Conjunctive Query Inseparability in OWL 2 QL is ExpTime-hard

Roman Kontchakov[1], Vladislav Ryzhikov[2], and Michael Zakharyaschev[1]

[1] Department of Computer Science and Information Systems, Birkbeck, University of London, U.K.
[2] Faculty of Computer Science, Free University of Bozen-Bolzano, Italy

Abstract. We settle an open question on the complexity of the following problem: given two OWL 2 QL TBoxes and a signature, decide whether these TBoxes return the same answers to any conjunctive query over any data formulated in the given signature. It has been known that the complexity of this problem is between PSpace and ExpTime. Here we show that the problem is ExpTime-complete and, in fact, deciding whether two OWL 2 QL knowledge bases (each with its own data) give the same answers to any conjunctive query is ExpTime-hard.

Keywords: Query inseparability, OWL 2 QL, complexity.

1 Introduction

In this paper, we show that the problem of deciding whether two given OWL 2 QL ontologies return the same answers to any conjunctive query in a given signature is ExpTime-hard. The notion of conjunctive query inseparability has recently been used in two settings: ontology engineering and maintenance, and knowledge base exchange.

It has been argued [6, 8, 4, 5] that conjunctive query inseparability is fundamental for many ontology engineering and maintenance tasks in the context of ontology-based data access (OBDA). Indeed, suppose we want to query data via some ontology $\mathcal{T}$, and the set of concepts and roles we are interested in comprises a signature Σ. From our point of view, any operations with $\mathcal{T}$ such as refining $\mathcal{T}$, extracting a smaller module, importing a module, updating $\mathcal{T}$ to a new version, etc. should preserve answers to conjunctive queries formulated in Σ. In other words, all such transformations of $\mathcal{T}$ should result in ontologies that are Σ-query inseparable from $\mathcal{T}$.

In the knowledge exchange framework [2], the problem of inseparability by the unions of conjunctive queries (UCQ) for two OWL 2 QL knowledge bases (KBs) appears in the context of computing a universal UCQ-solution. Let Σ_1 and Σ_2 be a *source* and a *target signature*, respectively. The two signatures are disjoint and connected by means of a declarative mapping specification, which is a TBox $\mathcal{T}_{12}$ relating concepts and roles in Σ_1 and Σ_2. A KB $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A}_2)$ in the target signature Σ_2 is called a *universal UCQ-solution* for a KB $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A}_1)$

in the source signature Σ_1 under $\mathcal{T}_{12}$ if $(\mathcal{T}_1 \cup \mathcal{T}_{12}, \mathcal{A}_1)$ and $(\mathcal{T}_2, \mathcal{A}_2)$ give the same answers to any UCQ formulated in Σ_2. Thus, in this case we are interested in Σ-query inseparability for OWL 2 QL KBs.

The Σ-query inseparability problem for ontologies (TBoxes) in the fragment of OWL 2 QL without role inclusions proves to be quite simple: it is NLogSpace-complete and can be solved in practice using various existing reasoners [6, 4]. The addition to this language conjunctions in the left-hand side of concept inclusions and (unqualified) number restrictions makes the problem coNP-complete [6]. But for the full OWL 2 QL the problem unexpectedly turned out to be much more challenging and required new logical tools compared to the previously analysed ontology languages. It was shown [4] that the interaction between role inclusions and inverse roles makes this problem PSpace-hard; on the other hand, the established upper bound was obtained by a reduction to the emptiness problem for alternating two-way automata, which belongs to ExpTime [9]. The proof of PSpace-hardeness mentioned above was also adapted to query inseparability of OWL 2 QL KBs [1].

In this paper, we prove, by encoding alternating Turing machines with polynomial tape, that the Σ-query inseparability problem for both OWL 2 QL KBs and ontologies is ExpTime-hard.

2 Σ-Query Entailment and Inseparability

We use the following (somewhat simplified) syntax of OWL 2 QL. It contains *individual names* a_i, *concept names* A_i, and *role names* P_i ($i \geq 1$). *Roles* R and *basic concepts* B are defined by the grammar:

$$R \quad ::= \quad P_i \quad \mid \quad P_i^-, \qquad\qquad B \quad ::= \quad \bot \quad \mid \quad A_i \quad \mid \quad \exists R.$$

A *TBox*, $\mathcal{T}$, is a finite set of *inclusions* of the form

$$B_1 \sqsubseteq B_2, \qquad B_1 \sqcap B_2 \sqsubseteq \bot, \qquad R_1 \sqsubseteq R_2, \qquad R_1 \sqcap R_2 \sqsubseteq \bot.$$

An *ABox*, $\mathcal{A}$, is a finite set of atoms of the form $A_k(a_i)$ or $P_k(a_i, a_j)$. The set of individual names in $\mathcal{A}$ is denoted by $\mathsf{ind}(\mathcal{A})$. $\mathcal{T}$ and $\mathcal{A}$ together form the *knowledge base* (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. The semantics for OWL 2 QL is defined in the usual way based on interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ [3]. We use $S_1 \equiv S_2$ as a shortcut for $S_1 \sqsubseteq S_2$ and $S_2 \sqsubseteq S_1$ (for both concepts and roles), and write $\mathcal{I} \models \alpha$ to say that an inclusion or assertion α is true in $\mathcal{I}$. An interpretation $\mathcal{I}$ is a *model* of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{T} \cup \mathcal{A}$; in this case we write $\mathcal{I} \models \mathcal{K}$. $\mathcal{K}$ is *consistent* if it has a model. A concept B is said to be $\mathcal{T}$-*consistent* if $(\mathcal{T}, \{B(a)\})$ has a model. $\mathcal{K} \models \alpha$ means that $\mathcal{I} \models \alpha$ for all models $\mathcal{I}$ of $\mathcal{K}$.

