Technische Universität Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

# LTCS–Report

## Optimal ABox Repair w.r.t. Static $\mathcal{EL}$ TBoxes: from Quantified ABoxes back to ABoxes (Extended Version)

Franz Baader, Patrick Koopmann,
Francesco Kriegel, Adrian Nuradiansyah

LTCS-Report 22-01

Updated on March 11, 2022. This is an extended version of an article accepted at ESWC 2022.

Postal Address:
Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
01062 Dresden

http://lat.inf.tu-dresden.de

Visiting Address:
Nöthnitzer Str. 46
Dresden

# Optimal ABox Repair w.r.t. Static $\mathcal{EL}$ TBoxes: from Quantified ABoxes back to ABoxes (Extended Version)

Franz Baader, Patrick Koopmann, Francesco Kriegel, and Adrian Nuradiansyah

Theoretical Computer Science, TU Dresden, Dresden, Germany
firstname.lastname@tu-dresden.de

**Abstract.** Errors in Description Logic (DL) ontologies are often detected when a reasoner computes unwanted consequences. The question is then how to repair the ontology such that the unwanted consequences no longer follow, but as many of the other consequences as possible are preserved. The problem of computing such optimal repairs was addressed in our previous work in the setting where the data (expressed by an ABox) may contain errors, but the schema (expressed by an $\mathcal{EL}$ TBox) is assumed to be correct. Actually, we consider a generalization of ABoxes called quantified ABoxes (qABoxes) both as input for and as result of the repair process. Using qABoxes for repair allows us to retain more information, but the disadvantage is that standard DL systems do not accept qABoxes as input. This raises the question, investigated in the present paper, whether and how one can obtain optimal repairs if one restricts the output of the repair process to being ABoxes. In general, such optimal ABox repairs need not exist. Our main contribution is that we show how to decide the existence of optimal ABox repairs in exponential time, and how to compute all such repairs in case they exist.

## 1 Introduction

Description Logics (DLs) [2] are a successful family of logic-based knowledge representation languages, which are employed in various application domains, but arguably their most prominent success was the adoption of the DL-based language OWL[1] as the standard ontology language for the Semantic Web. A DL knowledge base (aka ontology) consists of a TBox and an ABox. In the former, concepts can be used to state terminological constraints as so-called general concept inclusions (GCIs). For example, the concept $\exists \mathit{parent}.(\mathit{Famous} \sqcap \mathit{Rich})$ describes individuals that have a parent that is both famous and rich, and the GCI $\exists \mathit{friend}.\mathit{Famous} \sqsubseteq \mathit{Famous}$ states that individuals that have a famous friend are famous themselves. The expressiveness of a DL depends on which constructors for building concepts are available. The concepts in our example use the constructors conjunction ($\sqcap$) and existential restriction ($\exists r.C$), which together

[1] https://www.w3.org/OWL/

with the top concept ($\top$) are the ones available in the DL $\mathcal{EL}$, to which we restrict our attention here. While being quite inexpressive, $\mathcal{EL}$ is nevertheless frequently used for building ontologies,[2] and it has the advantage over more expressive DLs that reasoning is polynomial w.r.t. $\mathcal{EL}$ ontologies. In the ABox, one can relate named individuals with concepts and with each other. For example, the concept assertion $(\exists \mathit{parent}.\mathit{Rich})(\mathit{BEN})$ states that Ben has a rich parent, and the role assertion $\mathit{friend}(\mathit{BEN}, \mathit{JOHN})$ says that Ben has John as friend. If concept assertions are restricted to employing only concept names, like $\mathit{Famous}(\mathit{JOHN})$, rather than complex concepts, then the ABox is called simple. DL systems provide their users with inference services that automatically derive implicit consequences such as instance relationships. For example, given the ABox assertions and the GCI introduced above, we can derive that Ben is famous, i.e., that the assertions $\mathit{Famous}(\mathit{BEN})$ follows from this ontology.

Although DL reasoners are usually sound (i.e., only derive instance relationship that indeed follow from the ontology), a computed consequence may still be incorrect in the application domain, due to the fact that the modelling of the domain in the ontology is erroneous. The question is then how to repair the ontology such that one gets rid of the unwanted consequences, but retains as many consequences as possible. Classical repair approaches that are based on removing axioms from the ontology [8,11,15,16,18,19] are not optimal since, by removing large axioms, one may also lose information that does not contribute to the unwanted consequence. For example, if the concept assertion for John is $(\mathit{Famous} \sqcap \mathit{Rich})(\mathit{JOHN})$ rather than just $\mathit{Famous}(\mathit{JOHN})$, then to get rid of the consequence $\mathit{Famous}(\mathit{BEN})$ we need to remove the whole assertion, and thus unnecessarily also lose the information that John is rich.

Extending on our previous work in [5,7], we investigated in [3] how to compute optimal repairs in a setting where the ABox may contain errors, but the TBox is assumed to be a correct $\mathcal{EL}$ TBox, and thus remains unchanged. More precisely, we consider a generalization of ABoxes called quantified ABoxes (qABoxes) both as input for and as result of the repair process since this allows us to retain more consequences. Such a qABox is a simple ABox where, however, some of the individuals are anonymized, which is formally expressed by existentially quantifying over them. In [3], we introduce two different notions of repair, depending on which entailment relation between qABoxes is considered: classical logical entailment or $\mathsf{IQ}$-entailment, where the latter retains as many instance relationships as possible (but not necessarily answers to conjunctive queries). For the $\mathsf{IQ}$ case, we show that optimal $\mathsf{IQ}$-repairs always exist and can be computed in exponential time. In the worst case, such repairs may be exponentially large and there may be exponentially many of them. Reusing an example from the introduction of [3], let us assume that the input ABox contains the information that Ben has a parent, Jerry, that is both rich and famous, that the TBox contains the GCI $\mathit{Famous} \sqsubseteq \mathit{Rich}$, and that we want to remove the consequence $(\exists \mathit{parent}.(\mathit{Rich} \sqcap \mathit{Famous}))(\mathit{BEN})$. Using the optimized repair ap-

[2] For example, the large medical ontology SNOMED CT is an $\mathcal{EL}$ ontology.

proach of [3], we obtain the following qABox as one of the optimal IQ-repairs: $\exists\{y\}.\{\mathit{parent}(\mathit{BEN}, y), \mathit{Rich}(y), \mathit{Famous}(\mathit{JERRY}), \mathit{Rich}(\mathit{JERRY})\}$.

The advantage of using qABoxes rather than ABoxes for repair is that more information can be retained (e.g. the fact that Ben has a rich parent). The disadvantage is that, though anonymized individuals are part of the OWL standard, DL systems usually do not accept them as input. Thus, the question arises whether one can also obtain optimal repairs if one restricts the output of the repair process to being ABoxes. In the above example, the qABox obtained as an optimal IQ-repair can actually be expressed by an ABox with complex concept assertions: $\{(\exists \mathit{parent}.\mathit{Rich})(\mathit{BEN}), \mathit{Famous}(\mathit{JERRY}), \mathit{Rich}(\mathit{JERRY})\}$.

However, this is not always the case. As an example, consider the ABox $\mathcal{A} := \{\mathit{parent}(\mathit{BEN}, \mathit{JERRY}), \mathit{Rich}(\mathit{JERRY})\}$ and the TBox $\mathcal{T} := \{\exists \mathit{parent}.\mathit{Rich} \sqsubseteq \mathit{Famous}, \mathit{Famous} \sqsubseteq \exists \mathit{friend}.\mathit{Famous}, \exists \mathit{friend}.\mathit{Famous} \sqsubseteq \mathit{Famous}\}$, which together imply that Ben is famous. Assume that Ben wants to get rid of this consequence. The repair approach of [3] yields the following qABox as an optimal IQ-repair: $\exists\{x, y\}.\{\mathit{parent}(\mathit{BEN}, x), \mathit{Rich}(\mathit{JERRY}), \mathit{friend}(\mathit{BEN}, y), \mathit{friend}(y, y)\}$. This qABox retains the information that Ben has a parent (but not that Jerry is this parent) and that Ben is the starting point of an infinite *friend*-chain, i.e., Ben belongs to the concepts $C_n := (\exists \mathit{friend}.)^n\top$ for all $n \geq 1$. The latter is the reason why this qABox cannot be expressed by an IQ-equivalent ABox, which in turn is the reason why there is no optimal ABox repair. The culprit is obviously the cycle $\mathit{friend}(y, y)$. However, such cycles need not always cause problems. In fact, if we remove the third GCI $\exists \mathit{friend}.\mathit{Famous} \sqsubseteq \mathit{Famous}$ from the TBox, then the following qABox is an optimal IQ-repair:

$$\exists\{x, y\}.\{\mathit{parent}(\mathit{BEN}, x), \mathit{Rich}(\mathit{JERRY}), \\ \mathit{friend}(\mathit{BEN}, y), \mathit{friend}(y, y), \mathit{Famous}(y)\}.$$

This qABox can be expressed by an ABox that is IQ-equivalent to it w.r.t. the given TBox: $\{(\exists \mathit{parent}.\top)(\mathit{BEN}), \mathit{Rich}(\mathit{JERRY}), (\exists \mathit{friend}.\mathit{Famous})(\mathit{BEN})\}$. The reason is that, due to the existence of a famous friend of Ben, the GCI $\mathit{Famous} \sqsubseteq \exists \mathit{friend}.\mathit{Famous}$ now yields the infinite *friend*-chain.

These examples demonstrate that optimal ABox repairs may not always exist, and that it is not obvious to see when they do. The main contribution of the present paper is that we show how to decide the existence of optimal ABox repairs in exponential time, and how to compute all such repairs in case they exist. There may exist exponentially many such repairs, and each one may in the worst case be of double-exponential size. Our approach for showing these results roughly proceeds as follows. First, we observe that classical entailment between a qABox and an ABox coincides with so-called IRQ-entailment, which is slightly stronger than IQ-entailment by additionally taking role assertions between named individuals into account. Then, we show that both the canonical and the optimized IQ-repairs of [3] cannot only be used to obtain all optimal IQ-repairs, but also to compute all optimal IRQ-repairs. Subsequently, we introduce the notion of an optimal ABox approximation of a given qABox, and prove that the set of optimal ABox approximations of all optimal IRQ-repairs yields all optimal ABox repairs. A given qABox may not have an optimal ABox approximation, but if it does, then this approximation is unique up to equivalence and of at most exponential size. Then we investigate the problem of deciding the existence of optimal ABox approximations. The first step is to transfer the qABox into a specific form, called pre-approximation, which is saturated w.r.t. the TBox and consists of the original role assertions between named individuals and for each named individual $a$ a sub-qABox $\mathcal{B}_a$. We prove that the original qABox has an optimal ABox approximation iff all the named individuals $a$ have a most specific concept $C_a$ in $\mathcal{B}_a$ w.r.t. the TBox. The optimal ABox approximation is then obtained by replacing each $\mathcal{B}_a$ with $C_a(a)$ in the pre-approximation. We can then use the results stated in [20] to test the existence of the msc in polynomial time[3] and to generate the at most exponentially large msc. Given that the optimal $\mathsf{IRQ}$-repairs may be of exponential size, this yields the complexity upper bounds for testing the existence and computing optimal ABox repairs mentioned above. Due to space constraints for the submission to ESWC 2022, we could not give complete proofs of all our results there. The missing proofs can be found in this extended version after the references, starting on page 19.

## 2 Preliminaries

We start with introducing the DL $\mathcal{EL}$ as well as TBoxes and (quantified) ABoxes. Then we consider the entailment relations relevant for this paper.

The name space available for defining $\mathcal{EL}$ concepts and ABox assertions is given by a *signature* $\Sigma$, which is the disjoint union of sets $\Sigma_\mathsf{O}$, $\Sigma_\mathsf{C}$, and $\Sigma_\mathsf{R}$ of *object names*, *concept names*, and *role names*. Starting with concept names and the top concept $\top$, $\mathcal{EL}$ *concepts* are defined inductively: if $C, D$ are $\mathcal{EL}$ concepts and $r$ is a role name, then $C \sqcap D$ (conjunction) and $\exists r.C$ (existential restriction) are also $\mathcal{EL}$ concepts. An $\mathcal{EL}$ *general concept inclusion (GCI)* is of the form $C \sqsubseteq D$, an $\mathcal{EL}$ *concept assertion* is of the form $C(u)$, and a *role assertion* is of the form $r(u, v)$, where $C, D$ are $\mathcal{EL}$ concepts, $r \in \Sigma_\mathsf{R}$, and $u, v \in \Sigma_\mathsf{O}$. An $\mathcal{EL}$ *TBox* is a finite set of $\mathcal{EL}$ GCIs and an $\mathcal{EL}$ *ABox* is a finite set of $\mathcal{EL}$ concept assertions and role assertions. Such an ABox is called *simple* if all its concept assertions are of the form $A(u)$ with $A \in \Sigma_\mathsf{C}$. A *quantified ABox (qABox)* is of the form $\exists X.\mathcal{A}$ where $X$ is a finite subset of $\Sigma_\mathsf{O}$ and $\mathcal{A}$ is a simple ABox, which we call the *matrix* of $\exists X.\mathcal{A}$. We call the elements of $X$ *variables* and the other object names occurring in $\mathcal{A}$ *individuals*.[4] The set of individual names occurring in $\exists X.\mathcal{A}$ is denoted with $\Sigma_\mathsf{I}(\exists X.\mathcal{A})$, and the set of all object names (including the variables) with $\Sigma_\mathsf{O}(\exists X.\mathcal{A})$.

The semantics of the syntactic entities introduced above can either be defined directly using interpretations, or by a translation into first-order logic (FO). For the sake of brevity, we choose the latter approach (see [3] for the former). In the

[3] The proof for this polynomiality result in [20] is actually incorrect, but we show how to correct it.

