

Finite Model Theory of the Triguarded Fragment and Related Logics

Emanuel Kieroński
Institute of Computer Science
University of Wrocław
 Email: emanuel.kieronski@cs.uni.wroc.pl

Sebastian Rudolph
Computational Logic Group
Technische Universität Dresden
 Email: sebastian.rudolph@tu-dresden.de

Abstract—The Triguarded Fragment (TGF) is among the most expressive decidable fragments of first-order logic, subsuming both its two-variable and guarded fragments without equality. We show that the TGF has the finite model property (providing a tight doubly exponential bound on the model size) and hence finite satisfiability coincides with satisfiability known to be N2EXPTIME-complete. Using similar constructions, we also establish 2EXPTIME-completeness for finite satisfiability of the constant-free (tri-)guarded fragment with transitive guards.

1. Introduction

Ever since first-order logic (FOL) was found to have an undecidable satisfiability problem, researchers have attempted to identify expressive yet decidable fragments of FOL and pinpoint their complexity. Two of the most prominent fragments in this regard are FO^2 (the *two-variable fragment*) and GF (the *guarded fragment*).

For FO^2 , decidability is retained through reducing the number of available variables to 2, essentially restricting expressivity to independent pairwise interactions between domain elements. Decidability of FO^2 without equality was already established in the 1960s [15]; in the 1970s the result was extended to the case with equality [13]. NEXPTIME-completeness was established in the 1990s [8].

GF, which owes its decidability to the restricted “guarded” use of quantifiers, originated in the late 1990s [1]. Its satisfiability problem is 2EXPTIME-complete but drops to EXPTIME-completeness when the maximum predicate arity or the number of variables is bounded [7], [17].

Both FO^2 and GF possess the *finite model property* (FMP), meaning that any satisfiable sentence has a finite model. As a consequence, finite-model reasoning coincides with reasoning under arbitrary models for these fragments. For FO^2 , existence of a finite model of only exponential size in the sentence was actually the path to establishing the above mentioned complexity. For GF, the original FMP result gave rise to a triply exponential bound on the model size [7], whereas a tight doubly-exponential bound was established much more recently [2].

In an attempt to unify FO^2 and GF toward an even more expressive decidable FOL fragment, the *triguarded fragment* (TGF) was introduced [14], extending prior results [10] as well as refining and correcting previous ideas

related to “cross products” [3]. TGF relaxes the guardedness restrictions of GF by allowing non-guarded quantification of subformulas with up to two free variables. The price to pay for retaining decidability is that equality needs to be disallowed, or at least its use must be significantly restricted. TGF brings a new quality, as it allows one to express properties expressible in neither FO^2 nor GF. In particular it embeds one of the most important *prefix classes*, namely Gödel’s class without equality, consisting of prenex formulas of the shape $\exists \bar{x} \forall y_1 y_2 \exists \bar{z} \varphi$. Indeed given such a formula we can translate it to TGF by eliminating the initial prefix of existential quantifiers, replacing the variables \bar{x} by constants, and guarding the block of quantifiers $\exists \bar{z}$ by a dummy guard $G(y_1, y_2, \bar{z})$. Along the same lines, TGF settles an open question by ten Cate and Franceschet [17] about the decidability of formulas of the shape $\exists \bar{x} \forall y_1 y_2 \exists \bar{z} \psi$ where ψ is a guarded formula. In fact, checking satisfiability of TGF is N2EXPTIME-complete, dropping to 2EXPTIME when disallowing constants – as opposed to FO^2 and GF, where presence or absence of constants does not make a difference, complexity-wise – and to NEXPTIME if the arity of predicates is bounded.

One central question left wide open in the original work on TGF [14] is if TGF has the FMP (and thus, if finite model reasoning and the associated complexity is any different from the arbitrary-model case). In that paper, it is noted that neither technique used for establishing the FMP for FO^2 and GF seems to directly lend itself for solving the question for TGF, yet it is conjectured that the FMP holds. Indeed one of this paper’s core contributions is to answer this open question to the positive.

An important, practically relevant and theoretically challenging modelling feature is *transitivity* of a binary relation. Neither FO^2 , nor GF, and also not TGF allow for axiomatising transitivity. As a remedy, it has been suggested to provide a set of dedicated binary predicate names whose transitivity is “hard-wired” into the logic, that is, externally imposed by the semantics. As it turned out, when doing so, one has to be very careful not to lose decidability. Unrestricted use of transitive relations in GF is known to lead to undecidability [7], this even holds for $\text{GF}^2 (= \text{FO}^2 \cap \text{GF})$, the two-variable guarded fragment [5].

One way out is to restrain the use of transitive relations so that they only are allowed to occur in guards. Indeed, sat-

isfiability of GF+TG (*GF with transitive guards*) was shown to be decidable and, in fact, 2EXPTIME-complete [16], as was – more recently – satisfiability of TGF+TG [11]. Results for the finite model case are less extensive: so far, only finite satisfiability of GF²+TG was shown to be decidable and 2EXPTIME-complete [12]. We note that GF²+TG does not have the FMP: indeed, a typical infinity axiom saying that, for a transitive relation T , every element has a T -successor but is not related by T to itself is naturally expressible in GF²+TG. We remark that all the results concerning logics with TG assume the absence of constants. It is conjectured that adding constants to the picture is technically challenging but generally possible without hazarding decidability.

In this paper, we significantly advance the state of the art in finite model theory for the (tri)guarded fragment with and without transitive guards showing the following:

- TGF (with and without constants) has the FMP, thus finite satisfiability coincides with satisfiability known to be N2EXPTIME-complete with and 2EXPTIME-complete without constants.
- Finite satisfiability of constant-free GF+TG (with equality) is decidable and 2EXPTIME-complete.
- Finite satisfiability of constant-free TGF+TG (without equality) is decidable and 2EXPTIME-complete.
- All three results come with a tight upper bound on the size of the finite model which is doubly exponential in the formula length.

The results are established through novel, rather elaborate, carefully crafted model constructions coupled with meticulous inspections of existing proofs toward the extraction of tight bounds.