A *conjunctive query* (CQ) $\boldsymbol{q}(\boldsymbol{x})$ is a formula $\exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$, where φ is a conjunction of atoms of the form $A_k(z_1)$ or $P_k(z_1, z_2)$ with $z_i \in \boldsymbol{x} \cup \boldsymbol{y}$. A tuple $\boldsymbol{a} \subseteq \mathsf{ind}(\mathcal{A})$ is a *certain answer* to $\boldsymbol{q}(\boldsymbol{x})$ over $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if $\mathcal{I} \models \boldsymbol{q}(\boldsymbol{a})$ for all $\mathcal{I} \models \mathcal{K}$; in this case we write $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$. If $\boldsymbol{x} = \emptyset$, the CQ $\boldsymbol{q}$ is called *Boolean*; a certain answer to such a $\boldsymbol{q}$ over $\mathcal{K}$ is 'yes' if $\mathcal{K} \models \boldsymbol{q}$ and 'no' otherwise.

To define the main notions of this paper, consider two KBs $\mathcal{K}_1 = (\mathcal{T}_1, \mathcal{A})$ and $\mathcal{K}_2 = (\mathcal{T}_2, \mathcal{A})$. For example, the $\mathcal{T}_i$ are different versions of some ontology, or one of them is a refinement of the other by means of new axioms. The question we are interested in is whether they give the same answers to CQs formulated in a certain signature, say, in the common vocabulary of the $\mathcal{T}_i$ or in a vocabulary relevant to an application. More precisely, by a *signature*, Σ, we understand any finite set of concept and role names. A concept (inclusion, TBox, etc.) all concept and role names of which are in Σ is called a Σ-*concept* (*inclusion*, etc.). We say that $\mathcal{K}_1$ Σ-*query entails* $\mathcal{K}_2$ if, for *all* Σ-queries $\boldsymbol{q}(\boldsymbol{x})$ and all $\boldsymbol{a} \subseteq \mathsf{ind}(\mathcal{A})$, $\mathcal{K}_2 \models \boldsymbol{q}(\boldsymbol{a})$ implies $\mathcal{K}_1 \models \boldsymbol{q}(\boldsymbol{a})$. In other words: any certain answer to a Σ-query given by $\mathcal{K}_2$ is also given by $\mathcal{K}_1$. We call $\mathcal{K}_1$ and $\mathcal{K}_2$ Σ-*query inseparable* if they Σ-query entail each other, in which case we write $\mathcal{K}_1 \equiv_\Sigma \mathcal{K}_2$.

As the ABox is typically not fixed or known at the ontology design stage, we may have to compare the TBoxes over *arbitrary* Σ-ABoxes rather than a fixed one, which gives the following definition. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes and Σ a signature. $\mathcal{T}_1$ Σ-*query entails* $\mathcal{T}_2$ if $(\mathcal{T}_1, \mathcal{A})$ Σ-query entails $(\mathcal{T}_2, \mathcal{A})$ for any Σ-ABox $\mathcal{A}$. $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-*query inseparable* if they Σ-query entail each other, in which case we write $\mathcal{T}_1 \equiv_\Sigma \mathcal{T}_2$.

In the remainder of this section, we recap the semantic criteria of Σ-query entailment established in [4]. Recall first that to compute certain answers to any CQ $\boldsymbol{q}$ over a consistent KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, it is enough to find answers to $\boldsymbol{q}$ in the *canonical model* $\mathcal{C}_\mathcal{K}$ of $\mathcal{K}$.

Let $[R] = \{S \mid \mathcal{T} \models R \equiv S\}$. We write $[R] \leq_\mathcal{T} [S]$ if $\mathcal{T} \models R \sqsubseteq S$; thus, $\leq_\mathcal{T}$ is a partial order on the set $\{[R] \mid R \text{ a role in } \mathcal{T}\}$. For each $[R]$, we introduce a *witness* $w_{[R]}$ and define a *generating relation* $\rightsquigarrow_\mathcal{K}$ on the set of these witnesses together with $\mathsf{ind}(\mathcal{A})$ by taking:

$a \rightsquigarrow_\mathcal{K} w_{[R]}$ if $a \in \mathsf{ind}(\mathcal{A})$ and $[R]$ is $\leq_\mathcal{T}$-minimal such that $\mathcal{K} \models \exists R(a)$ and $\mathcal{K} \not\models R(a, b)$ for any $b \in \mathsf{ind}(\mathcal{A})$;
$w_{[S]} \rightsquigarrow_\mathcal{K} w_{[R]}$ if $[R]$ is $\leq_\mathcal{T}$-minimal with $\mathcal{T} \models \exists S^- \sqsubseteq \exists R$ and $[S^-] \neq [R]$.

