Amalgamation is PSPACE-hard *

Manuel Bodirsky¹, Simon Knäuer², and Jakub Rydval³

¹ Institute of Algebra, TU Dresden, Germany
manuel.bodirsky@tu-dresden.de
² Institute of Algebra, TU Dresden, Germany
simon.knauer@tu-dresden.de
³ Institute of Theoretical Computer Science, TU Dresden, Germany
jakub.rydval@tu-dresden.de

Abstract. The finite models of a universal sentence \( \Phi \) in a finite relational signature are the age of a homogeneous structure if and only if \( \Phi \) has the amalgamation property. We prove that the computational problem whether a given universal sentence \( \Phi \) has the amalgamation property is PSPACE-hard, even if \( \Phi \) is additionally Horn and the signature of \( \Phi \) only contains relation symbols of arity at most three. The decidability of the problem remains open.

Keywords: amalgamation property · universal · Horn

1 Introduction

A relational structure is called homogeneous if every isomorphism between finite substructures can be extended to an automorphism. In model theory, homogeneous structures with a finite relational signature are an important source of structures with a large automorphism group and quantifier elimination. Famous examples are the ordered rational numbers \((\mathbb{Q};<)\), the Random graph, or the countable homogeneous poset that embeds all finite posets.

Fraïssé proved that the age of a homogeneous structure \( \mathcal{B} \), i.e., the class of all finite structures that embed into \( \mathcal{B} \), forms an amalgamation class. Conversely, for every amalgamation class \( \mathcal{C} \) there exists a countably structure \( \mathcal{B} \) whose age equals \( \mathcal{C} \), and this structure \( \mathcal{B} \) is unique up to isomorphism. Hence, a homogeneous structure is uniquely given by its age.

All countable homogeneous posets have been classified by [Sch79], all countable homogeneous tournaments by Lachlan [Lac84], all countable undirected graphs by Lachlan and Woodrow [LW80], and all countable homogeneous digraphs by Cherlin [Che98]. There are only finitely many homogeneous tournaments and posets, and only countably many homogeneous graphs. It is an interesting question whether countable homogeneous structures can be classified in general. However, it is not entirely clear what one could mean by ‘classify’ in

* Supported by DFG GRK 1763 (QuantLA). Manuel Bodirsky has received funding from the European Research Council through the ERC Consolidator Grant 681988 (CSP-Infinity).
this context. There are uncountably many countable homogeneous digraphs, but there are only countably many that do not have the free amalgamation property, which is the simplest form of amalgamation. So Cherlin’s result about the homogeneous directed graphs can still be considered a classification of amalgamation classes despite the fact that there are uncountably many.

For many of the well-known examples of homogeneous structures \( \mathcal{B} \) the age of \( \mathcal{B} \) can be described by finitely many forbidden substructures. Let \( \mathcal{F} \) be a finite set of finite structures with finite relational signature \( \tau \) and let \( \mathcal{C} \) be the class of finite \( \tau \)-structures containing no isomorphic copy of a structure in \( \mathcal{F} \). For the special case where all relation symbols in \( \tau \) have arity at most two, Knight and Lachlan [KL87] (page 227) proved the following claim.

"Claim: (...) we can effectively check whether \( \mathcal{C} \) is an amalgamation class."

They proceed

"We do not know whether the claim is true for languages of arbitrary arity."

This problem is still open, and still relevant and of interest; it can also be found in the recent monograph of Cherlin [Che20] (Problem 6). While we cannot answer the question, we prove that the problem is \( \text{PSPACE} \)-hard. This is in sharp contrast to the binary case, where the same problem can be decided in polynomial time. Our proof is by a reduction from the universality problem for regular expressions, which is known to be \( \text{PSPACE} \)-complete. Our method of producing amalgamation classes of finite relational structures from universal regular grammars is an interesting source of examples for many open problems about homogeneous structures.

2 Preliminaries

The set \( \{1, \ldots, n\} \) is denoted by \( [n] \). We use the bar notation for tuples; for a tuple \( \bar{a} \) indexed by a set \( I \), the value of \( \bar{a} \) at the position \( i \in I \) is denoted by \( \bar{a}[i] \).

For a function \( f: A \to B \) and a tuple \( \bar{a} \in A^k \), we set \( f(\bar{a}) := (f(\bar{a}[1]), \ldots, f(\bar{a}[k])) \).

We always use capital Fraktur letters \( \mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \) etc to denote structures, and the corresponding capital Roman letters \( A, B, C, \) etc for their domains.

Let \( \tau \) be a finite relational signature. An atomic \( \tau \)-formula, or \( \tau \)-atom for short, is a formula of the form \( R(x_1, \ldots, x_n) \) for \( R \in \tau \) and (not necessarily distinct) variables \( x_1, \ldots, x_n \). Note that in this text equality atoms, i.e., expressions of the form \( x = y \), are not permitted. A class \( \mathcal{C} \) of relational \( \tau \)-structures has the amalgamation property (AP) if: whenever there exist embeddings \( e_i: \mathfrak{A} \hookrightarrow \mathfrak{B}_i \) (\( i \in [2] \)) for some \( \mathfrak{A}, \mathfrak{B}_1, \mathfrak{B}_2 \in \mathcal{C} \), then there exists \( \mathfrak{C} \in \mathcal{C} \) and embeddings \( f_i: \mathfrak{B}_i \hookrightarrow \mathfrak{C} \) (\( i \in [2] \)) such that \( f_1 \circ e_1 = f_2 \circ e_2 \). It has the strong amalgamation property if \( \mathfrak{C} \in \mathcal{C} \) and \( f_i: \mathfrak{B}_i \hookrightarrow \mathfrak{C} \) can be chosen so that \( f_1(B_1) \cap f_2(B_2) = f_1(e_1(A)) = f_2(e_2(A)) \). It has the free amalgamation property if additionally \( \mathfrak{C} \in \mathcal{C} \) can be chosen so that \( R^\mathfrak{C} = f_1(R^{\mathfrak{B}_1}) \cup f_2(R^{\mathfrak{B}_2}) \) for every \( R \in \tau \). A class \( \mathcal{C} \) of finite relational \( \tau \)-structures is called an (strong, free)
amalgamation class if it is closed under isomorphism, substructures, and has the (strong, free) AP.