[4] The variables correspond to what we have called anonymized individuals in the introduction, and the individuals to what we have called named individuals.

translation, the elements of $\Sigma_\mathsf{O}$, $\Sigma_\mathsf{C}$, and $\Sigma_\mathsf{R}$ are respectively viewed as constant symbols, unary predicate symbols, and binary predicate symbols. $\mathcal{EL}$ concepts $C$ are inductively translated into FO formulas $\phi_C(x)$ with one free variable $x$:

- concept $A$ for $A \in \Sigma_\mathsf{C}$ is translated into $A(x)$ and $\top$ into $A(x) \vee \neg A(x)$ for an arbitrary $A \in \Sigma_\mathsf{C}$;
- if $C, D$ are translated into $\phi_C(x)$ and $\phi_D(x)$, then $C \sqcap D$ is translated into $\phi_C(x) \wedge \phi_D(x)$ and $\exists r.C$ into $\exists y.(r(x,y) \wedge \phi_D(y))$, where $\phi_D(y)$ is obtained from $\phi_D(x)$ by replacing the free variable $x$ by a different variable $y$.

GCIs $C \sqsubseteq D$ are translated into sentences $\phi_{C\sqsubseteq D} := \forall x.(\phi_C(x) \to \phi_D(x))$ and TBoxes $\mathcal{T}$ into $\phi_\mathcal{T} := \bigwedge_{C\sqsubseteq D\in\mathcal{T}} \phi_{C\sqsubseteq D}$. Concept assertions $C(u)$ are translated into $\phi_C(u)$, role assertions $r(u,v)$ stay the same, and ABoxes $\mathcal{A}$ are translated into the conjunction $\phi_\mathcal{A}$ of the translations of their assertions. For a quantified ABox $\exists X.\mathcal{A}$, the elements of $X$ are viewed as first-order variables rather than constants, and its translation is $\exists\vec{x}.\phi_\mathcal{A}$, where $\vec{x}$ is the tuple of the variables in $X$ in arbitrary order.

Let $\alpha, \beta$ be (q)ABoxes, concept inclusions, or concept assertions (possibly not both of the same kind), and $\mathcal{T}$ an $\mathcal{EL}$ TBox. Then we say that $\alpha$ *entails* $\beta$ *w.r.t.* $\mathcal{T}$ (written $\alpha \models^\mathcal{T} \beta$) if the implication $(\phi_\alpha \wedge \phi_\mathcal{T}) \to \phi_\beta$ is valid according to the semantics of FO. Furthermore, $\alpha$ and $\beta$ are *equivalent w.r.t.* $\mathcal{T}$ (written $\alpha \equiv^\mathcal{T} \beta$), if $\alpha \models^\mathcal{T} \beta$ and $\beta \models^\mathcal{T} \alpha$. In case $\mathcal{T} = \emptyset$, we will sometimes write $\models$ instead of $\models^\emptyset$. If $\emptyset \models^\mathcal{T} C \sqsubseteq D$, then we also write $C \sqsubseteq^\mathcal{T} D$ and say that $C$ *is subsumed by* $D$ *w.r.t.* $\mathcal{T}$; in case $\mathcal{T} = \emptyset$ we simply say that $C$ is subsumed by $D$. If $\exists X.\mathcal{A} \models^\mathcal{T} C(a)$, then $a$ is called an *instance of* $C$ w.r.t. $\exists X.\mathcal{A}$ and $\mathcal{T}$. For ABoxes, the instance relation is defined analogously. Entailment between qABoxes w.r.t. an $\mathcal{EL}$ TBox is NP-complete, but the subsumption and the instance problem are polynomial [7].

Note that ABoxes are a special case of qABoxes. For simple ABoxes, this is the case where $X = \emptyset$. For general ABoxes, one can express complex concept assertions by introducing existentially quantified variables (e.g., $\{(A \sqcap \exists r.B)(a)\}$ is equivalent to $\exists\{x\}.\{A(a), r(a,x), B(x)\}$). For this reason, the entailment relations defined below for qABoxes are also well-defined for ABoxes.

**$\mathsf{IQ}$-entailment** If one is mainly interested in asking instance queries, i.e., in what kind of instance relations a qABox entails, then the following weaker form of entailment can be used [3, 7]. We say that the qABox $\exists X.\mathcal{A}$ $\mathsf{IQ}$-*entails* the qABox $\exists Y.\mathcal{B}$ w.r.t. the $\mathcal{EL}$ TBox $\mathcal{T}$ (written $\exists X.\mathcal{A} \models^\mathcal{T}_\mathsf{IQ} \exists Y.\mathcal{B}$) if every concept assertion $C(a)$ entailed w.r.t. $\mathcal{T}$ by the latter is also entailed w.r.t. $\mathcal{T}$ by the former. Whenever we compare two qABoxes $\exists X.\mathcal{A}$ and $\exists Y.\mathcal{B}$, we follow [7] and assume without loss of generality that they are *renamed apart*, which means that $X$ is disjoint with $\Sigma_\mathsf{O}(\exists Y.\mathcal{B})$ and $Y$ is disjoint with $\Sigma_\mathsf{O}(\exists X.\mathcal{A})$, and we further assume that the two qABoxes speak about the same set of individual names $\Sigma_\mathsf{I} := \Sigma_\mathsf{I}(\exists X.\mathcal{A}) \cup \Sigma_\mathsf{I}(\exists Y.\mathcal{B})$.

For the case of an empty TBox, it was shown in [7] that $\exists X.\mathcal{A} \models^\emptyset_\mathsf{IQ} \exists Y.\mathcal{B}$ iff there is a simulation from $\exists Y.\mathcal{B}$ to $\exists X.\mathcal{A}$. A *simulation* from $\exists Y.\mathcal{B}$ to $\exists X.\mathcal{A}$ is

**$\sqcap$-rule.** If $(C_1 \sqcap \ldots \sqcap C_n)(t) \in \mathcal{A}$, then remove this assertion from $\mathcal{A}$ and add the assertions $C_1(t), \ldots, C_n(t)$ to $\mathcal{A}$.

**$\exists$-rule.** If $(\exists r. C)(t) \in \mathcal{A}$, then remove this assertion from $\mathcal{A}$, add the two assertions $r(t, x_C)$ and $C(x_C)$ to $\mathcal{A}$, and add $x_C$ to $X$ if it is not already there.

**$\sqsubseteq$-rule.** If $t \in \Sigma_{\mathsf{O}}(\exists X. \mathcal{A})$, $C \sqsubseteq D \in \mathcal{T}$, $\mathcal{A} \models C(t)$, and $\mathcal{A} \not\models D(t)$, then add the assertion $D(t)$ to $\mathcal{A}$.

The $\sqcap$-rule has higher precedence than the $\exists$-rule, and the latter has higher precedence than the $\sqsubseteq$-rule.

Fig. 1: The IQ-saturation rules from [3].

a relation $\mathfrak{S} \subseteq \Sigma_{\mathsf{O}}(\exists Y. \mathcal{B}) \times \Sigma_{\mathsf{O}}(\exists X. \mathcal{A})$ such that $(a, a) \in \mathfrak{S}$ for each $a \in \Sigma_{\mathsf{I}}$ and, for each $(u, v) \in \mathfrak{S}$, $A(u) \in \mathcal{B}$ implies $A(v) \in \mathcal{A}$ and $r(u, u') \in \mathcal{B}$ implies that there exists an object $v' \in \Sigma_{\mathsf{O}}(\exists X. \mathcal{A})$ such that $(u', v') \in \mathfrak{S}$ and $r(v, v') \in \mathcal{A}$. Since checking the existence of a simulation can be done in polynomial time [10], the simulation characterization of IQ-entailment shows that IQ-entailment between qABoxes can be decided in polynomial time if $\mathcal{T} = \emptyset$ [7].

To extend these results to the case of a non-empty TBox, the notion of an IQ-saturation is introduced in [3]. The saturation rules given in Fig. 1 add new variables and assertions to the qABox if the existence of a corresponding element and the validity of the assertion is implied by the TBox. To be more precise, for each existential restriction $\exists r. C$ occurring in $\mathcal{T}$, a fresh variable $x_C$ not contained in the initial qABox is introduced. When applying the $\exists$-rule to an assertion of the form $(\exists r. C)(t)$, this variable is always used for the successor object. As pointed out in [3], IQ-saturation (i.e., the exhaustive application of the IQ-saturation rules) terminates in polynomial time and generates a qABox $\mathsf{sat}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X. \mathcal{A})$, which can be seen as a qABox representation of what is called the *canonical model* in [13, Section 5.2]. IQ-entailment for qABoxes w.r.t. an $\mathcal{EL}$ TBox is now characterized in [3] as follows.

**Theorem 1 ([3]).** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\exists X. \mathcal{A}$ and $\exists Y. \mathcal{B}$ qABoxes. Then the following statements are equivalent:*

- $\exists X. \mathcal{A} \models_{\mathsf{IQ}}^{\mathcal{T}} \exists Y. \mathcal{B}$,
- $\mathsf{sat}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X. \mathcal{A}) \models_{\mathsf{IQ}}^{\emptyset} \exists Y. \mathcal{B}$,
- *there is a simulation from $\exists Y. \mathcal{B}$ to $\mathsf{sat}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X. \mathcal{A})$.*

Since the IQ-saturation can be computed in polynomial time, this clearly shows that IQ-entailment for qABoxes w.r.t. an $\mathcal{EL}$ TBox can also be decided in polynomial time.

**IRQ-entailment** If we are not only interested in implied concept assertions, but also in implied role assertions, then IQ-entailment is not sufficient. Instead, we must use IRQ-entailment. We say that the qABox $\exists X. \mathcal{A}$ IRQ-*entails* the qABox $\exists Y. \mathcal{B}$ w.r.t. the $\mathcal{EL}$ TBox $\mathcal{T}$ (written $\exists X. \mathcal{A} \models_{\mathsf{IRQ}}^{\mathcal{T}} \exists Y. \mathcal{B}$) if every concept or role assertion entailed w.r.t. $\mathcal{T}$ by the latter is also entailed w.r.t. $\mathcal{T}$ by the former.

It is easy to see that a qABox cannot entail a role assertion involving a variable, and it can only entail a role assertion between individuals if its matrix contains this assertion. This yields the following characterization of $\mathsf{IRQ}$-entailment, which shows that $\mathsf{IRQ}$-entailment can be decided in polynomial time.

**Proposition 2.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\exists X.\mathcal{A}$ and $\exists Y.\mathcal{B}$ qABoxes. Then the following statements are equivalent:*

- $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IRQ}} \exists Y.\mathcal{B}$,
- $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IQ}} \exists Y.\mathcal{B}$ *and* $r(a,b) \in \mathcal{B}$ *implies* $r(a,b) \in \mathcal{A}$ *for all* $r \in \Sigma_{\mathsf{R}}$ *and* $a, b \in \Sigma_{\mathsf{I}}$.

Since ABoxes consist of concept and role assertions, we obtain the following characterization of entailment between a qABox and an ABox, which implies that this entailment can be decided in polynomial time.

**Proposition 3.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, $\exists X.\mathcal{A}$ a qABox, and $\mathcal{B}$ an ABox. Then $\exists X.\mathcal{A} \models^{\mathcal{T}} \mathcal{B}$ iff $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{B}$.*

## 3 Optimal ABox repairs and approximations

We first introduce the notion of an optimal repair w.r.t. an entailment relation, and show that the approaches for computing optimal $\mathsf{IQ}$-repairs described in [3] can also be used to compute optimal $\mathsf{IRQ}$-repairs. Then, we define optimal ABox approximations and show some useful properties for them. Finally, we introduce optimal ABox repairs, and describe how optimal ABox approximations can be used to obtain them from optimal $\mathsf{IRQ}$-repairs.

### 3.1 Optimal $\mathsf{IQ}$- and $\mathsf{IRQ}$-repairs

We start by recalling the definition of optimal repairs given in [3], but consider $\mathsf{IRQ}$ as an additional entailment relation.

**Definition 4.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and $\mathsf{QL} \in \{\mathsf{IRQ}, \mathsf{IQ}\}$.*

- *An $\mathcal{EL}$* repair request *is a finite set of $\mathcal{EL}$ concept assertions.*
- *Given a qABox $\exists X.\mathcal{A}$ and an $\mathcal{EL}$ repair request $\mathcal{R}$, a* $\mathsf{QL}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ *is a qABox $\exists Y.\mathcal{B}$ such that $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{QL}} \exists Y.\mathcal{B}$ and $\exists Y.\mathcal{B} \not\models^{\mathcal{T}} C(a)$ for all $C(a) \in \mathcal{R}$.*
- *Such a repair $\exists Y.\mathcal{B}$ is* optimal *if there is no $\mathsf{QL}$-repair $\exists Z.\mathcal{C}$ of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ such that $\exists Z.\mathcal{C} \models^{\mathcal{T}}_{\mathsf{QL}} \exists Y.\mathcal{B}$ and $\exists Y.\mathcal{B} \not\models^{\mathcal{T}}_{\mathsf{QL}} \exists Z.\mathcal{C}$.*

Two qABoxes are $\mathsf{QL}$-*equivalent* if they $\mathsf{QL}$-entail each other, and $\exists X.\mathcal{A}$ *strictly* $\mathsf{QL}$-*entails* $\exists Y.\mathcal{B}$ if $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{QL}} \exists Y.\mathcal{B}$ and $\exists Y.\mathcal{B} \not\models^{\mathcal{T}}_{\mathsf{QL}} \exists X.\mathcal{A}$. We say that a set $\mathfrak{R}$ of $\mathsf{QL}$-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ $\mathsf{QL}$-*covers* all $\mathsf{QL}$-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ if for every $\mathsf{QL}$-repair $\exists Y.\mathcal{B}$ of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ there exists an element $\exists Z.\mathcal{C}$ of $\mathfrak{R}$ such that $\exists Z.\mathcal{C} \models^{\mathcal{T}}_{\mathsf{QL}} \exists Y.\mathcal{B}$. It is easy to see that such a covering set $\mathfrak{R}$ must contain, up to QL-equivalence, all optimal QL-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$, and thus one can obtain from it, up to QL-equivalence, the set of all optimal QL-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ by removing elements that are strictly QL-entailed by another element. Clearly, this set still QL-covers all QL-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$.