Clearly, $\rightsquigarrow_\mathcal{K}$ can be computed in polynomial time in $|\mathcal{K}|$. A $\mathcal{K}$-*path* is a finite sequence $a w_{[R_1]} \cdots w_{[R_n]}$, $n \geq 0$, such that $a \in \mathsf{ind}(\mathcal{A})$ and, if $n > 0$, then $a \rightsquigarrow_\mathcal{K} w_{[R_1]}$ and $w_{[R_i]} \rightsquigarrow_\mathcal{K} w_{[R_{i+1}]}$, for $i < n$. Denote by $\mathsf{tail}(\sigma)$ the last element in the path σ. The canonical model $\mathcal{C}_\mathcal{K}$ of $\mathcal{K}$ is defined by taking $\Delta^{\mathcal{C}_\mathcal{K}}$ to be the set of all $\mathcal{K}$-paths and setting:

$$\begin{aligned}
a^{\mathcal{C}_\mathcal{K}} &= a, \text{ for all } a \in \mathsf{ind}(\mathcal{A}),\\
A^{\mathcal{C}_\mathcal{K}} &= \{a \in \mathsf{ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup\\
&\quad \{\sigma \cdot w_{[R]} \mid \mathcal{T} \models \exists R^- \sqsubseteq A\}, \text{ for all concept names } A,\\
P^{\mathcal{C}_\mathcal{K}} &= \{(a, b) \in \mathsf{ind}(\mathcal{A}) \times \mathsf{ind}(\mathcal{A}) \mid R(a, b) \in \mathcal{A} \text{ with } [R] \leq_\mathcal{T} [P]\} \cup\\
&\quad \{(\sigma, \sigma \cdot w_{[R]}) \mid \mathsf{tail}(\sigma) \rightsquigarrow_\mathcal{K} w_{[R]},\ [R] \leq_\mathcal{T} [P]\} \cup\\
&\quad \{(\sigma \cdot w_{[R]}, \sigma) \mid \mathsf{tail}(\sigma) \rightsquigarrow_\mathcal{K} w_{[R]},\ [R] \leq_\mathcal{T} [P^-]\}, \text{ for all role names } P.
\end{aligned}$$

Theorem 1. *For all consistent OWL 2 QL KBs $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, CQs $\boldsymbol{q}(\boldsymbol{x})$ and tuples $\boldsymbol{a} \subseteq \mathsf{ind}(\mathcal{A})$, we have $\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$ iff $\mathcal{C}_\mathcal{K} \models \boldsymbol{q}(\boldsymbol{a})$.*

Thus, to decide Σ-query entailment between KBs $\mathcal{K}_1$ and $\mathcal{K}_2$, it suffices to check whether $\mathcal{C}_{\mathcal{K}_2} \models \boldsymbol{q}(\boldsymbol{a})$ implies $\mathcal{C}_{\mathcal{K}_1} \models \boldsymbol{q}(\boldsymbol{a})$ for all Σ-queries $\boldsymbol{q}(\boldsymbol{x})$ and tuples $\boldsymbol{a}$. This relationship between $\mathcal{C}_{\mathcal{K}_2}$ and $\mathcal{C}_{\mathcal{K}_1}$ can be characterised semantically in terms of finite Σ-homomorphisms.

For an interpretation $\mathcal{I}$ and a signature Σ, the *Σ-types* $\boldsymbol{t}^{\mathcal{I}}_{\Sigma}(x)$ and $\boldsymbol{r}^{\mathcal{I}}_{\Sigma}(x, y)$, for $x, y \in \Delta^{\mathcal{I}}$, are given by:

$$\boldsymbol{t}^{\mathcal{I}}_{\Sigma}(x) = \{\Sigma\text{-concept } B \mid x \in B^{\mathcal{I}}\}, \quad \boldsymbol{r}^{\mathcal{I}}_{\Sigma}(x, y) = \{\Sigma\text{-role } R \mid (x, y) \in R^{\mathcal{I}}\}.$$

A *Σ-homomorphism* from an interpretation $\mathcal{I}$ to $\mathcal{I}'$ is a function $h\colon \Delta^{\mathcal{I}} \to \Delta^{\mathcal{I}'}$ such that $h(a^{\mathcal{I}}) = a^{\mathcal{I}'}$, for all individual names a interpreted in $\mathcal{I}$,

$$\boldsymbol{t}^{\mathcal{I}}_{\Sigma}(x) \subseteq \boldsymbol{t}^{\mathcal{I}'}_{\Sigma}(h(x)) \quad \text{and} \quad \boldsymbol{r}^{\mathcal{I}}_{\Sigma}(x, y) \subseteq \boldsymbol{r}^{\mathcal{I}'}_{\Sigma}(h(x), h(y)), \qquad \text{for all } x, y \in \Delta^{\mathcal{I}}.$$

It is known that answers to Σ-CQs are preserved under Σ-homomorphisms. Thus, if there is a Σ-homomorphism from $\mathcal{C}_{\mathcal{K}_2}$ to $\mathcal{C}_{\mathcal{K}_1}$, then $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$. However, it was observed [4] that the converse does not hold in general and a more subtle notion of homomorphism is required. We say that $\mathcal{I}$ is *finitely Σ-homomorphically embeddable into $\mathcal{I}'$* if, for every finite sub-interpretation $\mathcal{I}_1$ of $\mathcal{I}$, there exists a Σ-homomorphism from $\mathcal{I}_1$ to $\mathcal{I}'$.

Theorem 2 ([4]). *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be consistent OWL 2 QL KBs. Then $\mathcal{K}_1$ Σ-query entails $\mathcal{K}_2$ iff $\mathcal{C}_{\mathcal{K}_2}$ is finitely Σ-homomorphically embeddable into $\mathcal{C}_{\mathcal{K}_1}$.*

To obtain a similar criterion for TBoxes (rather than KBs), we use the fact that inclusions in OWL 2 QL, different from disjointness axioms, involve only *one* concept or role in the left-hand side and making sure that the TBoxes entail the same Σ-inclusions, one can show that it is enough to consider *singleton* Σ-ABoxes of the form $\{B(a)\}$.