If \( \mathfrak{A} \) is a \( \tau \)-structure and \( \Phi \) is a \( \tau \)-sentence, we write \( \mathfrak{A} \models \Phi \) if \( \mathfrak{A} \) satisfies \( \Phi \). If \( \Phi \) and \( \Psi \) are \( \tau \)-sentences, we write \( \Phi \models \Psi \) if every \( \tau \)-structure that satisfies \( \Phi \) also satisfies \( \Psi \). For universal sentences, this is equivalent to the statement that every finite \( \tau \)-structure that satisfies \( \Phi \) also satisfies \( \Psi \), by a standard application of the compactness theorem of first-order logic. If \( \Phi \) is a \( \tau \)-sentence, we denote the class of all finite models of \( \Phi \) by \( \llbracket \Phi \rrbracket_{\leq \omega} \). Clearly, every universal formula can be written as \( \forall x_1, \ldots, x_n, \psi \) where \( \psi \) is quantifier-free and in conjunctive normal form. Conjuncts of \( \psi \) are also called clauses; clauses are disjunctions of atomic formulas and negated atomic formulas. A clause \( \phi \) is a weakening of a clause \( \psi \) if every disjunct of \( \psi \) is also a disjunct of \( \phi \). We sometimes omit universal quantifiers in universal sentences (all first-order variables are then implicitly universally quantified). It is well-known that a class \( C \) of finite \( \tau \)-structures can be described by forbidding isomorphic copies of fixed finitely many finite \( \tau \)-structures if and only if \( C = \llbracket \Phi \rrbracket_{\leq \omega} \) for some universal \( \tau \)-sentence \( \Phi \). This allows us to reformulate the decision problem from the introduction as follows:

INPUT: A universal sentence \( \Phi \) in a finite relational signature.

QUESTION: Does \( \llbracket \Phi \rrbracket_{\leq \omega} \) have the AP?

A universal sentence \( \Phi \) is called Horn if its quantifier-free part is a conjunction of Horn clauses, i.e., disjunctions of positive or negative atomic formulas each having at most one positive disjunct. Every Horn clause can be written equivalently as an implication

\[
(\phi_1 \land \cdots \land \phi_n) \Rightarrow \phi_0
\]

where \( \phi_1, \ldots, \phi_n \) are atomic formulas (possibly \( n = 0 \)), and \( \phi_0 \) is either an atomic formula or \( \bot \) which stands for the empty disjunction. We refer to \( \phi_1 \land \cdots \land \phi_n \) as the premise, and to \( \phi_0 \) as the conclusion of the Horn clause. A Horn clause is called a tautology if the conclusion equals one of the conjuncts of the premise.

**Definition 1.** Let \( \Phi \) be a universal Horn sentence and \( \psi \) a Horn clause, both in a fixed relational signature \( \tau \). An SLD-derivation of \( \psi \) from \( \Phi \) of length \( k \) is a finite sequence of Horn clauses \( \psi^0, \ldots, \psi^k = \psi \) such that \( \psi^0 \) is a conjunct in \( \Phi \) and each \( \psi^i \) (\( 1 \leq i \leq k \)) is a (binary) resolvent of \( \psi^{i-1} \) and a conjunct \( \phi^i \) from \( \Phi \), i.e., for some atomic formula \( \psi^{i-1}_j \), we have

\[
\frac{\psi^{i-1}_1 \land \cdots \land \psi^{i-1}_j \land \cdots \land \psi^{i-1}_{n-1} \Rightarrow \psi^{i-1}_0}{\phi^i_1 \land \cdots \land \phi^i_{m^i} \Rightarrow \psi^{i-1}_j}
\]

There exists an SLD-deduction of \( \psi \) from \( \Phi \), written as \( \Phi \vdash \psi \), if \( \psi \) is a tautology or a weakening of a Horn clause that has an SLD-derivation from \( \Phi \) up to renaming variables.
The following theorem presents a fundamental property of universal Horn sentences.

**Theorem 1** (Theorem 7.10 in [NCdW97]). Let \( \Phi \) be a universal Horn sentence and \( \psi \) a Horn clause, both in a fixed signature \( \tau \). Then

\[
\Phi \models \psi \iff \Phi \vdash \psi.
\]

### 3 Universal Horn Sentences and the AP

This section contains some important observations about the AP in the context of universal Horn sentences which we later use in the proof of our main result. Let \( \tau \) be a relational signature. A Horn clause \( \phi \Rightarrow \psi \) is called complete if the graph with vertex set \( \{x_1, \ldots, x_n\} \) where \( x_i \) and \( x_j \) form an edge if they appear jointly in a conjunct of \( \phi \), forms a complete graph on all variables occurring in \( \phi \Rightarrow \psi \) in the usual graph-theoretic sense.

**Proposition 1.** Let \( \Phi \) be a conjunction of complete Horn clauses. Then \( \llbracket \Phi \rrbracket^{<\omega} \) has the AP.

**Proof.** The class \( \llbracket \Phi \rrbracket^{<\omega} \) even has the free amalgamation property. \( \square \)

A class \( C \) of relational \( \tau \)-structures has the one-point (strong) amalgamation property if it has the (strong) amalgamation property restricted to triples \((A, B_1, B_2)\) which satisfy \( |B_1| = |B_2| = |A| + 1 \).

**Definition 2 (Subsumption).** Let \( \Phi \) be a universal Horn sentence over the relational signature \( \tau \), and let \( \phi(\bar{x}) \) and \( \psi(\bar{x}, \bar{y}) \) be conjunctions of atomic \( \tau \)-formulas. We say that \( \phi(\bar{x}) \) subsumes \( \psi(\bar{x}, \bar{y}) \) with respect to \( \Phi \) and write \( \psi(\bar{x}, \bar{y}) \sqsubseteq_{\Phi} \phi(\bar{x}) \) if for every \( \chi \) that is \( \bot \) or an atomic \( \tau \)-formula with free variables among \( \bar{x} \)

\[
\Phi \models \forall \bar{x}, \bar{y} (\psi(\bar{x}, \bar{y}) \Rightarrow \chi(\bar{x})) \text{ implies } \Phi \models \forall \bar{x} (\phi(\bar{x}) \Rightarrow \chi(\bar{x})).
\]

Note that subsumption can be tested using SLD-deduction (Theorem [1]). In our proofs later it will be useful to take a proof-theoretic perspective on the amalgamation property.

**Lemma 1.** Let \( \Phi \) be a universal Horn sentence over the relational signature \( \tau \). Then the following are equivalent:

1. \( \llbracket \Phi \rrbracket^{<\omega} \) has the amalgamation property.
1'. \( \llbracket \Phi \rrbracket^{<\omega} \) has the one-point amalgamation property.
2. \( \llbracket \Phi \rrbracket^{<\omega} \) has the strong amalgamation property.
2'. \( \llbracket \Phi \rrbracket^{<\omega} \) has the one-point strong amalgamation property.
3. Suppose that \( \phi(\bar{x}), \phi_1(\bar{x}, y_1), \) and \( \phi_2(\bar{x}, y_2) \) are conjunctions of atomic formulas such that \( \phi(\bar{x}) \land \phi_i(\bar{x}, y_i) \sqsubseteq_{\Phi} \phi(\bar{x}) \) for \( i \in \{1, 2\} \). Then

\[
\phi(\bar{x}) \land \phi_1(\bar{x}, y_1) \land \phi_2(\bar{x}, y_2) \sqsubseteq_{\Phi} \phi(\bar{x}) \land \phi_1(\bar{x}, y_1).
\]
Proof. The equivalence (1) ⇔ (1) is well-known and holds in general; see, e.g., Proposition 2.3.17 in [Bod21]. The equivalence (2) ⇔ (2) is also clear as it follows the exact same principle.