In [3], two ways of computing such a covering set for IQ-repairs are described, the canonical IQ-repairs and the optimized IQ-repairs (see Proposition 8 and Theorem 14). Since these covering sets are of at most exponential cardinality, their elements are of at most exponential size, and IQ-entailment can be decided in polynomial time, this shows that, up to IQ-equivalence, the set of all optimal IQ-repairs can be computed in exponential time.

The canonical (optimized) IQ-repairs also yield covering sets for the IRQ case. The reason is basically that the approaches for constructing them introduced in [3] do not generate new role assertions between individuals and preserve as many of them as possible, although this is not required for IQ-entailment.

**Proposition 5.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, $\exists X.\mathcal{A}$ a qABox, and $\mathcal{R}$ an $\mathcal{EL}$ repair request. If $\mathfrak{R}$ is the set of all canonical or all optimized* IQ*-repairs obtained from this input according to the definitions in [3], then $\mathfrak{R}$ is a set of* IRQ*-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ that* IRQ*-covers all* IRQ*-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. In particular, up to* IRQ*-equivalence, the set of optimal* IRQ*-repairs can be computed in exponential time, and it* IRQ*-covers all* IRQ*-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$.*

Note that, though we have the same covering set $\mathfrak{R}$ in the IQ and in the IRQ case, the sets of optimal repairs obtained from it by removing strictly entailed elements need not coincide since different entailment relations are used during this removal. Since the requirements for IQ entailment are weaker than for IRQ entailment, it could be that a qABox may be removed from $\mathfrak{R}$ in the IQ case, but must be retained in IRQ case. Also notice that the proposition need not hold for arbitrary IQ-covering sets. Its proof uses properties of the canonical and the optimized IQ-repairs that need not hold for arbitrary covering sets.

*Example 6.* Consider the qABox $\exists\{x\}.\mathcal{A}$ for $\mathcal{A} = \{A(a), r(a,x), r(x,x)\}$, assume that the TBox is empty, and that the repair request is $\{A(a)\}$. An optimal IQ-repair $\exists\{x\}.\mathcal{A}'$ can be obtained from this qABox by removing the assertion $A(a)$ from $\mathcal{A}$, and this is also an optimal IRQ-repair. However, the ABox $\{r(a,a)\}$ is also an optimal IQ-repair since it is IQ-equivalent to $\exists\{x\}.\mathcal{A}'$, but it is not even an IRQ-repair since it is not IRQ-entailed by $\exists\{x\}.\mathcal{A}$.

### 3.2 Optimal ABox approximations

Given a qABox we are now interested in finding an ABox that approximates it as closely as possible in the sense that a minimal amount of information is lost. In the definition below, we use classical entailment. But note that, according to Proposition 3, this coincides with IRQ-entailment.

**Definition 7.** *Given a qABox* $\exists X.\mathcal{A}$ *and an* $\mathcal{EL}$ *TBox* $\mathcal{T}$*, we call an* $\mathcal{EL}$ *ABox* $\mathcal{B}$ *an* ABox approximation *of* $\exists X.\mathcal{A}$ *w.r.t.* $\mathcal{T}$ *if* $\exists X.\mathcal{A} \models^{\mathcal{T}} \mathcal{B}$*. The ABox approximation* $\mathcal{B}$ *of* $\exists X.\mathcal{A}$ *w.r.t.* $\mathcal{T}$ *is* optimal *if there is no ABox approximation* $\mathcal{C}$ *of* $\exists X.\mathcal{A}$ *w.r.t.* $\mathcal{T}$ *such that* $\mathcal{C} \models^{\mathcal{T}} \mathcal{B}$*, but* $\mathcal{B} \not\models^{\mathcal{T}} \mathcal{C}$*.*

Such an optimal ABox approximation need not exist. The qABox $\exists\{x\}.\mathcal{A}'$ with $\mathcal{A}' = \{r(a,x), r(x,x)\}$ is an example for this case. In fact, this qABox entails $((\exists r.)^n\top)(a)$ for all $n \geq 1$, which is not possible for an ABox entailed by $\exists\{x\}.\mathcal{A}'$ since such an ABox cannot contain role assertions and can contain only finitely many concept assertions. However, if an optimal ABox approximation exists, then it is unique up to equivalence. This is an easy consequence of the fact that the union of two ABox approximations is again an ABox approximation.

**Proposition 8.** *If* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *are optimal ABox approximations of the qABox* $\exists X.\mathcal{A}$ *w.r.t. the* $\mathcal{EL}$ *TBox* $\mathcal{T}$*, then* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *are equivalent w.r.t.* $\mathcal{T}$*.*

Optimal ABox approximations can now be characterized as follows.

**Theorem 9.** *The ABox* $\mathcal{B}$ *is an optimal ABox approximation of* $\exists X.\mathcal{A}$ *w.r.t.* $\mathcal{T}$ *iff* $\exists X.\mathcal{A}$ *and* $\mathcal{B}$ *are* $\mathsf{IRQ}$*-equivalent.*

*Proof.* First, assume that $\exists X.\mathcal{A}$ and $\mathcal{B}$ are $\mathsf{IRQ}$-equivalent w.r.t. $\mathcal{T}$. Then $\exists X.\mathcal{A} \models^{\mathcal{T}} \mathcal{B}$ by Proposition 3, and thus $\mathcal{B}$ is an ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$. If $\mathcal{C}$ is another ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$, then $\exists X.\mathcal{A} \models^{\mathcal{T}} \mathcal{C}$ by definition, and thus $\mathcal{B} \models^{\mathcal{T}} \mathcal{C}$ due to the assumed $\mathsf{IRQ}$-equivalence. This shows optimality of $\mathcal{B}$.

Second, assume that $\mathcal{B}$ is an optimal ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$ that is not $\mathsf{IRQ}$-equivalent with $\exists X.\mathcal{A}$. Then there is either a role assertion that belongs to $\mathcal{A}$, but not to $\mathcal{B}$, or a concept assertion that is entailed w.r.t. $\mathcal{T}$ by $\exists X.\mathcal{A}$, but not by $\mathcal{B}$. Adding this assertion to $\mathcal{B}$ yields an ABox $\mathcal{B}'$ that is an ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$. In addition, it satisfies $\mathcal{B}' \models^{\mathcal{T}} \mathcal{B}$, but not $\mathcal{B} \models^{\mathcal{T}} \mathcal{B}'$, which contradicts the assumed optimality of $\mathcal{B}$. □

An approach for deciding whether a given qABox has an optimal ABox approximation, and for computing it in case it exists, will be described in Section 4. But first, we show how optimal ABox approximations can be used to compute optimal ABox repairs.

### 3.3 Optimal ABox repairs

The repair approaches developed in [3] in general yield quantified ABoxes as output, even if the input is an ABox. We are now interested in producing repairs that are ABoxes. The approach developed below does not require the input to be an ABox. It actually assumes that the input is a qABox, which means that input ABoxes first need to be transformed into equivalent qABoxes.

**Definition 10.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, $\exists X.\mathcal{A}$ a qABox, and $\mathcal{R}$ an $\mathcal{EL}$ repair request. We call an $\mathcal{EL}$ ABox $\mathcal{B}$ an* ABox repair *of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ if $\exists X.\mathcal{A} \models^{\mathcal{T}} \mathcal{B}$ and $\mathcal{B} \not\models^{\mathcal{T}} C(a)$ for all $C(a) \in \mathcal{R}$. The ABox repair $\mathcal{B}$ of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ is* optimal *if there is no ABox repair $\mathcal{C}$ of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ such that $\mathcal{C} \models^{\mathcal{T}} \mathcal{B}$, but $\mathcal{B} \not\models^{\mathcal{T}} \mathcal{C}$.*

Our approach for computing optimal ABox repairs proceeds as follows: first, we compute the set of all optimal $\mathsf{IRQ}$-repairs of $\exists X.\mathcal{A}$, and then ABox-approximate the elements of this set. In the following, if we say that $\mathfrak{R}$ is the set of optimal $\mathsf{IRQ}$-repairs of a qABox, we mean that, for every optimal $\mathsf{IRQ}$-repair, $\mathfrak{R}$ contains one element of its $\mathsf{IRQ}$-equivalence class. Also, for a given qABox $\exists Y.\mathcal{B}$, we define

$$\mathsf{Oapp}^{\mathcal{T}}(\exists Y.\mathcal{B}) := \begin{cases} \{\mathcal{C}\} & \text{for an optimal ABox approx. } \mathcal{C} \text{ of } \exists Y.\mathcal{B} \text{ w.r.t. } \mathcal{T}, \\ \emptyset & \text{if no optimal ABox approx. of } \exists Y.\mathcal{B} \text{ w.r.t. } \mathcal{T} \text{ exists.} \end{cases}$$

**Theorem 11.** *Let $\exists X.\mathcal{A}$ be a qABox, $\mathcal{T}$ an $\mathcal{EL}$-TBox, $\mathcal{R}$ an $\mathcal{EL}$ repair request, and $\mathfrak{R}$ the set of optimal $\mathsf{IRQ}$-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. Then the set*

$$\bigcup_{\exists Y.\mathcal{B} \in \mathfrak{R}} \mathsf{Oapp}^{\mathcal{T}}(\exists Y.\mathcal{B})$$

*consists of all optimal ABox repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ up to equivalence.*

*Proof.* First, assume that the ABox $\mathcal{C}$ belongs to the union defined in the statement of the theorem. Then $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IRQ}} \exists Y.\mathcal{B} \models^{\mathcal{T}} \mathcal{C}$ for some qABox $\exists Y.\mathcal{B} \in \mathfrak{R}$ that has $\mathcal{C}$ as an optimal ABox approximation. This implies that $\mathcal{C}$ does not entail any of the concept assertions in $\mathcal{R}$ (since $\exists Y.\mathcal{B}$ does not) and that $\exists X.\mathcal{A} \models^{\mathcal{T}} \mathcal{C}$. Thus, $\mathcal{C}$ is an ABox repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. It remains to show that it is optimal. Assume to the contrary that $\mathcal{C}'$ is an ABox repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ such that $\mathcal{C}' \models^{\mathcal{T}} \mathcal{C}$, but $\mathcal{C} \not\models^{\mathcal{T}} \mathcal{C}'$. Since $\mathcal{C}$ and $\exists Y.\mathcal{B}$ are $\mathsf{IRQ}$-equivalent by Theorem 9, this is a contradiction to the fact that $\exists Y.\mathcal{B}$ is an optimal $\mathsf{IRQ}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ since $\mathcal{C}'$ would then be a better $\mathsf{IRQ}$-repair.

Second, assume that the ABox $\mathcal{C}$ is an optimal ABox repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. Then $\mathcal{C}$ is also an $\mathsf{IRQ}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$, and thus Proposition 5 yields that there is an optimal $\mathsf{IRQ}$-repair $\exists Y.\mathcal{B} \in \mathfrak{R}$ such that $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IRQ}} \exists Y.\mathcal{B} \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{C}$. We know by Proposition 3 that the second $\mathsf{IRQ}$-entailment is in fact an entailment, and thus $\mathcal{C}$ is an ABox approximation of $\exists Y.\mathcal{B}$. It remains to show that it is optimal. Assume to the contrary that $\mathcal{C}'$ is an ABox approximation of $\exists Y.\mathcal{B}$ such that $\exists Y.\mathcal{B} \models^{\mathcal{T}} \mathcal{C}' \models^{\mathcal{T}} \mathcal{C}$, but $\mathcal{C} \not\models^{\mathcal{T}} \mathcal{C}'$. But then $\mathcal{C}'$ is an ABox repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ (since $\exists Y.\mathcal{B}$ is a repair) that is better than $\mathcal{C}$, which contradicts our assumption that $\mathcal{C}$ is optimal. □

Once we have developed a method for computing the sets $\mathsf{Oapp}^{\mathcal{T}}(\exists Y.\mathcal{B})$, this theorem shows how to compute the set of all optimal ABox repairs of a given qABox. Such a method will be introduced in the next section. Before doing this, we want to point out that, in contrast to the set of optimal $\mathsf{IRQ}$-repairs, which covers all $\mathsf{IRQ}$-repairs, the set of optimal ABox repairs in general does not cover all ABox repairs.

*Example 12.* Consider the ABox $\mathcal{A} = \{A(a), r(a,b), B(b)\}$, the TBox $\mathcal{T} = \{B \sqsubseteq \exists r.B, \exists r.B \sqsubseteq B\}$ and the repair request $\mathcal{R} = \{(A \sqcap \exists r.B)(a)\}$. There are basically three options for IRQ-repairing $\mathcal{A}$: remove $A(a)$, remove $B(b)$, or remove $r(a,b)$. Since things implied by the TBox must also be taken into account, these three options yield the following optimal IRQ-repairs of $\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$:[5] $\mathcal{B}_1 = \{r(a,b), B(b)\}$ as well as $\exists\{x\}.\mathcal{B}_i$ for $i = 2, 3$, where $\mathcal{B}_2 = \{A(a), r(a,b), r(b,x), r(x,x)\}$ and $\mathcal{B}_3 = \{A(a), B(b), r(a,x), r(x,x)\}$. Of these three, $\mathcal{B}_1$ is already an ABox, and thus its own optimal ABox approximation, whereas the other two have no optimal ABox approximation. However, they have non-optimal ABox approximations, which are not necessarily covered by $\mathcal{B}_1$. For example, $\{A(a), r(a,b), (\exists r.\exists r.\top)(b)\}$ is an ABox approximation of $\exists\{x\}.\mathcal{B}_2$ and an ABox repair of $\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$, but since it contains $A(a)$, it is not entailed by $\mathcal{B}_1$.