Theorem 3 ([4]). *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be OWL 2 QL TBoxes. Then $\mathcal{T}_1$ Σ-query entails $\mathcal{T}_2$ iff the following conditions hold*:

- **(p)** *$\mathcal{T}_2 \models \alpha$ implies $\mathcal{T}_1 \models \alpha$, for all Σ-inclusions α*;
- **(h)** *$\mathcal{C}_{(\mathcal{T}_2, \{B(a)\})}$ is finitely Σ-homomorphically embeddable into $\mathcal{C}_{(\mathcal{T}_1, \{B(a)\})}$, for all $\mathcal{T}_1$-consistent Σ-concepts B.*

It was shown [4] that checking Σ-query entailment for OWL 2 QL TBoxes is PSpace-hard and can be done in ExpTime.

3 ExpTime-hardness

Our main result is the following theorem:

Theorem 4. *The Σ-query entailment and Σ-query inseparability problems are* ExpTime*-hard for OWL 2 QL KBs.*

Proof. The proof is by encoding alternating Turing machines with polynomial tape and using the well-known fact that APSPACE = EXPTIME. For more details on alternating Turing machines the reader is referred to [7].

Suppose we are given an alternating Turing machine $M = (\Gamma, Q, q_0, q_1, \delta)$, where Γ is a tape alphabet, Q a set of states partitioned into existential $Q_\exists$ and universal $Q_\forall$ states, $q_0 \in Q_\exists$ an initial state, $q_1 \in Q$ an accepting state, and

$$\delta\colon (Q \setminus \{q_1\}) \times \Gamma \to (Q \times \Gamma \times \{-1, 0, +1\})^2$$

is a transition function that, for any $q \in Q$ and $a \in \Gamma$, gives two alternative transitions. We assume that existential and universal states strictly alternate, that is, any transition from an existential state results in a universal state, and vice versa. We also assume that M terminates on every input and there is a polynomial function f such that M uses at most $f(m)$ tape cells on any input of length m. It will be convenient to extend the transition function δ by the instructions $\delta(q_1, a) = ((q_1, a, 0), (q_1, a, 0))$, for all $a \in \Gamma$, which have an effect of going into an infinite loop if M reaches the accepting state. Thus, M accepts an input $\boldsymbol{w}$ iff there is a run of the modified machine M' on $\boldsymbol{w}$ such that all branches of the run are infinite.

Our aim is to construct, given M and an input $\boldsymbol{w}$, two TBoxes, $\mathcal{T}_1$ and $\mathcal{T}_2$, such that M accepts $\boldsymbol{w}$ iff the canonical model $\mathcal{C}_2$ of $\mathcal{K}_2 = (\mathcal{T}_2, \{A(c)\})$ is finitely Σ-homomorphically embeddable into the canonical model $\mathcal{C}_1$ of $\mathcal{K}_1 = (\mathcal{T}_1, \{A(c)\})$, where Σ comprises the concept and role names in $\mathcal{T}_2$ (in fact, Σ can be any set of concept and role names including those from $\mathcal{T}_2$).

In the definition of the TBox $\mathcal{T}_1$ (but not $\mathcal{T}_2$), we use inclusions of the form $B \sqsubseteq \exists R.(C_1 \sqcap \dots \sqcap C_k)$ as an abbreviation for $B \sqsubseteq \exists R_0$, $R_0 \sqsubseteq R$ and $\exists R_0^- \sqsubseteq C_i$, for $1 \leq i \leq k$, where R_0 is a fresh role name. If C_i is a complex concept then $\exists R_0^- \sqsubseteq C_i$ is also treated as an abbreviation for the respective concept and role inclusions.

We define $\mathcal{T}_1$ and $\mathcal{T}_2$ in three steps. Let $n = f(|\boldsymbol{w}|)$.

Step 1. First we encode configurations and transitions of M' using $\mathcal{T}_1$. We represent a configuration (that is, the contents of every cell on the tape, the state and the position of the head) by a sequence of $(n+1)$ R-arrows, for some roles R, which will be called a *block*. More precisely, the first arrow in each block is auxiliary; it is used to distinguish the type of the block. The range of the remaining n arrows belongs to one of the concepts C_a, for $a \in \Gamma$. For example, if the range of the $(i+1)$th arrow in the block belongs to C_a then the ith cell of the tape contains a in the configuration defined by the block.

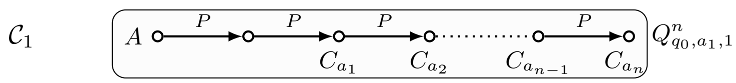

Fig. 1. Encoding a configuration by a P-block.