(2) ⇒ (1): This direction is trivial.

(1) ⇒ (3): Let \( \phi(x), \phi_1(x, y_1), \) and \( \phi_2(x, y_2) \) be as in (1). Let \( \Psi(x), \Psi_1(x, y_1), \) and \( \Psi_2(x, y_2) \) be the conjunctions of all atomic formulas implied by \( \Phi \land \phi(x), \Phi \land \phi_1(x, y_1), \) and \( \Phi \land \phi_2(x, y_2), \) respectively. If \( \Phi \land \Psi(x), \Phi \land \Psi_1(x, y_1), \) or \( \Phi \land \Psi_2(x, y_2) \) is unsatisfiable, then \( \Phi \land \Psi(x) \) is unsatisfiable by the subsumption assumption and we are done. So suppose that all three conjunctions are satisfiable. Define \( A, B_1, \) and \( B_2 \) as the structures whose domains consist of the variables \( \{ x[1], \ldots, y_1[1], \ldots \}, \) and \( \{ y_2, x[1], \ldots \}, \) respectively, and where \( \bar{z} \) is a tuple of a relation for \( R \in \tau \) if the conjunct \( R(\bar{z}) \) is contained in \( \Psi, \Psi_1, \) or \( \Psi_2, \) respectively. Since \( \phi(x) \land \phi_1(x, y_1) \subseteq \phi(x) \) for \( i \in \{1, 2\}, \) there exist embeddings \( e_i : A \rightarrow B_i \) for \( i \in \{1, 2\}. \) Note that, by construction, \( A, B_1, \) and \( B_2 \) satisfy every Horn clause in \( \Phi. \) Since \( \Phi \) is universal Horn, this implies that \( A, B_1, B_2 \in [\Phi]^{<\omega}. \) Since \( [\Phi]^{<\omega} \) has the one-point amalgamation property, there exists \( C \in [\Phi]^{<\omega} \) together with embeddings \( f_i : B_i \rightarrow C \) for \( i \in \{1, 2\} \) such that \( f_1 \circ e_1 = f_2 \circ e_2. \) By the construction of \( A, B_1, \) and \( B_2, \) it follows that \( \Phi \not\vdash \forall \bar{x}, y_1, y_2 (\phi(x) \land \phi_1(x, y_1) \land \phi_2(x, y_2) \Rightarrow \bot). \) Let \( \chi(x, y_1) \) be an atomic \( \tau \)-formula such that \( \Phi \models \forall \bar{x}, y_1, y_2 (\phi(x) \land \phi_1(x, y_1) \land \phi_2(x, y_2) \Rightarrow \chi(x, y_1)). \) By the construction of \( A, B_1, \) and \( B_2, \) and because \( f_1 \) and \( f_2 \) are homomorphisms, there exist a tuple \( \bar{z} \) over \( B_1 \) such that \( C \models \chi(f_1(\bar{z})). \) Since \( f_1 \) is an embedding, we must also have \( B_1 \models \chi(\bar{z}). \) Thus, by the construction of \( A, B_1, \) and \( B_2, \) it follows that \( \Phi \models \forall \bar{x}, y_1 (\phi(x) \land \phi_1(x, y_1) \Rightarrow \chi(x, y_1)). \)

(3) ⇒ (2): Let \( A, B_1, B_2 \in [\Phi]^{<\omega} \) be such that \( e_i : A \rightarrow B_i \) and \( B_1 \setminus e_i(A) = \{ y_1 \} \) for \( i \in \{1, 2\}. \) We construct a structure \( C \in [\Phi]^{<\omega} \) with \( f_i : B_i \rightarrow C, \) and \( f_1 \circ e_1 = f_2 \circ e_2 \) as follows. Without loss of generality we may assume that \( A = e_1(A) = e_2(A), \) i.e., \( A \) is the intersection of \( B_1 \) and \( B_2. \) Let \( \bar{x} \) be a tuple of variables representing the elements of \( B_1 \cap B_2 \) in some order, and let \( \phi(x), \phi_1(x, y_1), \phi_1(x, y_2) \) be the conjunctions of all atomic \( \tau \)-formulas which hold in \( A, B_1, B_2, \) respectively. Note that \( \phi_1(x, y_1) \subseteq \phi(x) \) for both \( i \in \{2\}. \) Let \( \Psi(x, y_1, y_2) \) be the conjunction of all atomic formulas implied by \( \Phi \land \phi(x) \land \phi_1(x, y_1) \land \phi_2(x, y_2). \) We claim that \( \Phi \land \Psi \) is satisfiable: otherwise, \( \Phi \models \forall \bar{x}, y_1, y_2 (\phi(x) \land \phi_1(x, y_1) \land \phi_2(x, y_2) \Rightarrow \bot), \) and then (3) implies that \( \Phi \models \forall \bar{x}, y_1, y_2 (\phi(x) \land \phi_1(x, y_1) \Rightarrow \bot), \) which is impossible since \( B_1 \models \Phi. \) Define \( C \) as the structure with domain \( \{ y_1, y_2, x[1], \ldots \} \) and such a tuple \( \bar{z} \in C^{<\omega} \) if and only if \( \Psi \) contains the conjunct \( R(\bar{z}). \) For \( i \in \{2\}, \) let \( f_i \) be the identity map. We claim that \( f_1 \) is an embedding from \( B_1 \) to \( C. \) It is clear from the construction of \( C \) that \( f_1 \) is a homomorphism. Suppose for contradiction that there exists \( \bar{z} \in C^{<\omega} \) and a tuple \( \bar{z} \in B_1 \) such that \( \bar{z} \not\in R_\Psi \) while \( f_1(\bar{z}) \in R_\Psi. \) For the sake of notation, we assume that \( i = 1; \) the case that \( i = 2 \) can be shown analogously. Note that the construction of \( C \) implies that \( \Phi \models \forall \bar{x}, y_1, y_2 (\phi(x) \land \phi_1(x, y_1) \land \phi_2(x, y_2) \Rightarrow R(\bar{z})), \) Then (3) implies that \( \Phi \models \forall \bar{x}, y_1, y_2 (\phi(x) \land \phi_1(x, y_1) \Rightarrow R(\bar{z})), \) a contradiction to \( B_1 \in [\Phi]^{<\omega}. \) So, \( f_1 \) is an embedding from \( B_1 \) to \( C. \) By the construction of
C we also clearly have that \( f_1 \circ e_1 = f_2 \circ e_2 \), which concludes the proof of the one-point strong amalgamation property. \( \square \)

The equivalence between the AP and the strong AP is essentially due to the fact that equality atoms are not permitted in \( \Phi \). We mention in passing that (2), (2'), and (3) are equivalent even if equality is permitted in \( \Phi \).