2. Logics

We work with signatures containing relation symbols of arbitrary positive arity and, possibly, constant symbols. We refer to structures using Fraktur capital letters $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \dots$, and to their domains using the corresponding Roman capitals A, B, C, \dots . Given a structure \mathfrak{A} and some $B \subseteq A$ we denote by $\mathfrak{A}|_B$ the restriction of \mathfrak{A} to its subdomain B .

We usually use a, b, \dots to denote domain elements of structures, \bar{a}, \bar{b}, \dots for tuples of domain elements, x, y, \dots for variables, \bar{x}, \bar{y}, \dots for tuples of variables, and c for constants, all of these possibly with decorations. For a tuple of variables \bar{x} we use $\psi(\bar{x})$ to denote that a formula (subformula) ψ has at most free variables from \bar{x} . Where convenient, tuples of elements will be treated as sets built out of its members.

For a structure \mathfrak{A} , a formula ψ with free variables \bar{x} , and a tuple \bar{a} of elements of A of the same length as \bar{x} , we will write $\mathfrak{A} \models \psi[\bar{a}]$ to denote that $\psi(\bar{x})$ is satisfied in \mathfrak{A} under the assignment $\bar{x} \mapsto \bar{a}$.

For $\ell > 0$, an (*atomic*) ℓ -*type* over a finite signature σ is a maximal consistent set of atomic or negated atomic formulas over σ in ℓ variables x_1, \dots, x_ℓ (we note that types contain equalities/inequalities and occurrences of constants if they are present in σ). A *type* is an ℓ -type for some ℓ . We

often identify a type with the formula obtained by taking the conjunction over its elements. A type is *guarded* if it contains a positive literal containing all its variables. Note that all 1-types are guarded as they all contain the atom $x_1 = x_1$.

Let \mathfrak{A} be a structure, and let \bar{a} be a tuple of its elements. We denote by $\text{tp}^{\mathfrak{A}}(\bar{a})$ the unique type *realized* in \mathfrak{A} by the tuple \bar{a} , i.e., the type $\alpha(\bar{x})$ such that $\mathfrak{A} \models \alpha[\bar{a}]$. We say that \bar{a} is *guarded* in \mathfrak{A} if \bar{a} is built out of a single element or there is a tuple of elements \bar{b} containing all the elements of \bar{a} and a relation symbol $P \in \sigma$ such that $\mathfrak{A} \models P[\bar{b}]$ (i.e., \bar{b} realizes a guarded type).

We will be particularly interested in types over signatures σ consisting of the relation symbols (and constants, if present) used in some given formula. A particularly important role will be played by 1-types and 2-types. Observe that, in the absence of constants, the number of 1-types is bounded by a function which is exponential in $|\sigma|$, and hence also in the length of the formula. This is because any 1-type just corresponds to a subset of σ . On the other hand, when at least one constant c is present, then the number of 1-types may be doubly exponentially large. This is because a 1-type must completely describe the substructure on a given element and the interpretation of c , and there are 2^{2^n} relations of arity n on a pair of elements.

A 2-type will be called *non-degenerate* if it contains $x_1 \neq x_2$. The number of 2-types may be doubly exponential in the length of the formula even in the absence of constants.

Given a formula φ , its *width* is the maximal number of free variables across all subformulas of φ , whereas for a signature σ , its *width* is the maximal arity among the symbols in σ .

Guarded fragment. The set of GF formulas is defined as the least set such that

- 1) every atomic formula belongs to GF,
- 2) GF is closed under the standard boolean connectives $\vee, \wedge, \neg, \Rightarrow, \Leftrightarrow$, and
- 3) if $\psi(\bar{x}, \bar{y}) \in \text{GF}$ then $\forall \bar{x}(\gamma(\bar{x}, \bar{y}) \Rightarrow \psi(\bar{x}, \bar{y}))$ and $\exists \bar{x}(\gamma(\bar{x}, \bar{y}) \wedge \psi(\bar{x}, \bar{y}))$ are in GF, where $\gamma(\bar{x}, \bar{y})$ is an atomic formula containing all the free variables of ψ .

The atoms γ relativising quantifiers in point (3) of the above definition are called the *guards* of the quantifiers. For convenience, we sometimes allow ourselves to leave quantifiers for subformulas with at most one free variable to be unguarded (formally speaking, they can be guarded by atoms $x = x$; such guards cause no problems even in those of our constructions in which equalities are generally forbidden).

In GF we admit the use of equality and constants, but function symbols of arity greater than zero are forbidden.

Triguarded fragment. TGF is an extension of equality-free GF in which quantification for subformulas with at most two variables need not be guarded. Formally, the set of TGF formulas is defined by taking the three syntax rules defining GF formulas (substituting in them 'GF' to 'TGF') and adding the following rule:

- 4) if $\psi(x, y)$ is in TGF, then $\exists x\psi(x, y)$ and $\forall x\psi(x, y)$ belong to TGF.

For convenience, instead of TGF we will mostly work with the equivalent logic GFU, *the guarded fragment with universal role*. We assume that signatures for GFU always contain the distinguished binary relation symbol U. The set of GFU formulas is then defined precisely as the set of GF formulas, but the set of admissible models is restricted to those which interpret U as the universally true relation. Structures interpreting U in this way will be called U-biquitous structures.

It should be clear that TGF and GFU have the same expressive power (modulo the presence of the extra predicate U). For example, the TGF-formula $\forall xy(P(x) \wedge Q(y) \Rightarrow \exists zR(x, y, z))$ can be transformed to the (up to U) equivalent GFU-formula $\forall xy(U(x, y) \Rightarrow (P(x) \wedge Q(y) \Rightarrow \exists zR(x, y, z)))$. In the opposite direction, GFU-formulas can be equivalently translated to TGF just by appending to them the conjunct $\forall xyU(x, y)$, thereby axiomatising U.

In our constructions, we will frequently interpret GFU formulas over non-U-biquitous structures. In this case, they are treated as usual GF formulas.