The initial block of $(n+1)$-many P-arrows represents the initial configuration, that is, symbols $a_1, \ldots, a_n$ written in the n cells of the tape (comprising the input $\boldsymbol{w}$ in the first m cells padded with the blanks) and the initial state q_0 (see Fig. 1):

$$A \sqsubseteq \exists P. \exists P. (C_{a_1} \sqcap \exists P. (C_{a_2} \sqcap \exists P. (\ldots \exists P. (C_{a_n} \sqcap Q^n_{q_0, a_1, 1}) \ldots))). \qquad (\mathcal{T}_1\text{-}1)$$

The concepts $Q^j_{q,a,k}$, for $0 \leq j \leq n$, define the remaining components of a configuration: $q \in Q$ is the current state and k, $1 \leq k \leq n$, is the current position of the head on the tape. In addition, $a \in \Gamma$ specifies the contents of the cell scanned by the head (this duplicates the information given by C_a but simplifies the construction). Finally, the index j is required to propagate the current state and head position along the points encoding the configuration. At the end of the tape, the concept $Q^n_{q,a,k}$ starts a separate branch for each of the two transitions: we take, for $q \in Q$ and $a \in \Gamma$ with $\delta'(q,a) = ((q_1, a_1, d_1), (q_2, a_2, d_2))$, where $d_1, d_2 \in \{+1, 0, -1\}$, the inclusions

$$Q^n_{q,a,k} \sqsubseteq \exists P. (X_1 \sqcap B^0_{q_1, a_1, k, d_1}) \sqcap \exists P. (X_2 \sqcap B^0_{q_2, a_2, k, d_2}), \quad \text{for } 1 \leq k \leq n, \quad (\mathcal{T}_1\text{-}2)$$

where X_1 and X_2 are two fresh concept names (distinguishing the two branches) and the $B^j_{q',a',k,d}$ are fresh concepts encoding the next state q' and the symbol a' to be placed to the kth cell in the next configuration with k being the current position of the head and d indicating the direction in which the head moves.

For all cells before the head, that is, for all j with $j+1 < k+d$ and $j+1 < k$, we use the following inclusions to pass the information along the sequences of arrows encoding the configuration:

$$B^j_{q,a,k,d} \sqsubseteq \exists P. (C_b \sqcap B^{j+1}_{q,a,k,d}), \qquad \text{for } b \in \Gamma \qquad (\mathcal{T}_1\text{-}3)$$

(these inclusions, in fact, generate, for each $b \in \Gamma$, a branch in $\mathcal{C}_1$ to represent the same cell but with a different symbol, b, tentatively assigned to the cell—we shall see later how the correct branch and the symbol are selected to match the cell contents in the preceding configuration). If $j+1 = k$ or $j+1 = k+d$ then the cell contents is changed according to the subscript of $B^j_{q,a,k,d}$ and the contents of the cell scanned by the head is then recorded in the subscript of $Q^j_{q,a,k}$:

$$B^{k-2}_{q,a,k,-1} \sqsubseteq \exists P. (C_b \sqcap \exists P. (C_a \sqcap \exists S_{a,0} \sqcap Q^k_{q,b,k-1})), \qquad \text{for } b \in \Gamma, \qquad (\mathcal{T}_1\text{-}4)$$

$$B^{k-1}_{q,a,k,0} \sqsubseteq \exists P. (C_a \sqcap \exists S_{a,0} \sqcap Q^k_{q,a,k}), \qquad (\mathcal{T}_1\text{-}5)$$

$$B^{k-1}_{q,a,k,+1} \sqsubseteq \exists P. (C_a \sqcap \exists S_{a,0} \sqcap \exists P. (C_b \sqcap Q^{k+1}_{q,b,k+1})), \qquad \text{for } b \in \Gamma. \qquad (\mathcal{T}_1\text{-}6)$$

The three situations are depicted in Fig. 2, where the $S_{a,0}$ arrows are not shown, and all arrows are labelled by P. The filled nodes represent the beginning of the pattern in each of the three inclusions $(\mathcal{T}_1\text{-}4)$–$(\mathcal{T}_1\text{-}6)$ and the subscript of the concept C_a in the label of the hatched nodes determines the respective subscript of the sequence of roles $Q^j_{q,a,k}$ that follows the pattern.

Note that there is only one branch for the cell whose contents is modified: the branch corresponds to the new symbol, a, in that cell (see explanations below).

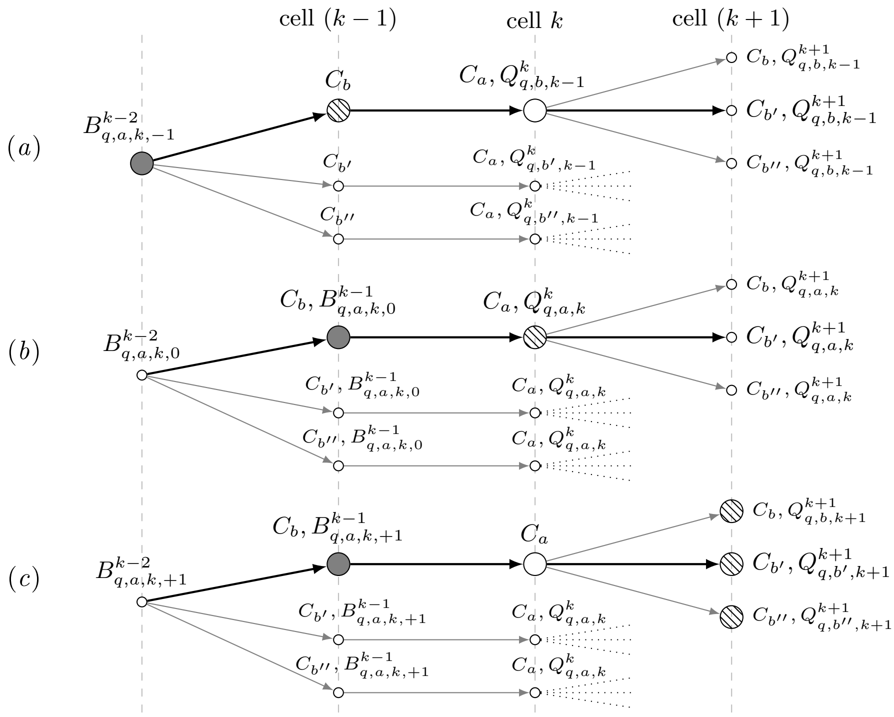

Fig. 2. Three types of transitions: head (a) moves left, (b) stays and (c) moves right.