## 4 PSPACE-hardness of the AP

In this section we prove that the problem of deciding whether the class of all finite models of a given universal Horn sentence \( \Phi \) has the amalgamation property is PSPACE-hard. Our proof is based on a reduction from the problem of deciding the universality of a given regular grammar.

A (left-) regular grammar is a 4-tuple \( G = (N, \Sigma, P, S) \) where

- \( N \) is a finite set of non-terminal symbols,
- \( \Sigma \) is a finite set of terminal symbols,
- \( P \) is a finite set of production rules that are of one of the following two forms:
  - \( A \rightarrow a \)
  - \( A \rightarrow Ba \)
  where \( A, B \in N \) and \( a \in \Sigma \), and
- \( S \in N \) is the start symbol.

For \( u \in (N \cup \Sigma)^* \) we write \( u \rightarrow_G v \) if there exist \( x \in \Sigma^* \) and \( (p \rightarrow q) \in P \) such that \( u = xp \) and \( v = xq \). The transitive closure of \( \rightarrow_G \) is denoted by \( \rightarrow_G^* \).

The language of \( G \) is \( L(G) := \{ w \in \Sigma^* \mid S \rightarrow_G^* w \} \). Note that with this definition the empty word, i.e., the word \( \epsilon \) of length 0, can never be an element of \( L(G) \); some authors use a modified definition that also allows rules that derive \( \epsilon \), but for our purposes the difference is not essential.

The idea of the reduction is to compute from a given regular grammar \( G \) a universal Horn sentence which consists of two parts, \( \Phi_1 \) and \( \Phi_2 \): the sentence \( \Phi_2 \) does not depend on \( G \) and entails many Horn clauses witnessing failure of the AP via Lemma 1; the sentence \( \Phi_1 \) can be computed efficiently from \( G \) and is such that \( \llbracket \Phi_1 \rrbracket_{\omega} \) has the free amalgamation property and prevents all the failures of the AP of \( \llbracket \Phi_1 \land \Phi_2 \rrbracket_{\omega} \) if and only if \( G \) is universal, i.e., \( L(G) = \Sigma^* \).

**Theorem 2.** For a given universal Horn sentence \( \Phi \) the question whether \( \llbracket \Phi \rrbracket_{\omega} \) has the AP is PSPACE-hard even if the signature is limited to at most ternary relation symbols.

**Proof.** The universality problem for regular expressions is known to be PSPACE-complete [AHU74] (Theorem 10.14, page 399). From every regular expression \( L \) over a finite alphabet \( \Sigma \), we can compute in polynomial time a left-regular grammar \( G = (N, \Sigma, P, S) \) such that \( L(G) = L \). Thus, deciding whether \( L(G) = \Sigma^* \) for a given left-regular grammar \( G \) is still PSPACE-hard. In this proof, we only consider left-regular grammars without production rules of the form \( A \rightarrow \)}
Sa. This is without loss of generality as such grammars are still able to generate all regular expressions. For \( k \geq 2 \), let

\[
C(x_1, \ldots, x_k) := \bigwedge_{i \in [k]} E(x_1, x_{i+1}).
\]

**Encoding regular grammars into amalgamation classes.** Let \( \tau_1 \) be the signature that contains the unary symbols \( I \) and \( T \), the binary symbol \( E \), and the binary relation symbol \( R_a \) for every element \( a \in N \cup \Sigma \). Let \( \Phi_1 \) be the universal Horn sentence that contains for every \( (A \to Ba) \in P \) with \( A \neq S \) the Horn clause

\[
I(y) \land C(y, x_i, x_{i+1}) \land R_B(y, x_i) \land R_a(x_i, x_{i+1}) \Rightarrow R_A(y, x_{i+1}),
\]

for every \( (S \to Ba) \in P \) the Horn clause

\[
I(y) \land T(x_n) \land C(y, x_{n-1}, x_n) \land R_B(y, x_{n-1}) \land R_a(x_{n-1}, x_n) \Rightarrow \bot,
\]

for every \( (A \to a) \in P \) with \( A \neq S \) the Horn clause

\[
I(y) \land C(y, x_1) \land R_a(y, x_1) \Rightarrow R_A(y, x_1),
\]

and for every \( (S \to a) \in P \) the Horn clause

\[
I(y) \land T(x_1) \land C(y, x_1) \land R_a(y, x_1) \Rightarrow \bot.
\]

Note that, due to the presence of \( C \), each conjunct of \( \Phi_1 \) is complete, which means that \( [\Phi_1]^{\leq \omega} \) has the AP by Proposition 1. The following correspondence can be shown via a straightforward induction.

**Claim 1.** For every \( w = a_1 \ldots a_n \in \Sigma^* \) and every \( A \in N \setminus \{S\} \)

\[
A \Rightarrow_G^* w \quad \text{if and only if} \quad \Phi_1 \models \forall x_1, \ldots, x_n, y \left( I(y) \land R_{a_1}(y, x_1) \land \bigwedge_{i \in [n-1]} (R_{a_i}(x_i, x_{i+1}) \land C(y, x_i, x_{i+1})) \Rightarrow R_A(y, x_n) \right).
\]

For \( A = S \), we have

\[
S \Rightarrow_G^* w \quad \text{if and only if} \quad \Phi_1 \models \forall x_1, \ldots, x_n, y \left( I(y) \land R_{a_1}(y, x_1) \land T(x_n) \land \bigwedge_{i \in [n-1]} (R_{a_i}(x_i, x_{i+1}) \land C(y, x_i, x_{i+1})) \Rightarrow \bot \right).
\]

**Proof.** “\( \Rightarrow \)” Suppose that \( A \Rightarrow_G^* a_1 \ldots a_n \) for \( A \in N \). Then there is a path in \( \Rightarrow_G \) from \( A \) to \( a_1 \ldots a_n \) of length \( \lambda \geq 1 \). We prove the statements by induction on \( \lambda \).

In the induction base \( \lambda = 1 \) we have \( (A \to a) \in P \) in which case \( 5 \) is a conjunct of \( \Phi_1 \).