## 4 Computing optimal ABox approximations

In this section, we assume that $\exists X.\mathcal{A}$ is a qABox and $\mathcal{T}$ an $\mathcal{EL}$ TBox. We will develop an approach for deciding whether $\exists X.\mathcal{A}$ has an optimal ABox approximation w.r.t. $\mathcal{T}$, which in the affirmative case also yields such an optimal approximation.

The first step is to saturate $\exists X.\mathcal{A}$ using the IQ-saturation rules of Fig. 1. In the following, let $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ denote a (fixed) qABox obtained by applying the IQ-saturation rules exhaustively to $\exists X.\mathcal{A}$. Note that the size of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ is polynomial in the size of the input $\exists X.\mathcal{A}$ and $\mathcal{T}$, and that $\exists X.\mathcal{A}$ and $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ are IQ-equivalent w.r.t. $\mathcal{T}$ by Theorem 1. In addition, it is easy to see that these two qABoxes contain the same individuals and the same role assertions between individuals. Thus, they are even IRQ-equivalent w.r.t. $\mathcal{T}$. As before, we use $\Sigma_\mathsf{I}$ to denote set of individuals of $\exists X.\mathcal{A}$.

In the next step, we transform $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ into a new qABox, called pre-approximation, whose matrix basically consists of the union of ABoxes $\mathcal{B}_a$ for each $a \in \Sigma_\mathsf{I}$, extended with the role assertions between individuals in $\mathcal{A}$. Each ABox $\mathcal{B}_a$ contains $a$ as the only individual name, and further contains a fully anonymized copy of the saturation $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$, which is connected with $a$ by indispensable role assertions.

**Definition 13.** *We call a role assertion $r(a,u)$ in $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ for $a \in \Sigma_\mathsf{I}$* indispensable *if there is no role assertion $r(a,b)$ for $b \in \Sigma_\mathsf{I}$ such that there is a simulation from $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ to itself that contains $(u,b)$.*

Since an individual always simulates itself, only role assertion $r(a,u)$ where $u$ is a variable can be indispensable. We are now ready to define the pre-approximation.

[5] The IQ-repairs computed by the approaches in [3] would contain more assertions, which are however redundant for IRQ-entailment w.r.t. $\mathcal{T}$.

**Definition 14.** *The* pre-approximation $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ *of* $\exists X.\mathcal{A}$ *w.r.t.* $\mathcal{T}$ *is defined as the quantified ABox* $\exists Y.\mathcal{B}$*, where*

$$
\begin{aligned}
Y &\coloneqq \{\, u' \mid u \text{ is an object name occurring in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \,\},\\
\mathcal{B} &\coloneqq \bigcup \{\, \mathcal{B}_a \mid a \text{ is an individual name in } \Sigma_{\mathsf{I}} \,\}\\
&\qquad \cup \{\, r(a,b) \mid r(a,b) \text{ occurs in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \text{ where } a,b \in \Sigma_{\mathsf{I}} \,\},\\
\mathcal{B}_a &\coloneqq \{\, A(a) \mid A(a) \text{ occurs in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \,\}\\
&\qquad \cup \{\, r(a,u') \mid r(a,u) \text{ occurs in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \text{ and is indispensable} \,\}\\
&\qquad \cup \{\, A(u') \mid A(u) \text{ occurs in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \,\}\\
&\qquad \cup \{\, r(u',v') \mid r(u,v) \text{ occurs in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \,\}.
\end{aligned}
$$

Obviously, the pre-approximation can be computed in polynomial time. In addition, it is IRQ-equivalent to $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$.

**Lemma 15.** *The qABoxes* $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ *and* $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ *are* IRQ-*equivalent w.r.t. the empty TBox* $\emptyset$*, and thus also w.r.t.* $\mathcal{T}$*.*

Since we already know that $\exists X.\mathcal{A}$ and $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ are IRQ-equivalent w.r.t. $\mathcal{T}$, this shows that $\exists X.\mathcal{A}$ is IRQ-equivalent to its pre-approximation w.r.t. $\mathcal{T}$. Consequently, an ABox $\mathcal{C}$ is an optimal ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$ iff it is one of the pre-approximation w.r.t. $\mathcal{T}$.

To test whether $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ has an optimal ABox approximation w.r.t. $\mathcal{T}$, it is sufficient to check whether, for all $a \in \Sigma_{\mathsf{I}}$, the individual $a$ has a most specific concept in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$.

**Definition 16.** *Let* $\mathcal{C}$ *be an* $\mathcal{EL}$ *ABox,* $\mathcal{T}$ *an* $\mathcal{EL}$ *TBox, and* $a$ *an individual name. The* $\mathcal{EL}$ *concept* $C$ *is a* most specific concept (msc) *of* $a$ *in* $\mathcal{C}$ *w.r.t.* $\mathcal{T}$ *if* $\mathcal{C} \models^{\mathcal{T}} C(a)$ *and* $\mathcal{C} \models^{\mathcal{T}} D(a)$ *implies* $C \sqsubseteq^{\mathcal{T}} D$ *for all* $\mathcal{EL}$ *concepts* $D$*.*

The most specific concept need not exist, but if it does, then it is unique up to equivalence w.r.t. $\mathcal{T}$. The ABox $\mathcal{C} := \{r(a,a)\}$ is a simple example where the msc of $a$ does not exist w.r.t. the empty TBox. In fact, $\mathcal{C} \models (\exists r.)^n \top$ for all $n \geq 1$, and it is easy to see that no $\mathcal{EL}$ concept can be subsumed by these infinitely many concepts. Note, however, that $\mathcal{C}$ has an optimal ABox approximation since it is itself an ABox. In this case, the pre-approximation is $\{r(a,a)\} \cup \mathcal{B}_a$ where $\mathcal{B}_a = \{r(a',a')\}$. There is no role assertion $r(a,a')$ since $r(a,a)$ is not indispensable. While $a'$ does not have an msc in $\mathcal{B}_a$, this is not what we are interested in. We want to know whether $a$ has one, and the answer is “yes” since $\top$ is an msc of $a$ in $\mathcal{B}_a$. The problem of testing for the existence of and computing the msc in $\mathcal{EL}$ was investigated in [20], where the following result is stated.

**Proposition 17 ([20]).** *Let* $\mathcal{C}$ *be an* $\mathcal{EL}$ *ABox,* $\mathcal{T}$ *an* $\mathcal{EL}$ *TBox, and* $a$ *an individual name. It can be decided in polynomial time whether* $a$ *has a most specific concept in* $\mathcal{C}$ *w.r.t.* $\mathcal{T}$*, and if the msc exists, then it can be computed in exponential time.*

The main idea underlying the proof of this proposition (rephrased into the setting of the present paper) is to *unravel* the IQ-saturation of $\mathcal{C}$ w.r.t. $\mathcal{T}$ into a concept $C_k$ an increasing number $k$ of steps, starting from $a$. After each step, one tests whether the ABox $\{C_k(a)\}$ IQ-entails $\exists X.\mathcal{C}$ w.r.t. $\mathcal{T}$, where $X$ consists of the object names in $\mathcal{C}$ different from $a$. In case this test succeeds, the concept $C_k$ is the msc of $a$ in $\mathcal{C}$ w.r.t. $\mathcal{T}$. This yields an effective test for the existence of the msc since the following can be shown: there is a polynomial $p$ such that the entailment test succeeds after at most $p(|\mathcal{C}|, |\mathcal{T}|)$ steps iff the msc exists.

For example, for the ABox $\mathcal{C}^{(1)} = \{r(a,a)\}$ and the TBox $\mathcal{T}^{(1)} = \emptyset$, the 0-step unraveling is $C_0^{(1)} = \top$, the 1-step unraveling is $C_1^{(1)} = \exists r.\top$, the two-step unraveling is $C_2^{(1)} = \exists r.\exists r.\top$, etc. It is easy to that there is no $k$ such that the entailment test succeeds. Thus, it does not succeed for $k(\mathcal{C}^{(1)}, \mathcal{T}^{(1)})$, which shows that $a$ does not have an msc. If instead we consider the ABox $\mathcal{C}^{(2)} = \{A(a), r(a,b), s(a,b), r(b,c), s(b,c), B(c)\}$ w.r.t. $\mathcal{T}^{(2)} = \emptyset$, then the 0-step unraveling is $C_0^{(2)} = A$, the 1-step unraveling is $C_1^{(2)} = A \sqcap \exists r.\top \sqcap \exists s.\top$, the 2-step unraveling is $C_2^{(2)} = A \sqcap \exists r.(\exists r.B \sqcap \exists s.B) \sqcap \exists s.(\exists r.B \sqcap \exists s.B)$, and the 3-step unraveling is identical to $C_2^{(2)}$. The entailment test succeeds for $k = 2$. It is easy to see that, whenever the unraveling becomes stable (which happens if no cycle in the ABox is reachable from $a$), then the entailment test succeeds. However, a reachable cycle in the ABox need not prevent the existence of the msc. For example, the individual $a$ has the msc $\exists r.B$ in $\mathcal{C}^{(3)} = \{r(a,b), r(b,b), B(b)\}$ w.r.t. $\mathcal{T}^{(3)} = \{B \sqsubseteq \exists r.B\}$.

As sketched until now, this method for deciding the existence of the msc does not yield a polynomial-time decision procedure. The reason is that, though the bound $k(\mathcal{C}, \mathcal{T})$ on the number of steps is polynomial, the unraveled concepts $C_k$ may become exponential even for $k \leq k(\mathcal{C}, \mathcal{T})$, as can be seen using an obvious generalization of our example ABox $\mathcal{C}^{(2)}$. This problem can be avoided by employing structure-sharing, which can be realized by representing the ABoxes $\{C_k(a)\}$ by IQ-equivalent qABoxes. In our second example, the ABox $\{C_2^{(2)}(a)\}$ can be represented by the more compact IQ-equivalent qABox $\exists\{x, y\}.\{A(a), r(a,x), s(a,x), r(x,y), s(x,y), B(y)\}$ (see the definition of the $k$-unraveling in [1] for how such an unraveling with structure sharing can be defined in general). It is easy to see that the qABoxes representing the ABoxes $\{C_k(a)\}$ are of polynomial size. Since IQ-entailment between qABoxes is polynomial, this yields the polynomiality result stated in the proposition. Note, however, that the msc obtained this way is still an unraveled concept $C_k$ without structure sharing, and thus may be of exponential size.

The following theorem shows that existence of the optimal ABox approximation can be reduced to existence of the msc (see the supplementary material for the proof).

**Theorem 18.** *Let $\exists X.\mathcal{A}$ be a qABox with set of individuals $\Sigma_{\mathsf{I}}$, let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, and let $\mathcal{B}_a$ for all $a \in \Sigma_{\mathsf{I}}$ be the ABoxes introduced in Definition 14. Then $\exists X.\mathcal{A}$ has an optimal ABox approximation w.r.t. $\mathcal{T}$ iff, for all individuals $a \in \Sigma_{\mathsf{I}}$, the msc of $a$ in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$ exists. If the latter condition is satisfied and*

*$C_a$ are these most specific concepts, then the following ABox is an optimal ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$:*

$$\{\, C_a(a) \mid a \in \Sigma_\mathsf{I} \,\} \cup \{\, r(a,b) \mid r(a,b) \text{ occurs in } \underline{\mathsf{sat}}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X.\mathcal{A}) \text{ where } a,b \in \Sigma_\mathsf{I} \,\}.$$

*In particular, the existence of an optimal ABox approximation can be tested in polynomial time and such an optimal approximation can be computed in exponential time if it exists.*

## 5 Computing optimal ABox repairs

We can now reap the benefits from the results shown in the previous two sections. Given a qABox $\exists X.\mathcal{A}$, an $\mathcal{EL}$ TBox $\mathcal{T}$, and an $\mathcal{EL}$ repair request $\mathcal{R}$, we can compute the set $\mathfrak{R}$ of optimal $\mathsf{IRQ}$-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ in exponential time. More precisely, by Proposition 5 this set contains at most exponentially many repairs, each of which has at most exponential size. Theorem 11 then says that the set of all optimal ABox repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ (up to equivalence) consists of the optimal ABox approximations w.r.t. $\mathcal{T}$ of those elements of $\mathfrak{R}$ for which such an optimal approximation exists. Finally, Theorem 18 shows how to decide existence of such optimal approximations and how to compute them if they exist. Since the elements of $\mathfrak{R}$ are already of exponential size, existence can be tested in exponential time and the size of the computed approximations is at most double-exponential.

**Theorem 19.** *Let $\exists X.\mathcal{A}$ be a qABox, $\mathcal{T}$ an $\mathcal{EL}$-TBox, and $\mathcal{R}$ an $\mathcal{EL}$ repair request. Then the existence of an optimal ABox repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ can be decided in exponential time, and the set of all such repairs can be computed in double-exponential time. This set contains at most exponentially many elements, each of which has at most double-exponential size.*

If the given qABox does not have an optimal repair or if we are looking for a repair not covered by an optimal one, our approach can also be used to compute non-optimal ABox repairs. In fact, consider an optimal $\mathsf{IRQ}$-repair that does not have an optimal ABox approximation. Then there are individuals $a$ whose msc in $\mathcal{B}_a$ does not exist. Following [17], we can then use the role-depth bounded msc instead, which is basically obtained by unraveling up to a fixed bound $k$ on the role-depth (i.e., the maximal nesting of existential restrictions). This way, we can produce a set of (possibly) non-optimal ABox repairs, which covers all ABox repairs whose concept assertions satisfy this bound on the role depth.