Finally, if $k \le j < n$ then the current state and head position with the symbol scanned by the head are simply propagated along the tape:

$$Q^j_{q,a,k} \sqsubseteq \exists P.(C_b \sqcap Q^{j+1}_{q,a,k}), \qquad \text{for } b \in \Gamma \qquad (\mathcal{T}_1\text{-}7)$$

(generating a separate branch for each symbol tentatively assigned to the cell; see cases (a) and (b) in Fig. 2).

Step 2. The axioms ($\mathcal{T}_1$-3)–($\mathcal{T}_1$-7) generate a separate P-successor for each symbol b in the alphabet Γ. The correct one will be chosen by a Σ-homomorphism from the canonical model $\mathcal{C}_2$ of $\mathcal{K}_2$ into the canonical model $\mathcal{C}_1$ of $\mathcal{K}_1$. To exclude wrong choices, we add the following concept and role inclusions to $\mathcal{T}_1$ and $\mathcal{T}_2$. Each element d_1 of $\mathcal{C}_1$ that represents a possible cell contents, say a, also generates a block of arrows $S_{b,0}, \dots, S_{b,n}$, for all symbols b different from a:

$$C_a \sqsubseteq D, \qquad (\mathcal{T}_1\text{-}8)$$

$$C_a \sqsubseteq \exists S_{b,0}, \qquad \text{for } b \in \Gamma \setminus \{a\}, \qquad (\mathcal{T}_1\text{-}9)$$

$$\exists S^-_{b,j-1} \sqsubseteq \exists S_{b,j}, \qquad \text{for } b \in \Gamma \text{ and } 0 < j \le n, \qquad (\mathcal{T}_1\text{-}10)$$

$$\exists S^-_{b,n} \sqsubseteq C_b, \qquad \text{for } b \in \Gamma. \qquad (\mathcal{T}_1\text{-}11)$$

On the other hand, $\mathcal{T}_2$ makes sure that every element d_2 of a concept D in $\mathcal{C}_2$ generates a block $Z_{a,0}, \dots, Z_{a,n}$, for each $a \in \Gamma$:

$$D \sqsubseteq \exists Z_{a,0}, \tag{$\mathcal{T}_2$-1}$$

$$\exists Z^-_{a,j-1} \sqsubseteq \exists Z_{a,j}, \qquad \text{for } 0 < j \leq n, \tag{$\mathcal{T}_2$-2}$$

$$\exists Z^-_{a,n} \sqsubseteq C_a. \tag{$\mathcal{T}_2$-3}$$

Suppose a Σ-homomorphism $h\colon \mathcal{C}_2 \to \mathcal{C}_1$ maps d_2 to d_1. The following role inclusions in $\mathcal{T}_1$ ensure then that all the $Z_{a,0}, \dots, Z_{a,n}$ blocks but one can be mapped by h to the respective blocks $S_{a,0}, \dots, S_{a,n}$ in $\mathcal{C}_1$ (see Fig. 3):

$$S_{a,j} \sqsubseteq Z_{a,j}, \qquad \text{for } a \in \Gamma \text{ and } 0 \leq j \leq n. \tag{$\mathcal{T}_1$-12}$$

So, the only remaining block is $Z_{a,0}, \dots, Z_{a,n}$ such that d_1 belongs to C_a in $\mathcal{C}_1$ (that is, the symbol a tentatively contained in this cell). Then the following role inclusions allow the sequence of $Z_{a,j}$ to be mapped along the inverses of P in $\mathcal{C}_1$, *but only* if the same cell contains a in the preceding configuration (that is, the element $n+1$ steps closer to the root in $\mathcal{C}_1$ belongs to C_a; see Fig. 3):

$$P^- \sqsubseteq Z_{a,j}, \qquad \text{for } a \in \Gamma \text{ and } 0 \leq j \leq n. \tag{$\mathcal{T}_1$-13}$$

Thus, the blocks $S_{b,0}, \dots, S_{b,n}$ act as 'sinks' for the respective blocks $Z_{b,0}, \dots, Z_{b,n}$ from $\mathcal{C}_2$.

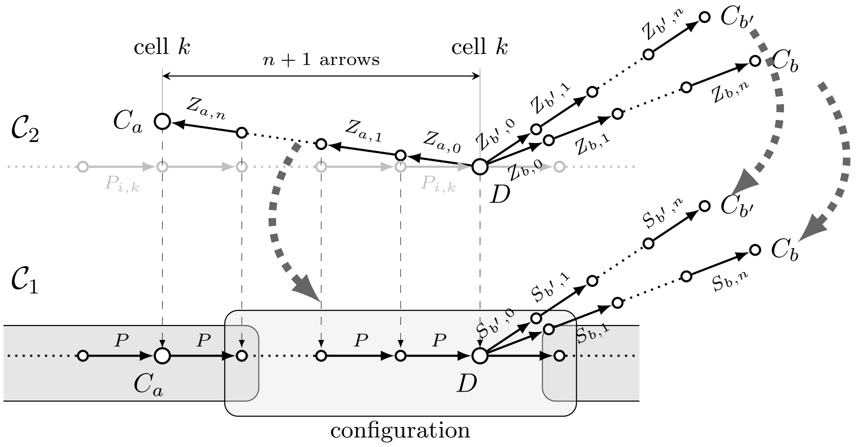

Fig. 3. Synchronising the contents of the cells in consecutive configurations.