In the induction step \( \lambda \rightarrow \lambda + 1 \), we assume that the claim holds for all paths of length \( \leq \lambda \), and that there exists a path of length \( \lambda + 1 \) from \( A \) to
\[ a_1 \ldots a_n, \text{i.e., there exists } (A \to B a_n) \in P \text{ and a path of length } \lambda \text{ from } B \text{ to } a_1 \ldots a_{n-1}. \] By the induction hypothesis, if \( A \neq S \) we have that

\[ \Phi_1 \models \forall x_1, \ldots, x_{n-1}, y \left( I(y) \land R_{a_1}(y, x_1) \right) \]

\[ \land \bigwedge_{i \in [n-2]} \left( R_{a_{i+1}}(x_i, x_{i+1}) \land C(y, x_i, x_{i+1}) \Rightarrow R_B(y, x_{n-1}) \right). \] \hspace{1cm} (7)

By the construction of \( \Phi_1 \), after renaming of variables we also have that

\[ \Phi_1 \models \forall x_{n-1}, x_n, y \left( I(y) \land R_B(y, x_{n-1}) \right) \]

\[ \land R_{a_n}(x_{n-1}, x_n) \land C(y, x_{n-1}, x_n) \Rightarrow R_A(y, x_n) \}. \] \hspace{1cm} (8)

We can now apply an SLD derivation step to (8) with (7) to obtain (5). The argument if \( A = S \) is analogous.

\[ \Leftarrow: \] Suppose that \( \Phi_1 \models \{5\} \). By Theorem 1, there is an SLD-deduction of (5) from \( \Phi_1 \). It cannot be the case that (5) is a tautology because each \( a_i \) is a terminal symbol and \( A \) is a non-terminal symbol. Thus, (5) is a weakening of a Horn clause \( \psi \) that has an SLD-derivation from \( \Phi_1 \), and we prove the claim by induction on the length \( \lambda \) of a shortest possible SLD-derivation for \( \psi \).

In the base case \( \lambda = 0 \), \( \psi \) must be a conjunct of \( \Phi_1 \). Since each \( a_i \) is a terminal symbol, \( \psi \) must be of the form (7). By the construction of \( \Phi_1 \), we get that \( (A \to a_1) \in P \) and thus \( A \to^{G*} a_1 \).

In the induction step \( \lambda \rightarrow \lambda + 1 \), we assume that the claim holds if \( \psi \) has an SLD-derivation of length \( \leq \lambda \). Suppose that \( \psi \) requires an SLD-derivation of length \( \lambda + 1 \). By the construction of \( \Phi_1 \), there must exist \( (A, Ba_n) \in P \) such that \( \Phi_1 \) contains a conjunct of the form (1) or (2) that is used in the last step in a shortest possible SLD-derivation of \( \psi \). Moreover, there exists an SLD-derivation of

\[ I(y) \land R_{a_1}(y, x_1) \land \bigwedge_{i \in [n-2]} \left( R_{a_{i+1}}(x_i, x_{i+1}) \land C(y, x_i, x_{i+1}) \Rightarrow R_B(y, x_{n-1}) \right) \] \hspace{1cm} (9)

from \( \Phi_1 \) of length \( \leq \lambda \). By the induction hypothesis, (9) is equivalent to \( B \to^{G*} a_1 \ldots a_{n-1} \). Therefore, \( A \to^{G*} a_1, \ldots, a_n. \)

Creating candidates for failure of the AP. Let \( \tau_2 \) be the signature which contains all symbols from \( \tau_1 \) except for the ones coming from \( N \) and additionally the binary symbol \( F \) and the ternary symbol \( Q \). For \( k \geq 2 \), let

\[ D(x_1, \ldots, x_k) := \bigwedge_{i \in [k-1]} F(x_1, x_{i+1}) \]

and

\[ P(x_1, \ldots, x_k) := D(y_1, x_1, \ldots, x_k) \land C(y_2, x_1, \ldots, x_k) \]
Amalgamation is PSPACE-hard

\[ \text{Fig. 1. A graphical depiction of } \mathcal{M}_5 \text{ for } n = 5. \]

The sentence \( \Phi_2 \) consists of the following Horn clauses for every \( a \in \Sigma \):

\[
I(y_2) \land R_a(y_2, x_1) \land P(x_1) \Rightarrow Q(y_1, y_2, x_1) \quad (10)
\]

\[
I(y_2) \land Q(y_1, y_2, x_i) \land R_a(x_i, x_{i+1}) \land P(x_i, x_{i+1}) \Rightarrow Q(y_1, y_2, x_{i+1}) \quad (11)
\]

\[
I(y_2) \land Q(y_1, y_2, x_n) \land T(x_n) \land P(x_n) \Rightarrow \bot \quad (12)
\]

The proof of the following claim is straightforward and left to the reader.

**Claim 2.** Let \( \phi(\bar{x}) \) be a conjunction of \((\tau_2 \setminus \{Q\})\)-atoms. Then

\( \Phi_2 \models \forall \bar{x} (\phi(\bar{x}) \Rightarrow \bot) \)

if any only if there exist \( n \in \mathbb{N} \) and \( a_1, \ldots, a_n \in \Sigma \) such that \( \phi \) has a subformula of the following form:

\[
I(y_2) \land R_{a_1}(y_2, x_1) \land P(x_1, \ldots, x_n) \land T(x_n) \land \bigwedge_{i \in [n-1]} R_{a_{i+1}}(x_i, x_{i+1}) \quad (13)
\]

Now we are ready for the main part of the proof. We set \( \Phi := \Phi_1 \land \Phi_2 \).

**Claim 3.** \([\Phi]^{<\omega}\) has the AP if and only if \( L(G) = \Sigma^* \).

**Proof.** \( \Rightarrow \): Suppose that \([\Phi]^{<\omega}\) has the AP. Let \( a_1 \ldots a_n \in \Sigma^* \). Consider the formulas \( \phi(x_1, \ldots, x_n) \), \( \phi_1(x_1, \ldots, x_n, y_1) \), and \( \phi_2(x_1, \ldots, x_n, y_2) \) given by

\[
\phi_1 := D(y_1, x_1, \ldots, x_n),
\]

\[
\phi_2 := C(y_2, x_1, \ldots, x_n) \land I(y_2) \land R_{a_1}(y_2, x_1),
\]

and \( \phi := T(x_n) \land \bigwedge_{i \in [n-1]} R_{a_{i+1}}(x_i, x_{i+1}) \).
By Claim 2, we have

$$
\Phi_2 \models \forall x_1, \ldots, x_n, y_1, y_2 (\phi \land \phi_1 \land \phi_2 \Rightarrow \bot).
$$