Logics with transitive guards. The *guarded fragment with transitive guards*, GF+TG, is the logic whose formulas are constructed over purely relational signatures containing distinguished binary symbols T_1, T_2, \dots . The syntax of GF+TG is defined as the syntax of GF, with the only difference that T_1, T_2, \dots can be used only as guards. The equality symbol is allowed. Regarding the semantics, we require admissible structures to interpret T_1, T_2, \dots as transitive relations.

The *triguarded fragment with transitive guards*, denoted TGF+TG, is obtained from GF+TG, as expected, by eliminating equality, and allowing quantification for subformulas with at most two free variables to be unguarded. As in the case of TGF, instead of TGF+TG we will mostly work with the equivalent logic GFU+TG, *the guarded fragment with universal role and transitive guards*, whose signatures contain the special binary symbol U. The syntax of GFU+TG is as the syntax of GF+TG, and the set of admissible models is restricted to U-biquitous ones interpreting T_1, T_2, \dots as transitive relations.

Normal form. We say that a GF (GFU, GF+TG, GFU+TG) formula is in *normal form* if it is of the shape

$$\bigwedge_i \forall \bar{x}(\gamma_i(\bar{x}) \Rightarrow \exists \bar{y}(\gamma'_i(\bar{x}, \bar{y}) \wedge \psi_i(\bar{x}, \bar{y}))) \wedge \bigwedge_j \forall \bar{x}(\gamma_j(\bar{x}) \Rightarrow \psi_j(\bar{x})) \quad (1)$$

where the γ_i and the γ'_i and γ_j are guards and the ψ_i, ψ_j are quantifier-free. The conjuncts indexed by i will be sometimes called $\forall\exists$ -conjuncts, while the conjuncts indexed by j will be called \forall -conjuncts. Note that in our normal form we do not explicitly include purely existential conjuncts like $\exists \bar{y}(\gamma(\bar{y}) \wedge \psi(\bar{y}))$, which sometimes appear in similar normal forms; nevertheless, we will occasionally allow ourselves to use them, as they can be always simulated by $\forall\exists$ -conjuncts $\forall x(x = x \Rightarrow \exists \bar{y}(G(x, \bar{y}) \wedge \gamma(\bar{y}) \wedge \psi(\bar{y})))$, for a fresh G .

Let φ be normal form formula, $\mathfrak{A} \models \varphi$, ζ_i the i -th $\forall\exists$ -conjunct of φ and \bar{a} a tuple of elements of A such that $\mathfrak{A} \models \gamma_i[\bar{a}]$. We then say that a tuple \bar{b} such that $\mathfrak{A} \models \gamma'_i(\bar{a}, \bar{b}) \wedge \psi_i(\bar{a}, \bar{b})$ is a *witness* for \bar{a} and ζ_i .

The following lemma will allow us, when dealing with (finite) satisfiability or analysing the size of minimal models of GF (GFU) or GF+TG (GFU+TG) formulas, to concentrate on normal form sentences of the shape as in (1). A proof of a very similar lemma can be found in [16] (see Lemma 2 there).

Lemma 1. Let φ_0 be a GF (GFU, GF+TG, GFU+TG) formula over a signature σ_0 . Then one can effectively compute a set $\Delta = \{\varphi'_1, \dots, \varphi'_d\}$ of normal form GF (GFU, GF+TG, GFU+TG) formulas over an extended signature $\sigma = \sigma_0 \cup \sigma_{\text{aux}}$ of size polynomial in $|\sigma_0|$ such that all the φ'_i are of length polynomial in $|\sigma_0|$, d is at most exponential in $|\sigma_0|$, $\bigvee_{s \leq d} \varphi'_s \models \varphi_0$ and every $\mathfrak{A} \models \varphi_0$ has a σ -expansion $\mathfrak{A}' \models \bigvee_{s \leq d} \varphi'_s$.

We conclude this subsection with two simple observations with straightforward proofs allowing one to build bigger models from existing ones. Both of them are intended to be used in the absence of constants. Their variants for the case with constants will be presented in the Appendix.

Let σ be a purely relational signature. Let $(\mathfrak{A}_i)_{i \in \mathcal{I}}$ be a family of σ -structures having disjoint domains. Their *disjoint union* is the structure \mathfrak{A} with domain $A = \bigcup_{i \in \mathcal{I}} A_i$ such that $\mathfrak{A}|_{A_i}$ is equal to \mathfrak{A}_i and for any tuple \bar{a} containing elements from at least two different A_i and any relation symbol $P \in \sigma$ of arity $|\bar{a}|$ we have $\mathfrak{A} \models \neg P(\bar{a})$.

Lemma 2. Let φ be a GF or GF+TG normal form formula over a purely relational signature. The disjoint union of any family of its models is also its model.

Let \mathfrak{A}_- be a σ -structure. Its *doubling* is the structure \mathfrak{A} built out of two copies of \mathfrak{A}_- . Formally, its domain is $A := A_- \times \{0, 1\}$ and for each $P \in \sigma$ we set $\mathfrak{A} \models P[(a_1, \ell_1), \dots, (a_k, \ell_k)]$ iff $\mathfrak{A}_- \models P[a_1, \dots, a_k]$ for all $a_i \in A_-$ and $\ell_i \in \{0, 1\}$.

Lemma 3. Let φ be a normal form GF, GFU, GF+TG, GFU+TG formula which does not use equality (or uses it only as trivial guards $x = x$) and let \mathfrak{A}_- be a model of φ . Then its doubling \mathfrak{A} is still a model of φ .

External constructions and procedures. In our work we will extensively use results on the complexity of guarded logics and on the size of their minimal finite models. We collect the relevant results in this paragraph. Generally, bounds in the original papers are formulated in terms of the length of the input formula. We additionally give some more specific estimations, implicit in the original works, obtained by careful yet routine analysis of the proofs. Some comments concerning these estimations can be found in the Appendix.

Theorem 4 ([2]). GF (with constants and equalities) has the finite model property. Every satisfiable formula has a model of size bounded doubly exponentially in its length. More specifically, the size of minimal models

of normal form formulas is bounded exponentially in the size of the signature and doubly exponentially in its width.

