Model Theory to Find Decidable and Tractable Description Logics with Concrete Domains. PhD thesis, Dresden University of Technology, Germany (2022). https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-799074.
50. Shamos, M.I.: Computational Geometry. PhD thesis, Yale University, United States (1978). http://euro.ecom.cmu.edu/people/faculty/mshamos/1978ShamosThesis.pdf.
51. Stockmeyer, L.J., Meyer, A.R.: Word Problems Requiring Exponential Time: Preliminary Report. In: Aho, A.V., Borodin, A., Constable, R.L., Floyd, R.W., Harrison, M.A., Karp, R.M., Strong, H.R. (eds.) Proceedings of the 5th Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1973, Austin, Texas, USA, pp. 1–9. ACM (1973). https://doi.org/10.1145/800125.804029
52. Tarski, A.: A lattice-theoretical fixpoint theorem and its applications. Pacific Journal of Mathematics **5**(2), 285–309 (1955). https://doi.org/10.2140/pjm.1955.5.285

53. Toussaint, G.T.: A simple linear algorithm for intersecting convex polygons. Vis. Comput. **1**(2), 118–123 (1985). https://doi.org/10.1007/BF01898355
54. Vatti, B.R.: A Generic Solution to Polygon Clipping. Commun. ACM **35**(7), 56–63 (1992). https://doi.org/10.1145/129902.129906
55. Yu, S., Zhuang, Q., Salomaa, K.: The State Complexities of Some Basic Operations on Regular Languages. Theor. Comput. Sci. **125**(2), 315–328 (1994). https://doi.org/10.1016/0304-3975(92)00011-F