Note that every Horn clause in $\Phi_2$ has both $E$- and $F$-atoms in its premise. Among $\phi$, $\phi_1$, and $\phi_2$, $E$-atoms are only present in $\phi_2$ and $F$-atoms are only present in $\phi_1$. Also note that every Horn clause in $\Phi_1$ does not contain any $F$-atoms. Thus, whenever we have

$$
\Phi \vdash \forall x_1, \ldots, x_n, y_1 (\phi \land \phi_1 \Rightarrow \chi)
$$

for some $\chi$ that is either $\bot$ or an atomic formula with free variables among $x_1, \ldots, x_n$, then we also have

$$
\Phi \vdash \forall x_1, \ldots, x_n (\phi \Rightarrow \chi).
$$

Using Theorem 1, we conclude that

$$
\phi \land \phi_1 \trianglelefteq \phi.
$$

(15)

Finally, note that every Horn clause in $\Phi$ contains an $I$-atom in its premise. Moreover, if a variable appears in an $I$-atom in the premise of a Horn clause whose conclusion is an atom, then this variable also appears in the atom from the conclusion. Since $y_2$ is the only variable which appears in an $I$-atom in $\phi \land \phi_2$, there exists an atomic formula with free variables among $x_1, \ldots, x_n$ such that

$$
\Phi \vdash \forall x_1, \ldots, x_n, y_2 (\phi \land \phi_2 \Rightarrow \chi)
$$

if and only if

$$
\Phi \vdash \forall x_1, \ldots, x_n, y_2 (\phi \land \phi_2 \Rightarrow \bot).
$$

By Theorem 1 together with Claim 1

$$
\Phi \vdash \forall x_1, \ldots, x_n, y_2 (\phi \land \phi_2 \Rightarrow \bot) \iff \Phi \models \forall x_1, \ldots, x_n, y_2 (\phi \land \phi_2 \Rightarrow \bot) \iff a_1 \ldots a_n \in L(G).
$$

Since $\phi \land \Phi$ is clearly satisfiable, it follows that

$$
\phi \land \phi_2 \trianglelefteq \phi.
$$

(16)

if and only if $a_1 \ldots a_n \notin L(G)$. Since $[\Phi]^{\omega}$ has the AP, it follows from Lemma 1 together with (14) and (15) that (16) may not be satisfied. Thus, $a_1 \ldots a_n \in L(G)$ and we are done.

"$\Leftarrow$": We prove the contrapositive and assume that $[\Phi]^{\omega}$ does not have the AP. Then there exists a counterexample to Item (3) in Lemma 1, i.e., there exists a Horn clause $\psi$ of the following form

$$
\phi(\bar{x}) \land \phi_1(\bar{x}, y_1) \land \phi_2(\bar{x}, y_2) \Rightarrow \chi
$$
where $\chi$ is either $\bot$ or an atomic $\tau$-formula with free variables $\bar{x}, y_1$ such that

$$
\Phi \models \forall \bar{x}, y_1, y_2 (\phi \land \phi_1 \land \phi_2 \Rightarrow \chi) \tag{17}
$$

and for both $i \in \{1, 2\}$

$$
\phi(\bar{x}) \land \phi_i(\bar{x}, y_i) \subseteq_{\phi} \phi(\bar{x}). \tag{19}
$$

We choose $\psi$ minimal with respect to the number of its atomic subformulas.

Our proof strategy is as follows. First we show that $\psi$ encodes a single word $w \in \Sigma^*$ in the sense of Claim 2. Then we show that the word $w$ may not be contained in $L(G)$, because otherwise a part of the counterexample would encode $w$ in the sense of Claim 1 which would lead to a contradiction.

**Observation 1.** The formula $\Phi \land \phi(\bar{x}) \land \phi_i(\bar{x}, y_i)$ is satisfiable for both $i \in \{1, 2\}$.

**Proof (of Observation 1).** Suppose for contradiction that $\Phi \land \phi(\bar{x}) \land \phi_i(\bar{x}, y_i)$ is unsatisfiable for some $i \in \{1, 2\}$. By (19) we have that $\Phi \models \forall \bar{x}(\phi(\bar{x}) \Rightarrow \bot)$. But then it follows trivially that $\phi(\bar{x}) \land \phi_1(\bar{x}, y_1) \land \phi_2(\bar{x}, y_2) \subseteq \phi \phi(\bar{x}) \land \phi_1(\bar{x}, y_1)$, a contradiction to our assumptions.

For a conjunction $\theta$ of $\tau$-atoms, we denote by $\theta'$ the conjunction obtained by removing all $(\tau_2 \setminus \tau_1)$-atoms from $\psi$.

**Observation 2.** $\Phi_2 \models \psi$ and $\psi$ only contains symbols from $\tau_2$.

**Proof (of Observation 2).** By Theorem 1 and (17) we have $\Phi \models \psi$. Since $Q \notin \tau_1$ and $R_A \notin \tau_2$ for every $A \in N$, by the construction of $\Phi_1$ and $\Phi_2$ we have either $\Phi_1 \models \psi$ or $\Phi_2 \models \psi$. Suppose for contradiction that $\Phi_1 \models \psi$. Note that $\chi(\bar{x}, y_1)$ cannot be a subformula of $\phi(\bar{x}) \land \phi_1(\bar{x}, y_1)$, by (18). If $\chi(\bar{x}, y_1)$ is a subformula of $\phi_2(\bar{x}, y_2)$, then (19) implies that $\Phi \models \forall \bar{x}, y_1 (\phi(\bar{x}) \Rightarrow \chi(\bar{x}, y_1))$, in contradiction to (18). Hence, $\chi(\bar{x}, y_1)$ is not a subformula of $\phi(\bar{x}) \land \phi_1(\bar{x}, y_1) \land \phi_2(\bar{x}, y_2)$. Thus, it follows from $\Phi_1 \models \psi$ and the fact that $\Phi_1$ only contains $\tau_1$-atoms that

$$
\Phi_1 \models \forall \bar{x}, y_1, y_2 (\phi' \land \phi_1' \land \phi_2' \Rightarrow \chi). \tag{20}
$$

We claim that, for both $i \in \{1, 2\}$,

$$
\phi'(\bar{x}) \land \phi_i'(\bar{x}, y_i) \subseteq_{\phi_1} \phi'(\bar{x}). \tag{21}
$$

Let $\eta$ be either a $\tau_1$-atom or $\bot$ such that $\Phi_1 \models \forall \bar{x}, y_i (\phi' \land \phi_i' \Rightarrow \eta)$ for some $i \in \{1, 2\}$. The latter is not possible by Observation 1. By (19), we have $\Phi \models \forall \bar{x}(\phi \Rightarrow \eta)$. If $\eta$ is a subformula of $\phi$, then trivially $\Phi_1 \models \forall \bar{x}(\phi' \Rightarrow \eta)$ and the claim holds. So suppose that this is not the case. Since $\eta$ is a $\tau_1$-atom that is not a subformula of $\phi$ and $\Phi \land \phi$ is satisfiable by Observation 1, by Theorem 1, $\eta$ can only be an $R_A$-atom for some $A \in N$ because other symbols from $\tau_1$ do not appear in the conclusion of any Horn implication in $\Phi$. But then, since $R_A$-atoms do not appear in any Horn implication from $\Phi_2$, by Theorem 1 we must have
we assume that the conclusion of \( \Phi_1 \models \forall x (\phi \Rightarrow \eta) \). Consequently, \( \Phi_1 \models \forall x (\phi' \Rightarrow \eta) \) because \( \{Q, F\} \)-atoms do not appear in any Horn implication from \( \Phi_1 \), again due to Theorem 1. Since \( \eta \) was chosen arbitrarily, (21) follows.