Theorem 5 ([7]). The satisfiability problem for GF (with constants and equalities) is 2EXPTIME-complete. More specifically, there is a procedure that, given a normal form formula, works in time bounded polynomially in the length of the input, exponentially in the size of the signature, and doubly exponentially in its width.

Theorem 6 ([12]). Every finitely satisfiable $\text{GF}^2 + \text{TG}$ formula (without constants, with equalities) has a model of size bounded doubly exponentially in its length. More specifically, for normal form formulas, the size of their minimal finite models is bounded exponentially in the number of the $\forall\exists$ -conjuncts of the input and doubly exponentially in the size of the signature.

Theorem 7 ([12]). The satisfiability problem for $\text{GF}^2 + \text{TG}$ (without constants, with equalities) is 2EXPTIME-complete. More specifically, there is a procedure that, given a normal form formula, works in time bounded polynomially in the length of its input, exponentially in the number of its $\forall\exists$ -conjuncts and doubly exponentially in the size of the signature.

3. Finite model construction for TGF (GFU)

Let us fix a GFU sentence φ in normal form, without equality, over a purely relational signature σ (we will explain how to cover the case of signatures containing constants later) and let \mathfrak{A} be a U-biquitous model of φ . Our goal is to build a finite U-biquitous model \mathfrak{A}' of φ .

3.1. Preparing building blocks

Let α be the set of 1-types realized in \mathfrak{A} . We construct a GF σ -sentence φ^* by appending to φ the following conjuncts:

$$\forall x \left(\bigvee_{\alpha \in \alpha} \alpha(x) \right) \quad (2)$$

$$\bigwedge_{\alpha, \alpha' \in \alpha} \exists xy (\alpha(x) \wedge \alpha'(y) \wedge U(x, y) \wedge U(y, x)) \quad (3)$$

$$\bigwedge_{P \in \sigma} \forall \bar{x} \left(P(\bar{x}) \Rightarrow \bigwedge_{1 \leq i, j \leq |\bar{x}|} U(x_i, x_j) \right) \quad (4)$$

saying, respectively, that only 1-types from α are realized, every pair of 1-types has a realization both-ways connected by U and every guarded pair of elements is connected by U. We can treat (2)–(4) as normal form conjuncts.

It is clear that φ^* , treated as a GF-formula, is satisfiable. In fact, \mathfrak{A} is its model. Thus, by the finite model property for GF, it also has a finite (not necessarily U-biquitous) model. We take such a finite model $\mathfrak{C}_- \models \varphi^*$, and let \mathfrak{C} be its doubling. As φ^* does not use equality (or, to be strict, needs it only for trivial guards $x = x$, omitted from (2)), we have by Lemma 3 that $\mathfrak{C} \models \varphi^*$.

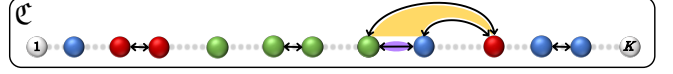


Figure 1. An example structure \mathfrak{C} . Different colours of nodes represent different 1-types. The black bidirectional edges depict U-connections; the orange connection represents a ternary atom; the violet connection – a binary one. While \mathfrak{C} needs not be U-biquitous, for any pair of node colours there is a pair of distinct nodes connected by U.

Moreover, \mathfrak{C} has another convenient property. Let us call elements $a, a' \in C$ *indistinguishable* in \mathfrak{C} if for any relation symbol $P \in \sigma$, any tuple $\bar{a}_1 \subseteq C$ and any tuple \bar{a}_2 obtained from \bar{a}_1 by replacing some occurrences of a by a' and some occurrences of a' by a we have that $\mathfrak{C} \models P[\bar{a}_1]$ iff $\mathfrak{C} \models P[\bar{a}_2]$. Then the following holds (see Fig. 1):

Claim 8. For any pair of 1-types $\alpha, \alpha' \in \alpha$ there is a pair of their distinct realizations a, a' in \mathfrak{C} such that $\mathfrak{C} \models U[a, a'] \wedge U[a', a]$. Moreover, if $\alpha = \alpha'$, then we even find indistinguishable a, a' with that property.

Proof: Let b, b' be elements witnessing the corresponding conjunct from subsentence (3) of φ^* in \mathfrak{C}_- . If $\alpha \neq \alpha'$ then b and b' are distinct and we can take $a = (b, 0)$ and $a' = (b', 0)$. If $\alpha = \alpha'$ then we take $a = (b, 0)$ and $a' = (b, 1)$. By the construction of \mathfrak{C} , a and a' have the required property. Note in particular that all 1-types in \mathfrak{C}_- contain $U(x, x)$ as they are realized in a U-biquitous model of φ . This implies that $\mathfrak{C} \models U[a, a'] \wedge U[a', a]$. \square

From this point on, the model \mathfrak{C}_- will not play any role. However, it will be convenient to build, using Lemma 2, yet another model $\mathfrak{B} \models \varphi^*$, this time as the disjoint union of five copies of \mathfrak{C} . Letting $K = |C|$, we assume that the domain of \mathfrak{B} is $B := \{1, \dots, 5K\}$; and that for $m = 0, \dots, 4$ the structure on $\{mK + 1, \dots, mK + K\}$ is isomorphic to \mathfrak{C} .

3.2. U-saturation

We now build a finite sequence of finite structures $\mathfrak{A}_0, \mathfrak{A}_1, \dots, \mathfrak{A}_f$, each of them being a model of φ^* and the last of them being a desired U-biquitous model \mathfrak{A}' of φ^* (and thus also of φ).

The domains of all these structures will be identical.

$$A_i = B \times \{1, \dots, 5K\} \times \{1, \dots, 5K\}.$$

The initial structure \mathfrak{A}_0 is defined as the disjoint union of $(5K)^2$ copies of \mathfrak{B} . Namely, for each $k, \ell \in \{1, \dots, 5K\}$ we make $\mathfrak{A}_0 \upharpoonright B \times \{k\} \times \{\ell\}$ isomorphic to \mathfrak{B} (via the natural projection $(b, k, \ell) \mapsto b$). By Lemma 2 we have that $\mathfrak{A}_0 \models \varphi^*$.