There are also cases where the existence of the optimal ABox approximation of the optimal $\mathsf{IRQ}$-repairs is guaranteed. In fact, if the qABox is acyclic and the TBox is cycle-restricted (i.e., there is no concept $C$ such that $C \sqsubseteq^{\mathcal{T}} \exists r_1.\cdots\exists r_k.C$, as defined in [3]), then the optimal $\mathsf{IRQ}$-repairs are acyclic, which implies that the ABoxes $\mathcal{B}_a$ in the pre-approximations are also acyclic. Consequently, all optimal $\mathsf{IRQ}$-repairs have an optimal ABox approximation. The following corollary is an easy consequence of this observation.

**Corollary 20.** *Let $\exists X.\mathcal{A}$ be an acyclic qABox, $\mathcal{T}$ a cycle-restricted $\mathcal{EL}$-TBox, and $\mathcal{R}$ an $\mathcal{EL}$ repair request. Then the set of optimal ABox repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ is non-empty, and it* IRQ*-covers all ABox repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$.*

## 6 Conclusion

Traditional repair approaches for DL-based ontologies, which compute maximal subsets of the ontology that do not have the unwanted consequences, are syntax-dependent and thus may remove too many consequences. Recently developed syntax-independent approaches for repairing DL ABoxes [3, 5, 7] compute optimal repairs that do not lose consequences unnecessarily, but they have the disadvantage that they produce quantified ABoxes rather than traditional ABoxes. In this paper we show how to overcome this problem by developing methods for computing optimal repairs that are traditional ABoxes. These methods are based on the computation of optimal IRQ-repairs, by adapting the approaches in [3] for computing optimal IQ-repairs, and then optimally approximating these qABoxes with ABoxes.

A perceived disadvantage of our approach could be that optimal ABox repairs need not exist, and even if they do, they need not cover all ABox repairs. However, by Corollary 20 this problem does not occur if the ABox is acyclic and the TBox is cycle-restricted. To see how often this corollary applies in practice, we checked the 80 large ontologies used in the experiments in [3]: 62 have cycle-restricted TBoxes, and of those only 7 have cyclic ABoxes. Thus, our Corollary 20 applies to 55 of the 80 ontologies considered in [3].

Another disadvantage could be the potentially double-exponential size of optimal ABox repairs. However, the first exponential comes from the computation of the optimal IQ-repairs, and the experiments in [3] indicate that this exponential blow-up does not occur in practice if the optimized approach for computing IQ-repairs is used. We do not yet have experimental results regarding the possible exponential blow-up due to the computation of ABox approximations, but would be surprised if this happened often in practice.

What is called “repair” in the DL community is closely related to what is called “contraction” in the Belief Change community. For classical repairs and also for the gentle repairs of [6], this connection was investigated in [14]. It would be interesting to see whether this investigation can be extended to our optimal ABox repairs. The original intention underlying our repair approach is that the ontology engineer chooses one of the computed optimal repairs as the new, repaired ABox. Alternatively, one could try to adapt the different repair semantics employed in inconsistency-tolerant query answering [9, 12] from classical repairs to our optimal repairs.

**Acknowledgements** This work was partially supported by the AI competence center ScaDS.AI Dresden/Leipzig and the Deutsche Forschungsgemeinschaft (DFG), Grant 430150274, and Grant 389792660 within TRR 248.

## References


1. Baader, F.: A graph-theoretic generalization of the least common subsumer and the most specific concept in the description logic $\mathcal{EL}$. In: Graph-Theoretic Concepts in Computer Science, 30th International Workshop, WG 2004, Bad Honnef, Germany, June 21-23, 2004, Revised Papers. Lecture Notes in Computer Science, vol. 3353, pp. 177–188. Springer (2004), https://doi.org/10.1007/978-3-540-30559-0_15
2. Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017), https://doi.org/10.1017/9781139025355
3. Baader, F., Koopmann, P., Kriegel, F., Nuradiansyah, A.: Computing optimal repairs of quantified ABoxes w.r.t. static $\mathcal{EL}$ TBoxes. In: Automated Deduction - CADE 28 - 28th International Conference on Automated Deduction, Virtual Event, July 12-15, 2021, Proceedings. Lecture Notes in Computer Science, vol. 12699, pp. 309–326. Springer (2021), https://doi.org/10.1007/978-3-030-79876-5_18
4. Baader, F., Koopmann, P., Kriegel, F., Nuradiansyah, A.: Computing optimal repairs of quantified ABoxes w.r.t. static $\mathcal{EL}$ TBoxes (extended version). LTCS-Report 21-01, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2021), https://doi.org/10.25368/2022.64
5. Baader, F., Kriegel, F., Nuradiansyah, A.: Privacy-preserving ontology publishing for $\mathcal{EL}$ instance stores. In: Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019, Proceedings. Lecture Notes in Computer Science, vol. 11468, pp. 323–338. Springer (2019), https://doi.org/10.1007/978-3-030-19570-0_21
6. Baader, F., Kriegel, F., Nuradiansyah, A., Peñaloza, R.: Making repairs in description logics more gentle. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018. pp. 319–328. AAAI Press (2018), https://aaai.org/ocs/index.php/KR/KR18/paper/view/18056
7. Baader, F., Kriegel, F., Nuradiansyah, A., Peñaloza, R.: Computing compliant anonymisations of quantified ABoxes w.r.t. $\mathcal{EL}$ policies. In: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part I. Lecture Notes in Computer Science, vol. 12506, pp. 3–20. Springer (2020), https://doi.org/10.1007/978-3-030-62419-4_1
8. Baader, F., Suntisrivaraporn, B.: Debugging SNOMED CT using axiom pinpointing in the description logic $\mathcal{EL}^+$. In: Proceedings of the Third International Conference on Knowledge Representation in Medicine, Phoenix, Arizona, USA, May 31st - June 2nd, 2008. CEUR Workshop Proceedings, vol. 410. CEUR-WS.org (2008), http://ceur-ws.org/Vol-410/Paper01.pdf
9. Bienvenu, M., Bourgaux, C.: Inconsistency-tolerant querying of description logic knowledge bases. In: Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering - 12th International Summer School 2016, Aberdeen, UK, September 5-9, 2016, Tutorial Lectures. Lecture Notes in Computer Science, vol. 9885, pp. 156–202. Springer (2016), https://doi.org/10.1007/978-3-319-49493-7_5
10. Henzinger, M.R., Henzinger, T.A., Kopke, P.W.: Computing simulations on finite and infinite graphs. In: 36th Annual Symposium on Foundations of Computer Science, Milwaukee, Wisconsin, USA, 23-25 October 1995. pp. 453–462. IEEE Computer Society (1995), https://doi.org/10.1109/SFCS.1995.492576

11. Kalyanpur, A., Parsia, B., Horridge, M., Sirin, E.: Finding all justifications of OWL DL entailments. In: The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Lecture Notes in Computer Science, vol. 4825, pp. 267–280. Springer (2007), https://doi.org/10.1007/978-3-540-76298-0_20
12. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant query answering in ontology-based data access. J. Web Semant. **33**, 3–29 (2015), https://doi.org/10.1016/j.websem.2015.04.002
13. Lutz, C., Wolter, F.: Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. J. Symb. Comput. **45**(2), 194–228 (2010), https://doi.org/10.1016/j.jsc.2008.10.007
14. Matos, V.B., Guimarães, R., Santos, Y.D., Wassermann, R.: Pseudo-contractions as gentle repairs. In: Description Logic, Theory Combination, and All That - Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday. Lecture Notes in Computer Science, vol. 11560, pp. 385–403. Springer (2019), https://doi.org/10.1007/978-3-030-22102-7_18
15. Meyer, T.A., Lee, K., Booth, R., Pan, J.Z.: Finding maximally satisfiable terminologies for the description logic $\mathcal{ALC}$. In: Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA. pp. 269–274. AAAI Press (2006), http://www.aaai.org/Library/AAAI/2006/aaai06-043.php
16. Parsia, B., Sirin, E., Kalyanpur, A.: Debugging OWL ontologies. In: Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005. pp. 633–640. ACM (2005), https://doi.org/10.1145/1060745.1060837
17. Peñaloza, R., Turhan, A.: A practical approach for computing generalization inferences in $\mathcal{EL}$. In: The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I. Lecture Notes in Computer Science, vol. 6643, pp. 410–423. Springer (2011), https://doi.org/10.1007/978-3-642-21034-1_28
18. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003. pp. 355–362. Morgan Kaufmann (2003), http://ijcai.org/Proceedings/03/Papers/053.pdf
19. Schlobach, S., Huang, Z., Cornet, R., van Harmelen, F.: Debugging incoherent terminologies. J. Autom. Reason. **39**(3), 317–349 (2007), https://doi.org/10.1007/s10817-007-9076-z
20. Zarrieß, B., Turhan, A.: Most specific generalizations w.r.t. general $\mathcal{EL}$-TBoxes. In: IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013. pp. 1191–1197. IJCAI/AAAI (2013), http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6709

# Supplementary Material

Before we can start to provide the proofs missing from the main text, we must introduce some notation. An $\mathcal{EL}$ atom (or just atom) is an $\mathcal{EL}$ concept of the form $A$ for $A \in \Sigma_\mathsf{C}$ or $\exists r.C$ for $r \in \Sigma_\mathsf{R}$ and an $\mathcal{EL}$ concept $C$. Given an $\mathcal{EL}$ concept $C$, we denote the set of atoms occurring in the top-level conjunction of $C$ with $\mathsf{Conj}(C)$ and all atoms occurring somewhere in $C$ with $\mathsf{Atoms}(C)$. Given an $\mathcal{EL}$ TBox $\mathcal{T}$ and an $\mathcal{EL}$ repair request $\mathcal{R}$, $\mathsf{Atoms}(\mathcal{R},\mathcal{T})$ denotes the set of atoms occurring somewhere in $\mathcal{T}$ or $\mathcal{R}$.

## A Proof of Proposition 5

In this section, we show that the canonical $\mathsf{IQ}$-repairs of [3] are also $\mathsf{IRQ}$-repairs and cover all $\mathsf{IRQ}$-repairs. As in Definition 7 in [3], every canonical $\mathsf{IQ}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ is induced by a seed function $s$ and is denoted by $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$.

**Proposition 21.** *For each seed function $s$, the induced canonical $\mathsf{IQ}$-repair $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$ is an $\mathsf{IRQ}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. Moreover, for each $\mathsf{IRQ}$-repair $\exists Y.\mathcal{B}$ of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$, there exists a repair seed function $s$ such that $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s) \models^\mathcal{T}_\mathsf{IRQ} \exists Y.\mathcal{B}$.*

*Proof.* Consider a repair seed function $s$. According to Proposition 8 in [3], it holds that $\exists X.\mathcal{A} \models^\mathcal{T}_\mathsf{IQ} \mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$ and $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s) \not\models^\mathcal{T} C(a)$ for each $C(a) \in \mathcal{R}$. Additionally, Definition 7 in [3] immediately implies that each role assertion in $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$ involving only individual names is also contained in $\exists X.\mathcal{A}$, and thus Proposition 2 yields that $\exists X.\mathcal{A} \models^\mathcal{T}_\mathsf{IRQ} \mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$. It follows that $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$ is also an $\mathsf{IRQ}$-repair.

Now let $\exists Y.\mathcal{B}$ be an $\mathsf{IRQ}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. Since $\models^\mathcal{T}_\mathsf{IQ}$ is weaker than $\models^\mathcal{T}_\mathsf{IRQ}$, it follows that $\exists Y.\mathcal{B}$ is also an $\mathsf{IQ}$-repair of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. By Proposition 8 in [3], there exists a repair seed function $s$ such that $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s) \models^\mathcal{T}_\mathsf{IQ} \exists Y.\mathcal{B}$. We are going to show that the latter is also an $\mathsf{IRQ}$-entailment, which boils down to checking that each role assertion in $\exists Y.\mathcal{B}$ involving only individual names is also contained in $\mathsf{rep}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A}, s)$, cf. Proposition 2. Beforehand, note that the repair seed function $s$ is defined in the proof of Proposition 8 in [4] as follows for each individual name $a$:[6]

$$s(a) \coloneqq \mathsf{Max}_{\sqsubseteq^\emptyset}\left(\{\, G \mid G \in \mathsf{Atoms}(\mathcal{R},\mathcal{T}),\ \exists Y.\mathcal{B} \not\models^\mathcal{T} G(a),\ \text{and } \exists X.\mathcal{A} \models^\mathcal{T} G(a) \,\}\right).$$

Consider some role assertion $r(a, b)$ in $\exists Y.\mathcal{B}$ where $a$ and $b$ are individual names. Since $\exists Y.\mathcal{B}$ is an $\mathsf{IRQ}$-repair of $\exists X.\mathcal{A}$, it holds that $\exists X.\mathcal{A} \models^\mathcal{T}_\mathsf{IRQ} \exists Y.\mathcal{B}$.

[6] We will implicitely use that, according to Theorem 3 in [3], $\exists X.\mathcal{A}$ entails $C(a)$ w.r.t. $\mathcal{T}$ iff (the matrix of) $\mathsf{sat}^\mathcal{T}_\mathsf{IQ}(\exists X.\mathcal{A})$ entails $C(a)$ (see Theorem 1).