By the construction of \( \Phi_1 \), (20) and Theorem 1 imply that either \( \chi \) is an \( R_A \)-atom for some \( A \in N \), or \( \chi \) is \( \perp \). In both cases, (21), (20), and (18) witness that \( [\Phi_1] \leq \psi \) does not have AP through an application of Lemma 1. But this is in contradiction to Proposition 1. Thus, \( \Phi_1 \vdash \psi \) does not hold, and \( \Phi_2 \vdash \psi \) holds instead. In this case, \( \phi \land \phi_2 \land \phi_3 \) can only contain symbols from \( \tau_2 \), since we could otherwise remove all \( (\tau_1 \setminus \tau_2) \)-atoms and get a contradiction to the minimality of \( \psi \). Similarly as above, by the construction of \( \Phi_2 \), Theorem 1 and \( \Phi_2 \vdash \psi \) imply that either \( \chi \) is a \( Q \)-atom, or \( \chi \) equals \( \perp \). Thus, \( \psi \) only contains \( \tau_2 \)-atoms. \( \Box \)

Observation 3. The formula \( \phi(\bar{x}) \land \phi_1(\bar{x}, y_1) \land \phi_2(\bar{x}, y_2) \) does not contain any \( Q \)-atom and \( \chi \) equals \( \perp \).

Proof (of Observation 3). Recall that, by Observation 2 and Theorem 1, \( \psi \) has an SLD-deduction from \( \Phi_2 \). By (17) and (18), \( \psi \) cannot be a tautology. By the minimality of \( \psi \), we may assume that there exists an SLD-derivation of \( \psi \) from \( \Phi_2 \). Consider any SLD-derivation \( \psi^0, \ldots, \psi^k \) of \( \psi \) from \( \Phi_2 \). Note that, by the construction of \( \Phi_2 \), for every \( i \in [k] \), if there exist variables \( z_1, z_2 \) in \( \psi^{i-1} \) such that

i. every \( Q \)-atom contains \( z_1 \) in its 1st and \( z_2 \) in its 2nd argument, respectively,
ii. every \( F \)-atom contains \( z_1 \) in its second argument, and
iii. every \( E \)-atom contains \( z_2 \) in its second argument,

then \( \psi^i \) also satisfies (ii), (iii), and (iii) for the same variables \( z_1, z_2 \). Since every possible choice of \( \psi^0 \) from \( \Phi_2 \) initially satisfies these three conditions, it follows via induction that (ii), (iii), and (iii) must hold for \( \psi = \psi^k \) for some \( z_1, z_2 \). Also note that (10) is the only Horn clause in \( \Phi_2 \) that is not complete, but the incompleteness is only due to one edge missing between two distinguished variables satisfying (ii) and (iii).

Suppose, for contradiction, that \( \{z_1, z_2\} \neq \{y_1, y_2\} \) holds for the pair \( z_1, z_2 \) satisfying (ii), (iii), and (iii) for \( \psi \), i.e., both \( z_1 \) and \( z_2 \) are among \( \bar{x}, y_i \) for some \( i \in \{1, 2\} \). Recall that, by Observation 1, \( \Phi_2 \land \phi(\bar{x}) \land \phi_i(\bar{x}, y_i) \) is satisfiable for both \( i \in \{1, 2\} \). By (19) together with (ii), (iii), and (iii), every \( Q \)-atom that can be derived from \( \Phi_2 \land \phi(\bar{x}) \land \phi_i(\bar{x}, y_i) \land \phi_2(\bar{x}, y_2) \) may only contain variables from \( \bar{x}, y_i \) for some \( i \in \{1, 2\} \) and can already be derived from \( \Phi_2 \land \phi(\bar{x}) \land \phi_i(\bar{x}, y_i) \) for the same \( i \in \{1, 2\} \). But then we get a contradiction to (18). Therefore, \( \{z_1, z_2\} = \{y_1, y_2\} \).

We claim that \( \psi^0 \) is of the form (12). Otherwise, \( \psi^0 \) is of the form (11) or (10), in which case the conclusion of \( \psi \) is a \( Q \)-atom. But then, since \( z_1, z_2 \) with \( \{z_1, z_2\} = \{y_1, y_2\} \) satisfy (ii) for \( \psi \), the conclusion of \( \psi \) would be an atom containing both variables \( y_1 \) and \( y_2 \). This leads to a contradiction to (17) where we assume that the conclusion of \( \psi \) may only contain variables from \( \bar{x}, y_1 \). The claim implies that \( \chi \) equals \( \perp \).
Since the premise of (14) does not contain any atomic subformula with both variables \( y_1 \) and \( y_2 \), the fact that \( z_1, z_2 \) with \( \{ z_1, z_2 \} = \{ y_1, y_2 \} \) satisfy (11) for \( \psi \) implies that \( \psi = \psi^k \) does not contain any \( Q \)-atoms at all.

As a consequence of Observation 2 and Observation 3 we have that \( \Phi_2 \models \psi \) where \( \psi \) is of the form \( \phi \land \phi_1 \land \phi_2 \Rightarrow \bot \) and \( \phi, \phi_1, \phi_2 \) are conjunctions of atomic \( \tau_2 \setminus \{ Q \} \)-formulas. Therefore, Claim 2 implies that there exist \( a_1, \ldots, a_n \in \Sigma \) such that \( \phi \land \phi_1 \land \phi_2 \) is of the form

\[
I(y_2) \land R_{a_1}(y_2, x_1) \land P(x_1, \ldots, x_n) \land T(x_n) \land \bigwedge_{i \in [n-1]} R_{a_{i+1}}(x_i, x_{i+1})
\]

(22)

where the variables need not all be distinct. More specifically,

- \( \phi(x) \) equals \( T(x_n) \land \bigwedge_{i \in [n-1]} R_{a_{i+1}}(x_i, x_{i+1}) \),
- \( \phi_1(x, y_1) \) equals \( D(y_1, x_1, \ldots, x_n) \), and
- \( \phi_2(x, y_2) \) equals \( C(y_2, x_1, \ldots, x_n) \land I(y_2) \land R_{a_1}(y_2, x_1) \).