It is helpful to think that each of the \mathfrak{A}_i is organized in a square table of size $5K \times 5K$. In particular every cell of \mathfrak{A}_0 contains a copy of \mathfrak{B} (which itself is a 5-fold copy of \mathfrak{C}), and in \mathfrak{A}_0 , there are no connections whatsoever between elements from different cells.

Outline of the construction. The whole process may be seen as a careful saturation of the initial model \mathfrak{A}_0 with

U-connections. In the passage from \mathfrak{A}_i to \mathfrak{A}_{i+1} we take a pair of distinct domain elements b_1, b_2 not connected by U yet. By Claim 8, we can find in \mathfrak{C} a pair of distinct elements a_1, a_2 that have the same 1-types as b_1, b_2 , but, in addition, are connected by U. We want to make the connection between b_1 and b_2 isomorphic to the connection between a_1 and a_2 , but after this, b_1, b_2 may start to satisfy the guard γ_i in one of the $\forall\exists$ -conjuncts and thus require witnesses. To provide such witnesses we connect the pair b_1, b_2 to one substructure located in one of the cells in \mathfrak{A}_{i+1} . Thereby, the challenge is to design a strategy which will allow us to perform a process of this kind without causing conflicts regarding the newly assigned connections. We now propose such a strategy.

Some notation. To describe our strategy in detail, let us introduce some further notation. We denote by $\mathfrak{B}_i^{k,\ell}$ the structure in the cell (k, ℓ) of \mathfrak{A}_i , that is the structure $\mathfrak{A}_i|B \times \{k\} \times \{\ell\}$. We recall that $\mathfrak{B}_0^{k,\ell}$ is isomorphic to \mathfrak{B} . We will sometimes say that an element (b, k, ℓ) is the b -th element of $\mathfrak{B}_i^{k,\ell}$. Further, for $m = 0, \dots, 4$, we denote by $\mathfrak{C}_i^{k,\ell,m}$ the structure $\mathfrak{B}_i^{k,\ell}| \{mK + 1, \dots, mK + K\} \times \{k\} \times \{\ell\}$. We recall that each $\mathfrak{C}_0^{k,\ell,m}$ is isomorphic to \mathfrak{C} . See Fig. 2.

Entry elements and their use. For any $1 \leq k, \ell \leq 5K$, let $\alpha^k = \text{tp}^{\mathfrak{B}}(k)$ and $\alpha^\ell = \text{tp}^{\mathfrak{B}}(\ell)$. For each such pair k, ℓ we now choose a pair of *entry elements* for each of the five structures in the cell (k, ℓ) of \mathfrak{A}_0 , that is for the structures $\mathfrak{C}_0^{k,\ell,m}$ ($m = 0, \dots, 4$).

By Claim 8, there are distinct elements $e_1, e_2 \in \mathfrak{C}$ such that $\mathfrak{C} \models \alpha^k[e_1] \wedge \alpha^\ell[e_2] \wedge U[e_1, e_2] \wedge U[e_2, e_1]$ and if $\alpha^k = \alpha^\ell$ then e_1 and e_2 are indistinguishable in \mathfrak{C} . We choose the entry elements $e_1^{k,\ell,m}, e_2^{k,\ell,m}$ to $\mathfrak{C}_0^{k,\ell,m}$ to be the corresponding copies of e_1 and e_2 in each of $\mathfrak{C}_0^{k,\ell,m}$ (recalling that the domains of all the \mathfrak{A}_i are the same, the elements $e_1^{k,\ell,m}, e_2^{k,\ell,m}$ belong to $C_i^{k,\ell,m}$, for all i). The entry elements will serve as a template for connecting some external pairs of elements to $\mathfrak{C}_i^{k,\ell,m}$. This will be done by the following construction.

By $+\mathfrak{C}_i^{k,\ell,m}$ we denote the structure with domain $C_i^{k,\ell,m} \cup \{b_1, b_2\}$ for some fresh elements b_1, b_2 such that $+\mathfrak{C}_i^{k,\ell,m} \upharpoonright C_i^{k,\ell,m} = \mathfrak{C}_i^{k,\ell,m}$ and for each $P \in \sigma$ and each tuple \bar{a} containing at least one of b_1, b_2 we have $+\mathfrak{C}_i^{k,\ell,m} \models P[\bar{a}]$ iff $\mathfrak{C}_0^{k,\ell,m} \models P[\mathfrak{h}(\bar{a})]$, where \mathfrak{h} is the function defined as $\mathfrak{h}(b_1) = e_1^{k,\ell,m}$, $\mathfrak{h}(b_2) = e_2^{k,\ell,m}$ and $\mathfrak{h}(a) = a$ for $a \in C_i^{k,\ell,m}$ (we emphasize that in this definition we copy the relations from $\mathfrak{C}_0^{k,\ell,m}$ and not from its possibly modified version $\mathfrak{C}_i^{k,\ell,m}$). In particular $+\mathfrak{C}_i^{k,\ell,m} \models \alpha^k[b_1] \wedge \alpha^\ell[b_2] \wedge U[b_1, b_2] \wedge U[b_2, b_1]$.