Note that the cell whose contents is changed generates the additional blocks $S_{a,0}, \dots, S_{a,n}$ to allow the respective blocks of $Z_{a,0}, \dots, Z_{a,n}$ from $\mathcal{C}_2$ to be sunk there (because there can be no match along the P-edges).

Step 3. It remains to encode the acceptance condition for the Turing machine M'. To this end, we extend $\mathcal{T}_2$ with four blocks of roles: for each $i = 0, \dots, 3$, we take role names $P_{i,0}, \dots, P_{i,n}$ and add to $\mathcal{T}_2$ the inclusions

$$\exists P_{i,j-1}^{-} \sqsubseteq \exists P_{i,j}, \text{ for } 0 \leq i \leq 3 \text{ and } 0 < j \leq n. \qquad (\mathcal{T}_2\text{-4})$$

The four blocks are arranged into an infinite tree-like structure by the following concept inclusions:

$$A \sqsubseteq \exists P_{0,0}, \qquad \exists P_{0,n}^{-} \sqsubseteq \exists P_{1,0}, \qquad \exists P_{0,n}^{-} \sqsubseteq \exists P_{2,0}, \qquad (\mathcal{T}_2\text{-5})$$

$$\exists P_{1,n}^{-} \sqsubseteq \exists P_{3,0}, \qquad \exists P_{2,n}^{-} \sqsubseteq \exists P_{3,0}, \qquad (\mathcal{T}_2\text{-6})$$

$$\exists P_{3,n}^{-} \sqsubseteq \exists P_{1,0}, \qquad \exists P_{3,n}^{-} \sqsubseteq \exists P_{2,0}. \qquad (\mathcal{T}_2\text{-7})$$

More precisely, the $P_{0,j}$ block can be thought of as the root, from which two branches start with $P_{1,j}$ and $P_{2,j}$ blocks, respectively. Each of these blocks is followed by a $P_{3,j}$ block. Then we again have the $P_{1,j}$ and $P_{2,j}$ blocks, and so on. Such a pattern is required to represent configurations of M' with alternating universal and existential states.

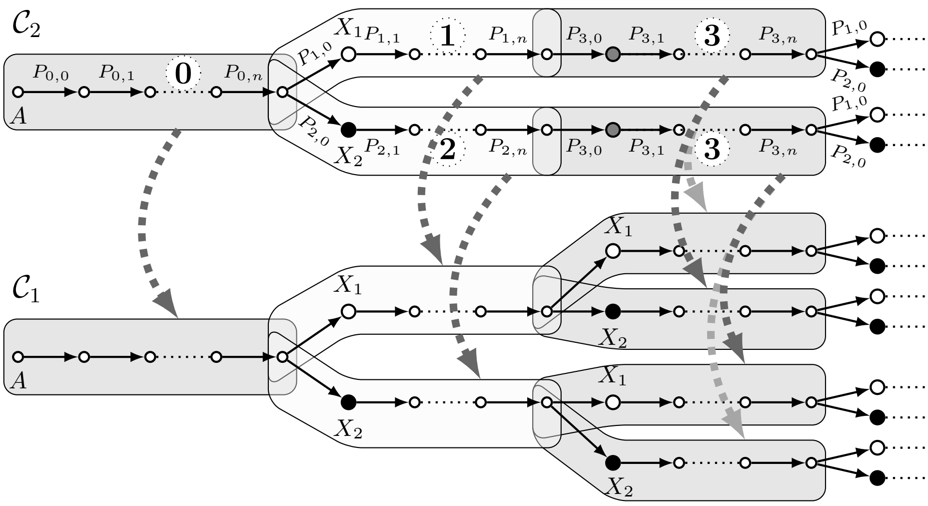

Fig. 4. Choosing a run all branches of which are infinite.

We extend $\mathcal{T}_1$ with the following role inclusions:

$$P \sqsubseteq P_{i,j}, \qquad \text{for } 0 \leq i \leq 3 \text{ and } 0 \leq j \leq n. \qquad (\mathcal{T}_1\text{-14})$$

Thus, each the initial $P_{0,i}$ block in $\mathcal{C}_2$ can be Σ-homomorphically mapped into the initial P-block in $\mathcal{C}_1$ (see Fig. 4). Its successor must be a universal state, which is reflected by the two blocks, $P_{1,j}$ and $P_{2,j}$, in $\mathcal{C}_2$. Both of these blocks should be Σ-homomorphically mapped into $\mathcal{C}_1$. The following concept inclusions in $\mathcal{T}_2$ ensure that the $P_{1,j}$ block is mapped to the P-block that begins with X_1 and the $P_{2,j}$ to the P-block that begins with X_2:

$$\exists P_{1,0}^- \sqsubseteq X_1, \qquad\qquad \exists P_{2,0}^- \sqsubseteq X_2, \qquad\qquad (\mathcal{T}_2\text{-8})$$

Each of the $P_{1,j}$ and $P_{2,j}$ blocks in $\mathcal{C}_2$ is followed by a $P_{3,j}$ block, which corresponds to an existential state: the $P_{3,j}$ block can be mapped to either of the two P-blocks (beginning with X_1 or X_2); see Fig. 4, where possible Σ-homomorphisms are shown by thick grey dashed arrows.