Note that, if \( L(G) = \Sigma^* \), then Claim 1 implies that

\[
\Phi_1 \models \forall x_1, \ldots, x_{n+1} \left( \phi \land \phi_2 \Rightarrow \bot \right).
\]

(23)

If some variables among \( x_1, \ldots, x_n \) are identified in (22), then we still have (23) even if we perform the same identification of variables. But then we get a contradiction to Observation 1. Thus, \( L(G) \neq \Sigma^* \), which concludes the proof of Claim 3.

We have thus found a reduction from the PSPACE-hard universality problem for \( G \) to the decidability problem of the AP for \( \llbracket \Phi \rrbracket_{<\omega} \); note that \( \Phi \) is universal Horn and can be computed from \( G \) in polynomial time.

Recall from Lemma 1 that a universal Horn sentence \( \Phi \) has the strong AP if and only if it has the AP.

**Corollary 1.** For a given universal sentence \( \Phi \) the question whether \( \llbracket \Phi \rrbracket_{<\omega} \) has the strong AP is PSPACE-hard even if \( \Phi \) is Horn and the signature is limited to ternary relation symbols.

Our proof of PSPACE-hardness has another interesting consequence. Namely, it shows that in general there is no subexponential upper bound on the size of a smallest triple without an amalgam.

**Corollary 2.** There is a sequence \( (\Phi_k)_{k \geq 3} \) of universal Horn sentences with at most ternary relation symbols such that, for each \( k \geq 3 \), \( \llbracket \Phi_k \rrbracket_{<\omega} \) does not have the AP, but the size of a smallest triple without an amalgam is in \( \Omega(2^{\left| \Phi_k \right|}) \).

**Proof.** By Theorem 33 in [EKSW05], there exists a sequence \( (G_k)_{k \geq 3} \) of left-regular grammars \( G_k = (N_k, \Sigma, P_k, S) \) such that the size of a smallest word rejected by \( G_k \) is in \( \Omega(2^{\left| G_k \right|}) \). For every \( k \geq 3 \), let \( \Phi_{G_k} \) be the Horn sentence
constructed in the proof of Theorem 2 from $G_k$. By the construction of $\Phi_{G_k}$, there exists $c > 0$ such that $|\Phi_{G_k}| \leq c \cdot |G_k|$ for every $k \geq 3$.

We claim that the size of a smallest triple $A, B_1, B_2 \in [\Phi_k]^{<\omega}$ without an amalgam is greater than or equal to the size of the smallest word $w \in \Sigma^* \setminus L(G_k)$. By “$\Leftarrow$” of the proof of Claim 3 combined with Lemma 1, each smallest counterexample to AP for $[\Phi_k]^{<\omega}$ is represented by a formula of the form (22). Let $\phi \land \phi_1 \land \phi_2$ be such a formula, and let $w := a_1 \ldots a_n$ be the word encoded by $\phi \land \phi_2$. We must have that $w \notin L(G_k)$. Otherwise, by Claim 1, $\Phi_{G_k} \land \phi \land \phi_2$ is unsatisfiable, which contradicts Observation 1.

Suppose, for contradiction, that some variables among $\{y_2, x_1, \ldots, x_n\}$ coincide in $\phi \land \phi_1 \land \phi_2$. Then $\phi \land \phi_2$ already encodes a subword $v := a_{i_1} \ldots a_{i_m}$ of $w$ along a shortest possible $R_{a_i}$-path from an $y_2$ to $x_n$. Again, we must have $v \notin L(G_k)$. Otherwise, by Claim 1, $\Phi_{G_k} \land \phi \land \phi_2$ is unsatisfiable, which contradicts Observation 1. But now we can fix an arbitrary shortest possible $R_{a_i}$-path from $y_2$ to $x_n$ and remove all subformulas from $\phi \land \phi_1 \land \phi_2$ which do not contain any variable occurring along this path in order to obtain a strictly smaller counterexample than $\phi \land \phi_1 \land \phi_2$. This leads to a contradiction to the minimality of $\phi \land \phi_1 \land \phi_2$. Thus, no variables among $\{y_2, x_1, \ldots, x_n\}$ may coincide in $\phi \land \phi_1 \land \phi_2$.

Now suppose, for contradiction, that there exists $b_1 \ldots b_m \in \Sigma^* \setminus L(G_k)$ with $m < n$. Since all variables in $\phi \land \phi_1 \land \phi_2$ are distinct, “$\Rightarrow$” of the proof of Claim 3 applied to $b_1 \ldots b_m$ yields us a strictly smaller counterexample than $\phi \land \phi_1 \land \phi_2$. Again, this leads to a contradiction to the minimality of $\phi \land \phi_1 \land \phi_2$. We conclude that $w$ is a shortest word in $\Sigma^* \setminus L(G_k)$, which is what we had to show.

5 Open Problems

The question of Knight and Lachlan about the decidability of the AP for universal sentences remains open even if the given sentence is Horn. We prove the PSPACE-hardness of this problem (Theorem 2) even if all the symbols in the signature have arity at most three. The following questions remain open.

**Question 1:** What is the complexity of the AP for universal classes where the signature is limited to ternary symbols?

**Question 2:** What is the complexity of the AP for universal Horn classes if the signature is limited to ternary symbols?

References

AHU74. Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.

Bod21. M. Bodirsky. *Complexity of Infinite-Domain Constraint Satisfaction*. Lecture Notes in Logic. Cambridge University Press, 2021.

Che98. G. L. Cherlin. The classification of countable homogeneous directed graphs and countable homogeneous $n$-tournaments. *AMS Memoir*, 131(621), January 1998.
Amalgamation is PSPACE-hard

Che20. G. L. Cherlin. Homogeneous ordered graphs and metrically homogeneous graphs, 2020. Preprint.

EKSW05. K. Ellul, B. Krawetz, J. Shallit, and M.-W. Wang. Regular expressions: New results and open problems. *J. Autom. Lang. Comb.*, 10(4):407–437, 2005.

KL87. J. Knight and A.H. Lachlan. Shrinking, stretching and codes for homogeneous structures. In *Classification Theory (Chicago, IL, 1985)*, Lecture Notes in Math., 1292, pages 192–228. Springer, Berlin-New York, 1987.

Lac84. A. H. Lachlan. Countable homogeneous tournaments. *Transactions of the American Mathematical Society (TAMS)*, 284:431–461, 1984.

LW80. A. H. Lachlan and R. E. Woodrow. Countable ultrahomogeneous undirected graphs. *Transactions of the AMS*, 262(1):51–94, 1980.

NCdW97. S.-H. Nienhuys-Cheng and R. de Wolf. *Foundations of Inductive Logic Programming*, volume 1228 of *Lecture Notes in Computer Science*. Springer, 1997.

Sch79. J. H. Schmerl. Countable homogeneous partially ordered sets. *Algebra Universalis*, 9:317–321, 1979.