From \mathfrak{A}_i to \mathfrak{A}_{i+1} . Assume now that the structure \mathfrak{A}_i has been defined, for some $i \geq 0$. If \mathfrak{A}_i is U-biquitous then we are done. Otherwise let b_1, b_2 be a pair of elements in A_i such that $\mathfrak{A}_i \models \neg U[b_1, b_2]$. For $s = 1, 2$ let k_s, ℓ_s, n_s be such that b_s is the n_s -th element of $\mathfrak{B}_i^{k_s, \ell_s}$. Let us choose $t \in \{0, \dots, 4\}$ such that $\mathfrak{C}_i^{n_1, n_2, t}$ does not contain the k_1 -th, ℓ_1 -th, k_2 -th or ℓ_2 -th element of $\mathfrak{B}_i^{n_1, n_2}$. Such a t must exist by the pigeon hole principle. We make the

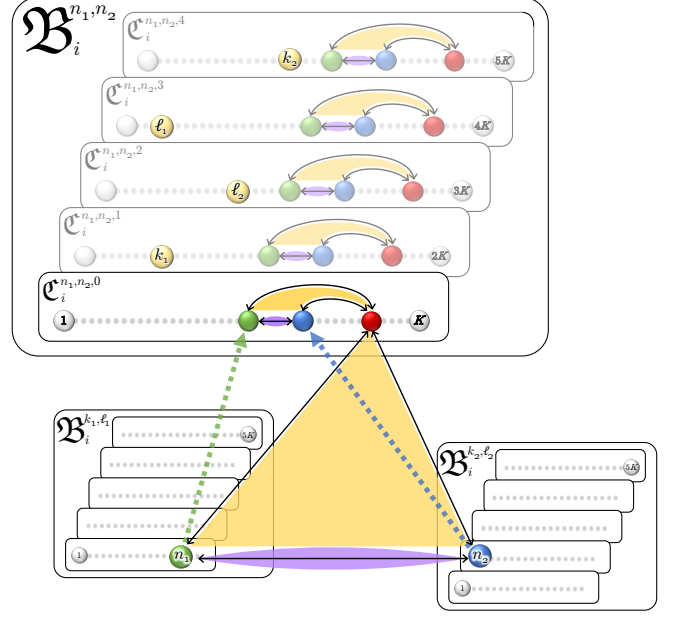


Figure 2. A single step of U-saturation. The green and blue elements in $\mathfrak{B}_i^{n_1, n_2}$ are the entry elements (to the $\mathfrak{C}_0^{n_1, n_2, m}$). The orange and violet connections, as well as the black U-connections inside $\mathfrak{C}_i^{n_1, n_2, 0}$ are already present in $\mathfrak{C}_0^{n_1, n_2, 0}$ (isomorphic to \mathfrak{C}). They give rise to the orange, violet and black connections joining various cells of the table. For transparency, not all newly arising connections are shown.

structure $\mathfrak{A}_{i+1}|C_i^{n_1, n_2, t} \cup \{b_1, b_2\}$ isomorphic to $+\mathfrak{C}_i^{n_1, n_2, t}$. The rest of the structure \mathfrak{A}_i remains untouched. Fig. 2 illustrates the described step. Note that the orange ternary atom from $\mathfrak{C}_i^{n_1, n_2, 0}$ is inherited from $\mathfrak{C}_0^{n_1, n_2, 0}$. Indeed a quick inspection shows that our construction never adds new local ternary atoms, where by a *local atom* we mean an atom in one of the substructures $\mathfrak{C}_i^{k, \ell, m}$. Since $\mathfrak{C}_0^{n_1, n_2, 0}$ satisfies (4) also the black connections shown in $\mathfrak{C}_i^{n_1, n_2, 0}$ are present already in $\mathfrak{C}_0^{n_1, n_2, 0}$. Later we will explain that our construction never modifies guarded types, so also the violet connection is present already in $\mathfrak{C}_0^{n_1, n_2, 0}$.

3.3. Correctness of the construction.

We argue that for all i we have $\mathfrak{A}_i \models \varphi^*$. Note first that our construction never modifies the 1-types of elements.

Claim 9. For every i we have that $\text{tp}^{\mathfrak{A}_i}(a) = \text{tp}^{\mathfrak{A}_0}(a)$.

Proof: The proof goes by induction. Assume that for all a and some i we have that $\text{tp}^{\mathfrak{A}_i}(a) = \text{tp}^{\mathfrak{A}_0}(a)$. In the passage from \mathfrak{A}_i to \mathfrak{A}_{i+1} we modify only some substructure $\mathfrak{A}_i|C_i^{k, \ell, m} \cup \{b_1, b_2\}$, where b_1 is the k -th element of its cell and b_2 is the ℓ -th element of its cell. By the inductive assumption they retain in \mathfrak{A}_i their 1-types from \mathfrak{A}_0 which are, α^k and α^ℓ , respectively. In this step, we do not modify $\mathfrak{C}_i^{k, \ell, m}$ at all, so in particular its elements retain their 1-types. The 1-type of b_1 (b_2) is set to be equal to the type of the first (second) entry element of $\mathfrak{C}_0^{k, \ell, m}$ which is, by our

definition, of type α^k (α^ℓ). So also b_1 and b_2 do not change their 1-types. \square

The following claim is crucial for the correctness of our construction.

Claim 10. Let $i > 0$ and assume $\mathfrak{A}_i \models \varphi^*$. Then every guarded tuple \bar{a} of domain elements in \mathfrak{A}_i (including the tuples guarded by U) retains its type in \mathfrak{A}_{i+1} , that is: $\text{tp}^{\mathfrak{A}_{i+1}}(\bar{a}) = \text{tp}^{\mathfrak{A}_i}(\bar{a})$.

Proof: \mathfrak{A}_{i+1} is obtained from \mathfrak{A}_i by making changes only in the substructure with domain $C_i^{n_1, n_2, t} \cup \{b_1, b_2\}$ (where n_1, n_2, t, b_1, b_2 are as in the description of the construction of \mathfrak{A}_{i+1}). The substructure $\mathfrak{C}_i^{n_1, n_2, t}$ itself is not touched at all.

By the conjunct (4) of φ^* we have that the pair b_1, b_2 cannot be guarded in \mathfrak{A}_i . Thus any tuple guarded in \mathfrak{A}_i which could potentially change its type in \mathfrak{A}_{i+1} must contain exactly one of b_1, b_2 and (possibly) some elements of $C_i^{n_1, n_2, t}$. Consider one such tuple \bar{a} .

If \bar{a} is built exclusively from b_1 or exclusively from b_2 then the claim follows from Claim 9.