Proposition 2 yields that $\exists X.\mathcal{A}$ contains $r(a,b)$ as well. Recall that $y_{a,s(a)}$ is a synonym for $a$ in $\mathsf{rep}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X.\mathcal{A}, s)$ and likewise for $y_{b,s(b)}$ and $b$. Thus we will show that $\mathsf{rep}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X.\mathcal{A}, s)$ contains the role assertion $r(y_{a,s(a)}, y_{b,s(b)})$ — according to Definition 7 in [3] we must show that, for each $\exists r.E \in s(a)$ where $\exists X.\mathcal{A} \models^{\mathcal{T}} E(b)$, the repair type $s(b)$ contains some atom subsuming $E$.

Let $\exists r.E \in s(a)$ where $\exists X.\mathcal{A} \models^{\mathcal{T}} E(b)$. We obtain that $\exists Y.\mathcal{B} \not\models^{\mathcal{T}} \exists r.E(a)$. Since $\exists Y.\mathcal{B}$ contains $r(a,b)$, we infer that $\exists Y.\mathcal{B} \not\models^{\mathcal{T}} E(b)$. It follows that there is an atom $F \in \mathsf{Conj}(E)$ where $\exists Y.\mathcal{B} \not\models^{\mathcal{T}} F(b)$, and where $\exists X.\mathcal{A} \models^{\mathcal{T}} F(b)$. We conclude that $s(b)$ contains either $F$ or some atom that subsumes $F$, i.e., $s(b)$ contains an atom subsuming $E$. □

Next, we turn our attention to the optimized IQ-repairs as per Definition 12 in [3]. As it turns out, each of them is IRQ-equivalent to the canonical IQ-repair induced by the same repair seed function.

**Proposition 22.** *For each repair seed function $s$, the optimized* IQ*-repair induced by $s$ is* IRQ*-equivalent to the canonical* IQ*-repair induced by $s$.*

*Proof.* Since the optimized IQ-repair is a sub-qABox of the canonical IQ-repair induced by the same repair seed function, it immediately follows that the former is IRQ-entailed by the latter.

Let $s$ be a repair seed function. Proposition 13 in [3] shows that the optimized IQ-repair induced by $s$ IQ-entails the canonical IQ-repair induced by $s$. Now consider a role assertion $r(a,b)$ in the canonical repair induced by $s$ where $a$ and $b$ are individual names. Then $a$ and $b$ are contained in the set $\Sigma_{\mathsf{I}} \cup Y_m$ from which the matrix of the optimized IQ-repair is constructed (see Definition 12 in [3]), and so this role assertion $r(a,b)$ occurs in the optimized IQ-repair induced by $s$ as well. An application of Proposition 2 yields the IRQ-entailment. □

We are now ready to prove the main result on the relationship between IQ- and IRQ-repairs.

**Proposition 5.** *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, $\exists X.\mathcal{A}$ a qABox, and $\mathcal{R}$ an $\mathcal{EL}$ repair request. If $\mathfrak{R}$ is the set of all canonical or all optimized* IQ*-repairs obtained from this input according to the definitions in [3], then $\mathfrak{R}$ is a set of* IRQ*-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$ that* IRQ*-covers all* IRQ*-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$. In particular, up to* IRQ*-equivalence, the set of optimal* IRQ*-repairs can be computed in exponential time, and it* IRQ*-covers all* IRQ*-repairs of $\exists X.\mathcal{A}$ for $\mathcal{R}$ w.r.t. $\mathcal{T}$.*

*Proof.* The first claim follows from Proposition 21 for the case where $\mathfrak{R}$ is the set of all canonical IQ-repairs, and from Proposition 22 for the case where $\mathfrak{R}$ is the set of all optimized IQ-repairs. Filtering out the non-optimal repairs from $\mathfrak{R}$ needs exponential time, since $\mathfrak{R}$ contains at most exponentially many repairs of at most exponential size each (see [3]), and since the IRQ-entailment relation $\models_{\mathsf{IRQ}}^{\mathcal{T}}$ can be decided in polynomial time (see Section 2). □

We have seen above that the set of all optimal IRQ-repairs can be constructed by, firstly, computing the set of *all* canonical repairs and, secondly, filtering out the repairs that are not minimal w.r.t. $\mathsf{IRQ}$-entailment. As there exist exponentially many repairs, such an approach is rather expensive since exponentially many comparisons with the $\mathsf{IRQ}$-entailment relation $\models^{\mathcal{T}}_{\mathsf{IRQ}}$ are necessary, even though $\models^{\mathcal{T}}_{\mathsf{IRQ}}$ is decidable in polynomial time. As an alternative, we can filter earlier, namely on the set of repair seed functions.

For the case of an empty TBox, it was shown in Section 4 in [7] that there is a partial order $\leq$ on the seed functions such that $\leq$-minimal seed functions correspond to the optimal $\mathsf{IQ}$-repairs. Using Lemma XII in [3], it can be shown that the same partial order $\leq$ characterizes the optimal $\mathsf{IQ}$-repairs also when a TBox is present. Specifically, it holds that $s \leq t$ if and only if $\mathsf{rep}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}, s) \models^{\mathcal{T}}_{\mathsf{IQ}} \mathsf{rep}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}, t)$.

By slightly extending the definition of $\leq$ such that it also checks that each role assertion $r(a, b)$ in $\mathsf{rep}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}, t)$ where $a, b \in \Sigma_{\mathsf{I}}$ also occurs in $\mathsf{rep}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}, s)$, we obtain a polynomial-time-decidable partial order $\leq'$ on the seed functions such that $s \leq' t$ if and only if $\mathsf{rep}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}, s) \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathsf{rep}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}, t)$. It follows that a set of optimal $\mathsf{IRQ}$-repairs can be obtained by, firstly, computing the set of all $\leq'$-minimal seed functions and, secondly, computing the canonical or optimized $\mathsf{IQ}$-repairs induced by them.

## B Proof of Proposition 8

**Proposition 8.** *If $\mathcal{B}_1$ and $\mathcal{B}_2$ are optimal ABox approximations of the qABox $\exists X. \mathcal{A}$ w.r.t. the $\mathcal{EL}$ TBox $\mathcal{T}$, then $\mathcal{B}_1$ and $\mathcal{B}_2$ are equivalent w.r.t. $\mathcal{T}$.*

*Proof.* Assume that $\mathcal{B}_1$ and $\mathcal{B}_2$ are optimal ABox approximations of $\exists X. \mathcal{A}$ w.r.t. $\mathcal{T}$ such that $\mathcal{B}_1 \not\models^{\mathcal{T}} \mathcal{B}_2$. Then it is easy to see that $\mathcal{B}_1 \cup \mathcal{B}_2$ is an ABox approximation of $\exists X. \mathcal{A}$ w.r.t. $\mathcal{T}$ such that $\mathcal{B}_1 \cup \mathcal{B}_2 \models^{\mathcal{T}} \mathcal{B}_1$, but $\mathcal{B}_1 \not\models^{\mathcal{T}} \mathcal{B}_1 \cup \mathcal{B}_2$. This contradicts the assumed optimality of $\mathcal{B}_1$. □

## C Proof of Lemma 15

**Lemma 15.** *The qABoxes $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A})$ and $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X. \mathcal{A})$ are $\mathsf{IRQ}$-equivalent w.r.t. the empty TBox $\emptyset$, and thus also w.r.t. $\mathcal{T}$.*

*Proof.* By the very construction, both qABoxes contain the same role assertions involving only individual names. It remains to show that both are $\mathsf{IQ}$-equivalent. To do so, we need to find simulations between them. Firstly, consider the following relation.

$$\mathfrak{S} := \left\{ (t, a) \;\middle|\; \begin{array}{l} t \text{ is an object name in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}) \\ \text{and } a \text{ is an individual name such that } t \precsim a \end{array} \right\}$$
$$\cup \{ (t, u') \mid t \text{ and } u \text{ are object names in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A}) \text{ such that } t \precsim u \}$$

We show that $\mathfrak{S}$ is a simulation from $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A})$ to $\exists Y. \mathcal{B}$.

1. Obviously, $\mathfrak{S}$ contains the pair $(a, a)$ for each individual name $a$.
2. We now deal with the concept assertions.
   (a) Let $(t, a) \in \mathfrak{S}$ where $t$ is an object name and $a$ is an individual name such that $t \precsim a$. If the saturation contains $A(t)$, then it contains $A(a)$ as well. Thus the pre-approximation also contains $A(a)$.
   (b) If $(t, u') \in \mathfrak{S}$ where $t$ and $u$ are object names in the saturation such that $t \precsim u$, and $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A})$ contains $A(t)$, then the saturation contains $A(u)$ as well and so, by the very definition, the pre-approximation contains $A(u')$.
3. It remains to consider the role assertions.
   (a) Let $(t, a) \in \mathfrak{S}$ where $t$ is an object name in the saturation and $a$ is an individual name such that $t \precsim a$. Further assume that $r(t, u)$ is a role assertion in the saturation. Since $t$ is simulated by $a$, there is an object name $v$ such that $u \precsim v$ and the saturation contains $r(a, v)$.
      – If $v$ is an individual name, then $\mathfrak{S}$ contains $(u, v)$ and the pre-approximation contains $r(a, v)$.
      – If $v$ is a variable and $r(a, v)$ is indispensable, then $\mathfrak{S}$ contains $(u, v')$ and the pre-approximation contains $r(a, v')$.
      – Otherwise, $v$ is a variable and $r(a, v)$ is dispensable. So the saturation contains a role assertion $r(a, b)$ where $v \precsim b$. The former yields that $r(a, b)$ is contained in the pre-approximation, and the latter implies that $u \precsim b$, i.e., $\mathfrak{S}$ contains $(u, b)$.
   (b) Let $(t, u') \in \mathfrak{S}$ where $t$ and $u$ are object names in the saturation such that $t \precsim u$, and further assume that $r(t, v)$ is a role assertion in the saturation. From $t \precsim u$ we infer that there is an object name $w$ such that $v \precsim w$ and $r(u, w)$ is a role assertion in the saturation. We further conclude that the relation $\mathfrak{S}$ contains $(v, w')$, and that the pre-approximation contains the role assertion $r(u', w')$.

Secondly, it is a finger exercise to verify that the following relation is a simulation in the converse direction, i.e., from the pre-approximation to the saturation.

$$\begin{aligned}\mathfrak{T} \coloneqq & \{\, (a, a) \mid a \text{ is an individual name} \,\} \\ & \cup \{\, (t', t) \mid t \text{ is an object name in the saturation} \,\}\end{aligned}$$ □

## D Proof of Theorem 18

Before we can prove the theorem, we must show two (rather technical) lemmas.

**Lemma 23.** *Assume that* $\exists X. \mathcal{A} =: \exists X_0. \mathcal{A}_0 \to \exists X_1. \mathcal{A}_1 \to \cdots \to \exists X_n. \mathcal{A}_n \coloneqq \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X. \mathcal{A})$ *is a sequence of qABoxes obtained by exhaustively applying the saturation rules to* $\exists X. \mathcal{A}$*, where* $\exists X_i. \mathcal{A}_i \to \exists X_{i+1}. \mathcal{A}_{i+1}$ *denotes that one saturation rule was applied once to* $\exists X_i. \mathcal{A}_i$ *in order to produce* $\exists X_{i+1}. \mathcal{A}_{i+1}$*. Further let* $x_C$ *be a variable in* $\exists X_i. \mathcal{A}_i$ *for some index* $i \in \{0, \ldots, n\}$*. Then the following two statements hold:*

1. *If the matrix of* $\exists X_i.\mathcal{A}_i$ *contains the concept assertion* $A(x_C)$ *for* $A \in \Sigma_\mathsf{C}$*, then*
   (a) *either* $A \in \mathsf{Conj}(C)$
   (b) *or there is a concept inclusion* $E \sqsubseteq F$ *in* $\mathcal{T}$ *and there is some index* $j < i$ *such that the matrix of* $\exists X_j.\mathcal{A}_j$ *entails* $E(x_C)$ *and* $A \in \mathsf{Conj}(F)$*.*
2. *If the matrix of* $\exists X_i.\mathcal{A}_i$ *contains the role assertion* $r(x_C, x_D)$*, then*
   (a) *either* $\exists r.D \in \mathsf{Conj}(C)$
   (b) *or there is a concept inclusion* $E \sqsubseteq F$ *in* $\mathcal{T}$ *and there is some index* $j < i$ *such that the matrix of* $\exists X_j.\mathcal{A}_j$ *entails* $E(x_C)$ *and* $\exists r.D \in \mathsf{Conj}(F)$*.*

*Proof.* Both statements are easy consequences of the definition of the three saturation rules (see Fig. 1).

First, note that the variable $x_C$ can only be created by an application of the $\exists$-rule, which simultaneously creates the assertion $C(x_C)$. There are three possibilities how the concept assertion $A(x_C)$ could have been created: $C = A$ and thus the assertion is created when $x_C$ is introduced; or $A \in \mathsf{Conj}(C)$ and a subsequent application of the $\sqcap$-rule creates $A(x_C)$; or later on the $\sqsubseteq$-rule is applied when $x_C$ matches the premise of a concept inclusion $E \sqsubseteq F$ in $\mathcal{T}$ and $A \in \mathsf{Conj}(F)$, which first creates the assertion $F(x_C)$ and, possibly after an application of the $\sqcap$-rule, the assertion $A(x_C)$.