It remains to add to $\mathcal{T}_2$ concept inclusions making sure that each of the domain elements representing cells of the tape in a non-initial configuration belongs to the concept D, thus enforcing synchronisation of the cell contents:

$$\exists P_{i,j}^- \sqsubseteq D, \qquad\qquad \text{for } 1 \leq i \leq 3 \text{ and } 0 < j \leq n. \qquad\qquad (\mathcal{T}_2\text{-9})$$

One can show now that $\mathcal{T}_1$ and $\mathcal{T}_2$ are as required: M accepts $\boldsymbol{w}$ iff the canonical model $\mathcal{C}_2$ of $\mathcal{K}_2 = (\mathcal{T}_2, \{A(c)\})$ is finitely Σ-homomorphically embeddable into the canonical model $\mathcal{C}_1$ of $\mathcal{K}_1 = (\mathcal{T}_1, \{A(c)\})$, where Σ contains the concept and role names in $\mathcal{T}_2$. It remains to use Theorem 2 and the fact that APSpace = ExpTime.

Next, we show that Σ-query inseparability is also ExpTime-hard. To this end, let Σ be the signature of $\mathcal{T}_2$. Then we take $\mathcal{T}_1^1$ and $\mathcal{T}_2^2$ be to the 'localised' TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively: namely, $\mathcal{T}_k^k$ is the result of attaching the superscript k to each symbol (concept or role name) in the TBox $\mathcal{T}_k$ (such symbols are local for the respective TBox, that is, the signatures of $\mathcal{T}_1^1$ and $\mathcal{T}_2^2$ do not intersect). Then we take $\mathcal{T}_k^\Sigma$ to link those local symbols to the signature Σ: each $\mathcal{T}_k^\Sigma$ consists of all inclusions of the form $S^k \sqsubseteq S$, for symbols S in Σ. Now, let $\mathcal{T}_k' = \mathcal{T}_k^k \cup \mathcal{T}_k^\Sigma$. Consider the problem of Σ-query inseparability of $(\mathcal{T}_1', \{A^1(a)\})$ and $(\mathcal{T}_1' \cup \mathcal{T}_2', \{A^1(a), A^2(a)\})$. Evidently, the canonical model of $(\mathcal{T}_1', \{A^1(a)\})$ is contained in the canonical model of $(\mathcal{T}_1' \cup \mathcal{T}_2', \{A^1(a), A^2(a)\})$. The converse Σ-query entailment is only possible if there is a Σ-homomorphism from the canonical model of $(\mathcal{T}_2', \{A^2(a)\})$ into the canonical model of $(\mathcal{T}_1', \{A^1(a)\})$, which amounts to checking whether $\mathcal{T}_2$ is Σ-query entailed by $\mathcal{T}_1$. ❑

The ExpTime lower bound for Σ-query entailment in the case of OWL 2 QL TBoxes can be proved, using Theorem 3, by a minor modification of $\mathcal{T}_1$ and $\mathcal{T}_2$ above: for each of the role names R involved in $\mathcal{T}_2$ (and so in $\mathcal{T}_1$), we take a fresh role R' and add an extra role inclusion $R \sqsubseteq R'$ to the each of the two TBoxes; and take Σ to include the concepts A, X_1, X_2 and the C_a, for $a \in \Gamma$, together with the newly introduced role names R'. Thus we obtain:

Theorem 5. *The Σ-query entailment and Σ-query inseparability problems for OWL 2 QL TBoxes are* ExpTime*-complete.*

4 Conclusion

In this paper, we have proved that the Σ-query entailment and inseparability problems for both OWL 2 QL KBs and TBoxes are ExpTime-hard. This lower bound is tight for OWL 2 QL TBoxes matching the upper bound previously established in [4]. We are working on a game-theoretic proof showing that it is tight for the case of KBs as well.

References

1. Arenas, M., Botoeva, E., Calvanese, D., Ryzhikov, V.: Exchanging OWL 2 QL knowledge bases. In: Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI 2013) (2013)
2. Arenas, M., Botoeva, E., Calvanese, D., Ryzhikov, V., Sherkhonov, E.: Exchanging description logic knowledge bases. In: Proc. of the 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2012). AAAI Press (2012)
3. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press (2003), (2nd edition, 2007)
4. Konev, B., Kontchakov, R., Ludwig, M., Schneider, T., Wolter, F., Zakharyaschev, M.: Conjunctive query inseparability of OWL 2 QL TBoxes. In: Proc. of the 25th AAAI Conf. on Artificial Intelligence (AAAI 2011). AAAI Press (2011)
5. Konev, B., Ludwig, M., Walther, D., Wolter, F.: The logical difference for the lightweight description logic EL. Journal of Artificial Intelligence Reserach (JAIR) 44, 633–708 (2012)
6. Kontchakov, R., Wolter, F., Zakharyaschev, M.: Logic-based ontology comparison and module extraction, with an application to DL-Lite. Artificial Intelligence 174, 1093–1141 (2010)
7. Kozen, D.: Theory of Computation. Springer (2006)
8. Lutz, C., Wolter, F.: Deciding inseparability and conservative extensions in the description logic EL. Journal of Symbolic Computation 45(2), 194–228 (2010)
9. Vardi, M.Y.: Reasoning about the past with two-way automata. In: Proc. of the 25th Int. Colloquium on Automata, Languages and Programming (ICALP'98). Lecture Notes in Computer Science, vol. 1443, pp. 628–641. Springer (1998)