Consider the case when \bar{a} contains exactly one of b_1, b_2 and at least one other element. Then \bar{a} contains either elements from two different cells of the $5K \times 5K$ table, or elements from two different substructures $\mathfrak{C}_i^{n_1, n_2, m}$ in the cell (n_1, n_2) (the substructure containing b_1/b_2 and the substructure with $m = t$); in both cases it is not guarded in \mathfrak{A}_0 . So, its type had to be modified in the passage from \mathfrak{A}_j to \mathfrak{A}_{j+1} for some $j < i - 1$ (it is also possible that it was defined in several such passages; in this case assume that $j \rightarrow j + 1$ is the last of them).

W.l.o.g. assume that out of b_1, b_2 the tuple \bar{a} contains b_1 . Thus the two cells which contain the elements of \bar{a} are (n_1, n_2) which contains $C_i^{n_1, n_2, t}$, and (k_1, l_1) which contains b_1 (as we noted, it is possible that this is actually the same cell). Moreover, in the structure $\mathfrak{B}_i^{n_1, n_2}$ from the cell (n_1, n_2) its k_1 -th and l_1 -th elements are not members of \bar{a} , which is ensured by our choice of t . Hence, by our strategy, none of the elements of $\bar{a} \setminus \{b_1\}$ was a member of a pair of elements which was connected to $\mathfrak{C}_s^{k_1, l_1, m}$ for any s , and thus it must be the case that the element b_1 , together with some other element b_3 (having the same 1-type as b_2), were connected to $\mathfrak{C}_j^{n_1, n_2, t}$ when forming \mathfrak{A}_{j+1} . If the 1-types $\alpha^{n_1}, \alpha^{n_2}$ of b_1 and, resp., b_3 are different, then the truth values of the atoms containing b_1 in \mathfrak{A}_{j+1} were defined in accordance with the truth values of the tuples containing the entry element $e_1^{n_1, n_2, t}$ in the structure \mathfrak{A}_0 , exactly as they are defined in \mathfrak{A}_{i+1} . So, there are no conflicts in this case. If $\alpha^{n_1} = \alpha^{n_2}$, it may happen that the atoms containing b_1 in \mathfrak{A}_{j+1} are defined in accordance with the truth values of the tuples containing the entry element $e_2^{n_1, n_2, t}$. In this case however there are also no conflicts since the entry elements of the structures $\mathfrak{C}_0^{n_1, n_2, t}$ are indistinguishable when $\alpha^{n_1} = \alpha^{n_2}$. \square

By straightforward induction we get:

Claim 11. Every guarded tuple of elements in \mathfrak{A}_0 retains its type in \mathfrak{A}_{i+1} .

We are ready to show that $\mathfrak{A}_i \models \varphi^*$ implies $\mathfrak{A}_{i+1} \models \varphi^*$.

Claim 12. If $\mathfrak{A}_i \models \varphi^*$ then $\mathfrak{A}_{i+1} \models \varphi^*$.

Proof: Let us observe first that $\mathfrak{A}_{i+1} \models \varphi$. For this consider any $\forall\exists$ -conjunct of φ : $\zeta = \forall \bar{x}(\gamma_i(\bar{x}) \Rightarrow \exists \bar{y}(\gamma'_i(\bar{x}, \bar{y}) \wedge \psi_i(\bar{x}, \bar{y})))$ and assume $\mathfrak{A}_{i+1} \models \gamma_i[\bar{a}]$ for some tuple \bar{a} . We consider two cases:

(a) $\mathfrak{A}_i \models \gamma_i[\bar{a}]$. Since $\mathfrak{A}_i \models \varphi$, we have in particular that $\mathfrak{A}_i \models \gamma'_i[\bar{a}, \bar{b}] \wedge \psi_i[\bar{a}, \bar{b}]$ for some tuple \bar{b} . As γ'_i is an atomic formula, the tuple $\bar{a}\bar{b}$ is guarded in \mathfrak{A}_i and by Claim 10 it retains its type in \mathfrak{A}_{i+1} . Hence $\mathfrak{A}_{i+1} \models \gamma'_i[\bar{a}, \bar{b}] \wedge \psi_i[\bar{a}, \bar{b}]$. It follows that $\mathfrak{A} \models \zeta$.

(b) $\mathfrak{A}_i \not\models \gamma_i[\bar{a}]$. In this case the fact $\gamma_i[\bar{a}]$ appeared first in \mathfrak{A}_{i+1} . Recall the construction of \mathfrak{A}_{i+1} and the notation used there. Let $\mathfrak{h} : \{b_1, b_2\} \cup C_i^{n_1, n_2, t} \rightarrow C_i^{n_1, n_2, t}$ be the function returning $e_s^{n_1, n_2, t}$ (for $s = 1, 2$) for b_s and returning a for $a \in C_i^{n_1, n_2, t}$. By the definition of \mathfrak{A}_{i+1} we have that $\mathfrak{C}_0^{n_1, n_2, t} \models \gamma_i[\mathfrak{h}(\bar{a})]$. As $\mathfrak{C}_0^{n_1, n_2, t} \models \zeta$ there is a tuple \bar{b} ($= \mathfrak{h}(\bar{b})$) in $C_0^{n_1, n_2, t}$ such that $\mathfrak{C}_0^{n_1, n_2, t} \models \gamma'_i[\mathfrak{h}(\bar{a}), \mathfrak{h}(\bar{b})] \wedge \psi_i[\mathfrak{h}(\bar{a}), \mathfrak{h}(\bar{b})]$. Since γ'_i is an atom, the tuple $\mathfrak{h}(\bar{a})\mathfrak{h}(\bar{b})$ is guarded in \mathfrak{A}_0 . For any tuple $\bar{a}_0 \subseteq \bar{a}\bar{b}$ not containing any of b_1, b_2 we have that $\bar{a}_0 = (\mathfrak{h}(\bar{a}_0)) \subseteq \mathfrak{h}(\bar{a})\mathfrak{h}(\bar{b})$, so the type of $\bar{a}_0 = (\mathfrak{h}(\bar{a}_0))$ from \mathfrak{A}_0 is retained in \mathfrak{A}_{i+1} by Claim 11. For tuples $\bar{a}_0 \subseteq \bar{a}\bar{b}$ containing b_1 and/or b_2 , their type in \mathfrak{A}_{i+1} is the same as the type of $\mathfrak{h}(\bar{a}_0)$ in \mathfrak{A}_0 , by the definition of \mathfrak{A}_{i+1} . In both cases, the type of \bar{a}_0 in \mathfrak{A}_{i+1} is the same as the type of $\mathfrak{h}(\bar{a}_0)$ in \mathfrak{A}_0 . It follows that $\mathfrak{A}_{i+1} \models \gamma'_i[\bar{a}, \bar{b}] \wedge \psi_i[\bar{a}, \bar{b}]$, and thus $\mathfrak{A}_{i+1} \models \zeta$.