Similarly, there are three ways how the role assertion $r(x_C, x_D)$ can have been created: $C = \exists r.D$, in which case an application of the $\exists$-rule creates the assertion; or $\exists r.D \in \mathsf{Conj}(C)$, i.e., first the $\sqcap$-rule is applied and a subsequent application of the $\exists$-rule produces the role assertion; or there is a concept inclusion $F \sqsubseteq G$ in $\mathcal{T}$ such that $x_D$ has eventually satisfied $F$ and $\exists r.E$ is a top-level conjunct in $G$, so that first applying the $\sqsubseteq$-rule at $x_D$ and then the $\sqcap$-rule followed by the $\exists$-rule produced the role assertion. □

**Lemma 24.** *Let* $x_C$ *be a variable in the saturation* $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$*. Then the following two statements hold:*

1. *If the matrix of* $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ *entails* $D(x_C)$*, then* $C \sqsubseteq^{\mathcal{T}} D$*.*
2. *If the matrix of* $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ *entails* $C(t)$*, then* $x_C \precsim t$*, i.e., there is a simulation on* $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ *that contains* $(x_C, t)$*.*

*Proof.* 1. We show the claim by induction on $D$ nested within an induction along the sequence of rule applications that produces $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ from $\exists X.\mathcal{A}$, say $\exists X.\mathcal{A} =: \exists X_0.\mathcal{A}_0 \to \exists X_1.\mathcal{A}_1 \to \cdots \to \exists X_n.\mathcal{A}_n \coloneqq \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$. The case where $D = \top$ is trivial, and the case where $D$ is a conjunction easily follows from the inner induction hypothesis.
- Assume that $D = A$ is a concept name where the matrix of $\exists X_i.\mathcal{A}_i$ entails $A(x_C)$. Lemma II in [4] shows that $\exists X_i.\mathcal{A}_i$ contains $A(x_C)$. According to Statement 1 in Lemma 23 there are two cases to consider.
  (a) If $A \in \mathsf{Conj}(C)$, then $C \sqsubseteq^{\mathcal{T}} A$ follows immediately.
  (b) If $A \in \mathsf{Conj}(F)$ for some concept inclusion $E \sqsubseteq F$ in $\mathcal{T}$ where the matrix of $\exists X_j.\mathcal{A}_j$ entails $E(x_C)$ for some index $j < i$, then the outer induction hypothesis yields that $C \sqsubseteq^{\mathcal{T}} E$. We conclude that $C \sqsubseteq^{\mathcal{T}} A$.

– Let $D = \exists r.\, E$ be an existential restriction where the matrix of $\exists X_i.\, \mathcal{A}_i$ entails $\exists r.\, E(x_C)$. By means of Lemma II in [4] we get a role assertion $r(x_C, x_F)$ in $\exists X_i.\, \mathcal{A}_i$ where $E(x_F)$ is entailed by $\exists X_i.\, \mathcal{A}_i$. The inner induction hypothesis yields that $F \sqsubseteq^{\mathcal{T}} E$. Regarding the role assertion, Statement 2 in Lemma 23 tells us that there are two cases as follows.
  (a) If $\exists r.\, F \in \mathsf{Conj}(C)$, then we immediately infer that $C \sqsubseteq^{\mathcal{T}} \exists r.\, E$.
  (b) If $\exists r.\, F \in \mathsf{Conj}(H)$ for a concept inclusion $G \sqsubseteq H$ in $\mathcal{T}$ such that the matrix of $\exists X_j.\, \mathcal{A}_j$ entails $G(x_C)$ for some index $j < i$, then the outer induction hypothesis yields that $C \sqsubseteq^{\mathcal{T}} G$. It follows that $C \sqsubseteq^{\mathcal{T}} \exists r.\, E$.

2. We show that the following relation $\mathfrak{W}$ is a simulation on $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$.

$$\begin{aligned}\mathfrak{W} \coloneqq \{\, (u, u) \mid u \in \Sigma_{\mathsf{O}}(\exists X.\, \mathcal{A}) \,\} \\ \cup \{\, (x_D, u) \mid \text{the matrix of } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A}) \text{ entails } D(u) \,\}\end{aligned}$$

Clearly, $\mathfrak{W}$ contains the pair $(x_C, t)$. By its very definition, $\mathfrak{W}$ contains $(a, a)$ for each individual name $a$, i.e., $\mathfrak{W}$ satisfies Condition (S1). We proceed with proving that $\mathfrak{W}$ fulfills the other two conditions of a simulation as well. For the pairs $(u, u)$ both are trivial. Now consider a pair $(x_D, u) \in \mathfrak{W}$, i.e., the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $D(u)$.

(S2) Let $A(x_D)$ in $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$. According to Statement 1 in Lemma 23, there are two cases.
  (a) In the first case, it holds that $A \in \mathsf{Conj}(D)$. Thus the fact that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $D(u)$ yields, using Lemma II in [4], that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ contains $A(u)$, and we are done.
  (b) In the second case, there is a concept inclusion $E \sqsubseteq F$ in $\mathcal{T}$ where the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $E(x_D)$[7] and $A \in \mathsf{Conj}(F)$. Statement 1 yields that $D \sqsubseteq^{\mathcal{T}} E$.
  Since the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $D(u)$, $D \sqsubseteq^{\mathcal{T}} E$, $E \sqsubseteq F \in \mathcal{T}$, and $A \in \mathsf{Conj}(F)$, we infer that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $A(u)$ w.r.t. $\mathcal{T}$.
  Since no saturation rule is applicable to $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$, Theorem 3 in [3] implies that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $A(u)$. Finally Lemma II in [4] allows us to conclude that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ contains $A(u)$.

(S3) Consider a role assertion $r(x_D, x_E)$ in $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$. According to Statement 2 in Lemma 23, it suffices to deal with the following two cases.
  (a) In the first case, we have $\exists r.\, E \in \mathsf{Conj}(D)$. It follows that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $\exists r.\, E(u)$.

[7] To be formally correct, there must be a partial saturation (before the final application of a saturation rule) the matrix of which entails $E(x_D)$. By an induction along the sequence of rule applications, it is easy to show that the final saturation entails each partial saturation before it. Thus we can infer that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\, \mathcal{A})$ entails $E(x_D)$.

(b) In the second case, there is a concept inclusion $F \sqsubseteq G$ such that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ entails $F(x_D)$ and $\exists r.E \in \mathsf{Conj}(G)$. This implies $D \sqsubseteq^{\mathcal{T}} F$ by Statement 1. We further conclude that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ entails $\exists r.E(u)$ w.r.t. $\mathcal{T}$. Since no saturation rule can be applied to $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$, Theorem 3 in [3] implies that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ entails $\exists r.E(u)$.

In both cases, Lemma II in [4] implies the existence of a role assertion $r(u,v)$ in $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ such that $E(v)$ is entailed by the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$. The latter implies $(x_E, v) \in \mathfrak{W}$ as needed.

$$\begin{array}{ccc} x_D & \xrightarrow{\;r\;} & x_E \\ {\scriptstyle\mathfrak{W}}\downarrow & & \downarrow{\scriptstyle\mathfrak{W}} \\ u & \xrightarrow[\;r\;]{} & v \end{array}$$

□

**Theorem 18.** *Let $\exists X.\mathcal{A}$ be a qABox with set of individuals $\Sigma_{\mathsf{I}}$, let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, and let $\mathcal{B}_a$ for all $a \in \Sigma_{\mathsf{I}}$ be the ABoxes introduced in Definition 14. Then $\exists X.\mathcal{A}$ has an optimal ABox approximation w.r.t. $\mathcal{T}$ iff, for all individuals $a \in \Sigma_{\mathsf{I}}$, the msc of $a$ in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$ exists. If the latter condition is satisfied and $C_a$ are these most specific concepts, then the following ABox is an optimal ABox approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$:*

$$\{\, C_a(a) \mid a \in \Sigma_{\mathsf{I}} \,\} \cup \{\, r(a,b) \mid r(a,b) \text{ occurs in } \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \text{ where } a,b \in \Sigma_{\mathsf{I}} \,\}.$$

*In particular, the existence of an optimal ABox approximation can be tested in polynomial time and such an optimal approximation can be computed in exponential time if it exists.*

*Proof.* In the following, assume that $Y$, $\mathcal{B}$, and $\mathcal{B}_a$ for each individual name $a$ are defined as in Definition 14.

We start with the if direction and therefore assume that, for each individual name $a$, the concept $C_a$ is the most specific concept of $a$ in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$. We denote the ABox in the theorem by $\mathcal{C}$, and we will show that it is an optimal approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$, i.e., that $\exists X.\mathcal{A} \equiv^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{C}$.

By assumption, it holds that $\mathcal{B}_a \models^{\mathcal{T}} C_a(a)$ for each individual name $a$ and so it follows that $\exists Y.\mathcal{B}_a \models^{\mathcal{T}} C_a(a)$. Since each $\exists Y.\mathcal{B}_a$ is a sub-qABox of $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$, we thus infer that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}} \{\, C_a(a) \mid a \text{ is an individual name} \,\}$. Furthermore, the ABox $\mathcal{C}$ and the pre-approximation $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ contain the same role assertions. We conclude that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}} \mathcal{C}$ and thus that also $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{C}$, using Proposition 3.

An application of Lemma 15 now yields that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{C}$. By means of Proposition 2 it follows that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}}_{\mathsf{IQ}} \mathcal{C}$ and that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ contains all role assertions occurring in $\mathcal{C}$. Compared to $\exists X.\mathcal{A}$, the saturation $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ does not contain any additional role assertions involving only individual names. So already $\exists X.\mathcal{A}$ contains all role assertions occurring in $\mathcal{C}$. Furthermore, we infer from $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}}_{\mathsf{IQ}} \mathcal{C}$ by means of Theorem 3 in [3] that

$\mathsf{sat}^{\mathcal{T}}_{\mathsf{IQ}}(\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})) \models^{\emptyset}_{\mathsf{IQ}} \mathcal{C}$. Since no saturation rule is applicable to $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$, the two saturations $\mathsf{sat}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ and $\mathsf{sat}^{\mathcal{T}}_{\mathsf{IQ}}(\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}))$ must be equal. We obtain that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \models^{\emptyset}_{\mathsf{IQ}} \mathcal{C}$, and another application of Theorem 3 in [3] yields that $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IQ}} \mathcal{C}$. In summary, we conclude that $\exists X.\mathcal{A} \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{C}$.

It remains to show the converse entailment, namely that $\mathcal{C} \models^{\mathcal{T}}_{\mathsf{IRQ}} \exists X.\mathcal{A}$. Since $\exists X.\mathcal{A}$ and $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ are IRQ-equivalent w.r.t. $\mathcal{T}$ (see Lemma 15), we can equivalently show that $\mathcal{C} \models^{\mathcal{T}}_{\mathsf{IRQ}} \mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ holds. It is easy to see that $\mathcal{C}$ and $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ contain the same role assertions involving only individual names. It remains to show IQ-entailment. Therefore, consider a concept assertion $C(a)$ such that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models^{\mathcal{T}} C(a)$. By Lemma 22 in [13] (or rather Lemma V in [4]) there exists a concept $D$ such that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models D(a)$ and $D \sqsubseteq^{\mathcal{T}} C$. We now consider the top-level conjuncts of $D$ one by one.

1. Consider a concept name $A \in \mathsf{Conj}(D)$. It holds that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models A(a)$ and so Lemma II in [4] yields that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ contains $A(a)$. Looking at Definition 14, we infer that $A(a)$ must be contained in $\mathcal{B}_a$, and thus $\mathcal{B}_a \models A(a)$. Since $C_a$ is the most specific concept of $a$, we conclude that $C_a \sqsubseteq^{\mathcal{T}} A$ and further that $\{C_a(a)\} \models^{\mathcal{T}} A(a)$. Since $\mathcal{C}$ contains $\{C_a(a)\}$ as a sub-ABox, we obtain that $\mathcal{C} \models^{\mathcal{T}} A(a)$.
2. Consider an existential restriction $\exists r.E \in \mathsf{Conj}(D)$. It holds that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \models \exists r.E(a)$. According to Lemma II in [4], there exists a role assertion $r(a,v)$ in $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ such that $E(v)$ is entailed by the matrix of $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$. In accordance with Definition 14, we proceed with a case distinction on the role assertion $r(a,v)$.
   (a) Assume that $v = b$ is an individual name. Then the induction hypothesis yields that $\mathcal{C} \models^{\mathcal{T}} E(b)$. Since $\mathcal{C}$ also contains $r(a,b)$, we conclude that $\mathcal{C} \models^{\mathcal{T}} \exists r.E(a)$.
   (b) Otherwise, we have $v = t'$ where $r(a,t)$ is indispensable. Specifically, $r(a,t')$ is contained in $\mathcal{B}_a$. Within the whole matrix $\mathcal{B}$, we cannot reach from $t'$ any objects outside the part $\mathcal{B}_a$. Since the matrix of $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ entails $E(t')$, it thus holds that already the part $\mathcal{B}_a$ must entail $E(t')$. It follows that $\mathcal{B}_a \models \exists r.E(a)$. As $C_a$ is the most specific concept of $a$, we conclude that $C_a \sqsubseteq^{\mathcal{T}} \exists r.E$ and further that $\{C_a(a)\} \models^{\mathcal{T}} \exists r.E(a)$. Since the singleton $\{C_a(a)\}$ is a sub-ABox of $\mathcal{C}$, it follows that $\mathcal{C} \models^{\mathcal{T}} \exists r.E(a)$.

In summary, we have shown that $\mathcal{C} \models^{\mathcal{T}} F(a)$ for each $F \in \mathsf{Conj}(D)$, which implies that $\mathcal{C} \models^{\mathcal{T}} D(a)$. Together with $D \sqsubseteq^{\mathcal{T}} C$, this yields $\mathcal{C} \models^{\mathcal{T}} C(a)$ as needed.

We continue with the only-if direction. Specifically, assume that $\mathcal{C}$ is an optimal approximation of $\exists X.\mathcal{A}$ w.r.t. $\mathcal{T}$. We first construct a fixed saturation $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$ and from it the pre-approximation $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$ as per Definition 14,[8] where we denote the parts as $Y'$, $\mathcal{B}'$, and $\mathcal{B}'_a$ for each individual name $a$.

We subdivide the proof into the following three steps:

I. For each individual $a$, the most specific concept of $a$ in $\mathcal{B}'_a$ w.r.t. $\mathcal{T}$ exists.
II. For each individual $a$, the sub-qABoxes $\exists Y.\mathcal{B}_a$ and $\exists Y'.\mathcal{B}'_a$ are $\mathsf{IQ}$-equivalent w.r.t. $\mathcal{T}$.
III. For each individual $a$, the most specific concept of $a$ in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$ exists.

By assumption, it holds that $\exists X.\mathcal{A} \equiv^{\mathcal{T}}_{\mathsf{IRQ}} \mathcal{C}$ and so we conclude that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A}) \equiv^{\emptyset}_{\mathsf{IRQ}} \underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$. By means of Lemma 15 it follows that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A}) \equiv^{\emptyset}_{\mathsf{IRQ}} \mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$. Specifically, this implies $\mathsf{IQ}$-equivalence and so there is a simulation $\mathfrak{U}$ from $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ to $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$ as well as a simulation $\mathfrak{V}$ in the converse direction. Furthermore, let $\mathfrak{S}$ and $\mathfrak{T}$ be the simulations in the proof of Lemma 15 from $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\exists X.\mathcal{A})$ to $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\exists X.\mathcal{A})$ and in the converse direction, respectively, and likewise let $\mathfrak{S}'$ and $\mathfrak{T}'$ be the simulations between $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$ and $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$.

I. Now let $a$ be an individual name. Define the concept

$$C'_a \coloneqq \bigsqcap\{\, C \mid C(a) \in \mathcal{C} \,\} \sqcap \bigsqcap\{\, E \mid D \sqsubseteq E \in \mathcal{T} \text{ and } \mathcal{C} \models^{\mathcal{T}} D(a) \,\},$$

and then let $C_a$ be obtained from $C'_a$ by removing each top-level conjunct $\exists r.F$ where $\mathcal{C}$ contains a role assertion $r(a,b)$ such that $b$ is an individual name and $\mathcal{C} \models^{\mathcal{T}} F(b)$. We are going to show that $C_a$ is the most specific concept of $a$ in $\mathcal{B}'_a$ w.r.t. $\mathcal{T}$.

1. By construction, it holds that $\mathcal{C} \models^{\mathcal{T}} C'_a(a)$ and thus also that $\mathcal{C} \models^{\mathcal{T}} C_a(a)$. Using Theorem 3 in [3], we further conclude that $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C}) \models C_a(a)$ and so Lemma 15 yields that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C}) \models C_a(a)$.
   (a) Consider a concept name $A \in \mathsf{Conj}(C_a)$. By Lemma II in [4] we infer that $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$ contains the concept assertion $A(a)$. According to Definition 14, this assertion $A(a)$ can only be contained in $\mathcal{B}'_a$, and thus $\mathcal{B}'_a \models A(a)$.
   (b) Now consider an existential restriction $\exists r.F \in \mathsf{Conj}(C_a)$. Lemma II in [4] yields a role assertion $r(a,v)$ in $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$ such that $F(v)$ is entailed by $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$. The construction of $C_a$ from $C'_a$ ensures that $v$ must be a variable.[9] Specifically, it holds that $v = t'$ where $r(a,t)$ is an indispensable role assertion (in $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$). With similar arguments as in Case 2b above, it follows that $\mathcal{B}'_a \models \exists r.F(a)$.[10]

[8] Formally, we beforehand need to transform $\mathcal{C}$ into an $\mathsf{IRQ}$-equivalent quantified ABox, namely by exhaustively applying the $\sqcap$-rule and the $\exists$-rule in Figure 1.

[9] Formally: Assume that $v = b$ were an individual name. Then $r(a,b)$ would also be contained in $\mathcal{C}$. Furthermore, $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C}) \models F(b)$ implies $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C}) \models F(b)$ by Lemma 15, and we conclude $\mathcal{C} \models^{\mathcal{T}} F(b)$ by Theorem 3 in [3] — a contradiction.

[10] Specifically, $r(a,t')$ is contained in $\mathcal{B}'_a$. Within the whole matrix $\mathcal{B}'$, we cannot reach from $t'$ any objects outside the part $\mathcal{B}'_a$. Since the matrix of $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$ entails

We have shown that $a$ is an instance of each top-level conjunct of $C_a$ w.r.t. $\mathcal{B}'_a$, and so it holds that $\mathcal{B}'_a \models C_a(a)$. Clearly, this implies $\mathcal{B}'_a \models^{\mathcal{T}} C_a(a)$.

2. Consider a concept assertion $G(a)$ such that $\mathcal{B}'_a \models^{\mathcal{T}} G(a)$. With Lemma 22 in [13] we infer that there is a concept $H$ such that $\mathcal{B}'_a \models H(a)$ and $H \sqsubseteq^{\mathcal{T}} G$. We are going to show that $C_a \sqsubseteq^{\mathcal{T}} H$, which then implies $C_a \sqsubseteq^{\mathcal{T}} G$ as needed.
   (a) Let $A \in \mathsf{Conj}(H)$. Then $\mathcal{B}'_a$ contains the concept assertion $A(a)$, cf. Lemma II in [4]. From Definition 14, it follows that the saturation $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$ contains $A(a)$. Thus, either $A(a)$ is already contained in (the qABox-normalization of) $\mathcal{C}$, or it has been created by an application of the $\sqsubseteq$-rule to $a$ possibly followed by an application of the $\sqcap$-rule. That is, either there is a concept assertion $C(a)$ in $\mathcal{C}$ such that $A \in \mathsf{Conj}(C)$, or there is a concept inclusion $D \sqsubseteq E$ in $\mathcal{T}$ such that $A \in \mathsf{Conj}(E)$ and, during the construction of the saturation, $a$ eventually satisfies $D$, i.e., $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C}) \models D(a)$. The latter implies $\mathcal{C} \models^{\mathcal{T}} D(a)$, cf. Theorem 3 in [3]. In both cases, it follows that $A$ is a top-level conjunct in $C_a$, and thus $C_a \sqsubseteq^{\mathcal{T}} A$.
   (b) Let $\exists r.K \in \mathsf{Conj}(H)$. By Lemma II in [4], there is a role assertion $r(a,t')$ in $\mathcal{B}'_a$ such that $\mathcal{B}'_a$ entails $K(t')$. Specifically, the saturation $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$ contains the indispensable role assertion $r(a,t)$ that must correspond either to an existential restriction in $\mathsf{Conj}(C)$ for some $C(a)$ in $\mathcal{C}$ or to an existential restriction in $\mathsf{Conj}(E)$ for some $D \sqsubseteq E$ in $\mathcal{T}$ where $a$ eventually matches $D$ during the construction of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$. That is, we have $t = x_L$ where either $\exists r.L \in \mathsf{Conj}(C)$ for some $C(a) \in \mathcal{C}$, or $\exists r.L \in \mathsf{Conj}(E)$ such that $D \sqsubseteq E \in \mathcal{T}$ and $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C}) \models D(a)$. Specifically, this means that $\exists r.L$ is a top-level conjunct in $C'_a$.
   Next, we show that $\exists r.L$ is also a top-level conjunct in $C_a$, which boils down to proving that $\mathcal{C}$ does not contain a role assertion $r(a,b)$ such that $\mathcal{C} \models^{\mathcal{T}} L(b)$. Assume that there were such an assertion. According to Theorem 3 in [3], $\mathcal{C} \models^{\mathcal{T}} L(b)$ would imply $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C}) \models L(b)$. By means of Statement 2 in Lemma 24 we would infer that $x_L \precsim b$ — a contradiction to indispensability of $r(a,x_L)$.
   Now note that the simulation $\mathfrak{T}'$ from $\mathsf{pre\text{-}approx}^{\mathcal{T}}_{\mathsf{IRQ}}(\mathcal{C})$ to $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$ contains $(t',t)$, cf. the proof of Lemma 15. It follows that $\mathcal{B}'_a \models K(t')$ implies that the matrix of $\underline{\mathsf{sat}}^{\mathcal{T}}_{\mathsf{IQ}}(\mathcal{C})$ entails $K(x_L)$. According to Statement 1 of Lemma 24, the latter yields $L \sqsubseteq^{\mathcal{T}} K$. We have shown above that $\exists r.L$ is a top-level conjunct in $C_a$, and so we conclude that $C_a \sqsubseteq^{\mathcal{T}} \exists r.K$.

II. Next, we show that $\exists Y.\mathcal{B}_a$ and $\exists Y'.\mathcal{B}'_a$ are $\mathsf{IQ}$-equivalent for each individual name $a$. First of all, we show that (copies of) indispensable role assertions can only be simulated by (copies of) indispensable role assertions.

---

$F(t')$, it thus holds that already the part $\mathcal{B}'_a$ must entail $F(t')$. It follows that $\mathcal{B}'_a \models \exists r.F(a)$.

Consider a role assertion $r(a,t')$ in $\mathcal{B}_a$, where $r(a,t)$ is indispensable in $\underline{\mathsf{sat}}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X.\mathcal{A})$. The simulation $\mathfrak{U}$ provides us with an object $v$ such that $r(a,v)$ is in $\mathsf{pre\text{-}approx}_{\mathsf{IRQ}}^{\mathcal{T}}(\mathcal{C})$ and $(t',v) \in \mathfrak{U}$. If $v = b$ were an individual name, then the role assertion $r(a,b)$ would also be contained in $\underline{\mathsf{sat}}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X.\mathcal{A})$. Furthermore, we would have $(t,t') \in \mathfrak{S}$, $(t',b) \in \mathfrak{U}$, and $(b,b) \in \mathfrak{V}$, and would conclude that $(t,b) \in \mathfrak{S} \circ \mathfrak{U} \circ \mathfrak{V}$ where the composition $\mathfrak{S} \circ \mathfrak{U} \circ \mathfrak{V}$ is a simulation on $\underline{\mathsf{sat}}_{\mathsf{IQ}}^{\mathcal{T}}(\exists X.\mathcal{A})$, i.e., $t' \precsim b$ holds — a contradiction since $r(a,t)$ is indispensable. We conclude that $v = u'$ must be a variable where $r(a,u)$ is indispensable in $\underline{\mathsf{sat}}_{\mathsf{IQ}}^{\mathcal{T}}(\mathcal{C})$. Summing up, (copies of) indispensable role assertions contained in $\mathcal{B}_a$ can only be simulated by (copies of) indispensable role assertions contained in $\mathcal{B}'_a$. The converse can be shown analogously.

It is now a finger exercise to show that the restriction

$$\mathfrak{U}_a \coloneqq \mathfrak{U} \cap (\Sigma_{\mathsf{O}}(\exists Y.\mathcal{B}_a) \times \Sigma_{\mathsf{O}}(\exists Y'.\mathcal{B}'_a))$$

is a simulation from $\exists Y.\mathcal{B}_a$ to $\exists Y'.\mathcal{B}'_a$, and likewise that the restriction $\mathfrak{V}_a \coloneqq \mathfrak{V} \cap (\Sigma_{\mathsf{O}}(\exists Y'.\mathcal{B}'_a) \times \Sigma_{\mathsf{O}}(\exists Y.\mathcal{B}_a))$ is a simulation from $\exists Y'.\mathcal{B}'_a$ to $\exists Y.\mathcal{B}_a$, both as needed.

III. We have shown above that $C_a$ is the most specific concept of $a$ in $\mathcal{B}'_a$ w.r.t. $\mathcal{T}$. Since $\exists Y'.\mathcal{B}'_a$ and $\exists Y.\mathcal{B}_a$ are $\mathsf{IQ}$-equivalent w.r.t. $\mathcal{T}$ and $a$ is an individual name in both qABoxes, we infer that $C_a$ is also the most specific concept of $a$ in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$. A formal proof is as follows.
   (1) From $\exists Y.\mathcal{B}_a \models_{\mathsf{IQ}}^{\mathcal{T}} \exists Y'.\mathcal{B}'_a \models^{\mathcal{T}} C_a(a)$ we infer that $\mathcal{B}_a \models^{\mathcal{T}} C_a(a)$.
   (2) Consider a concept assertion $M(a)$ where $\mathcal{B}_a \models^{\mathcal{T}} M(a)$. With $\exists Y'.\mathcal{B}'_a \models_{\mathsf{IQ}}^{\mathcal{T}} \exists Y.\mathcal{B}_a$ it follows that $\mathcal{B}'_a \models^{\mathcal{T}} M(a)$. Since $C_a$ is the most specific concept of $a$ in $\mathcal{B}'_a$ w.r.t. $\mathcal{T}$, we infer that $C_a \sqsubseteq^{\mathcal{T}} M$. □

In order to illustrate that, for each individual name $a$, the most specific concept of $a$ in $\mathcal{B}_a$ w.r.t. $\mathcal{T}$ cannot be simply read off from an optimal approximation, we provide the following example.

*Example 25.* The quantified ABox, which is to be approximated, is $\exists\emptyset.\{r(a,b), A(a)\}$. The TBox is $\mathcal{T} \coloneqq \{\exists r.\top \sqsubseteq A\}$. Note that $\exists X.\mathcal{A}$ is already saturated w.r.t. $\mathcal{T}$, and there are no indispensable role assertions. Then, the ABox $\{r(a,b)\}$ is already an optimal approximation, but the msc of $a$ in $\mathcal{B}_a = \{A(a)\}$ is $A$, which cannot be read off from $\{r(a,b)\}$.