The reasoning for the \forall -conjuncts is similar but simpler (actually, \forall -conjuncts are special case of $\forall\exists$ -conjuncts).

As we noted the conjuncts (2)–(4) are normal form conjuncts and thus we can argue about them exactly as about the conjuncts of φ . \square

That our construction terminates follows from Claim 10. Indeed, it implies that all pairs of elements connected by U in \mathfrak{A}_i remain connected by U in \mathfrak{A}_{i+1} . On the other hand at least one new U-connection appears in \mathfrak{A}_{i+1} : the one between the elements b_1 and b_2 . As the number of elements in the domain of our structures is fixed and finite, after a finite number of steps we end up in a structure \mathfrak{A}_f in which any two elements are connected by U. Since $\mathfrak{A}_0 \models \varphi^*$, Claim 12 implies, by induction, that $\mathfrak{A}_f \models \varphi^*$ and in particular $\mathfrak{A}_f \models \varphi$. Thus, we may take $\mathfrak{A}' := \mathfrak{A}_f$ as the desired U-biquitous model of φ .

3.4. Adding constants

The proof of the FMP for GFU (TGF) presented above can be extended without major problems to the case of signatures containing constants. Here we outline the basic idea, for a more detailed description of the construction see Appendix B.

Given a structure \mathfrak{A} interpreting a signature with constants we call the subset $\hat{A} \subseteq A$ consisting of the interpretations of all constants the *named part* of \mathfrak{A} . We set $\check{A} := A \setminus \hat{A}$ and call it the *unnamed part* of \mathfrak{A} .

We proceed as previously. We take a satisfiable normal form formula φ , expand it to φ^* and take a (not necessarily

U-biquitous) finite model $\mathcal{C}_- \models \varphi^*$. In the absence of constants we extensively used constructions building bigger models out of many copies of some existing ones (Lemmas 2, 3). As this time we cannot reproduce the named part of models, those constructions have to be replaced by ones which multiply only their unnamed parts. That is, when producing \mathcal{C} we double only \mathcal{C}_- , when producing \mathcal{B} we form the disjoint union of five copies of \mathcal{C} , and when producing \mathcal{A}_0 we form the disjoint union of $5K \times 5K$ copies of \mathcal{B} . In each of the above steps, all the copies of the unnamed part of the input model are attached to a single, shared copy of its named part, in such a way that the restriction of the resulting structure to the union of any copy of the unnamed part and the copy of the named part is isomorphic to the input model. In effect, the named part of \mathcal{A}_0 is inherited from the initial model \mathcal{C}_- . It is not difficult to show that after each of the above steps we still have a model of φ^* . In particular $\mathcal{A}_0 \models \varphi^*$.

Next, we perform the U-saturation process. Generally, it goes as previously: we find a pair of elements b_1, b_2 not connected by U, join them by U and connect them to the appropriate cell of the table to provide necessary witnesses. We note only, that this time, this step involves defining the truth values of relations on tuples consisting of the b_i , the elements from the cell to which the b_i are connected and, possibly, the interpretations of constants. The process leads eventually to a U-biquitous model $\mathcal{A}' \models \varphi^*$.

3.5. Size of models

We now estimate the size of finite models that can be produced by a use of our construction.

Assume we want to construct a finite model of a satisfiable formula φ_0 over a signature σ_0 . We first convert φ_0 into a disjunction of normal form formulas as guaranteed by Lemma 1, and choose a satisfiable normal form disjunct φ (over an extended signature σ). We take an arbitrary model $\mathcal{A} \models \varphi_0$. Next we append to φ the auxiliary conjuncts obtaining a normal form formula φ^* , send φ^* to a black box producing a finite but generally non-U-biquitous model $\mathcal{C}_- \models \varphi^*$, form models \mathcal{C} , \mathcal{B} , \mathcal{A}_0 and saturate \mathcal{A}_0 to get finally an U-biquitous model \mathcal{A}' . By Lemma 1, $|\varphi|$ is polynomial in $|\varphi_0|$. $|\varphi^*|$ is exponential in $|\varphi|$ in the case without constants and doubly exponential in the case with constants. This follows from the fact that $|\varphi^*|$ contains the conjuncts (2) and (3) whose size is polynomial in the number of 1-types over σ . So, we need to be careful and avoid estimating the size of \mathcal{A}_0 only in terms of the length of φ^* .

As the external black box procedure we can use any procedure constructing a finite model of a satisfiable GF formula. Let us assume that we use the model produced by the construction from [2]. By Thm. 4 the size of this model is bounded exponentially in the size and doubly exponentially in the width of the signature of φ^* , which is the same as the signature of φ , σ . As the size and the width of σ are bounded by $|\varphi|$ which, by Lemma 1, is polynomial in $|\varphi_0|$, eventually our bound on the size of \mathcal{C}_0 is doubly exponential in the size of the input formula $|\varphi_0|$.

Recall that $|C| = 2|C_0|$, $|B| = 5|C| = 10|C_0|$, and for all i : $|A_i| = |B|^2 \cdot |B| = (10|C_0|)^3$ which is still doubly exponential in $|\varphi_0|$. Thus we get:

Theorem 13. Every satisfiable TGF (GFU) formula φ (with or without constants) has a finite model of size bounded doubly exponentially in the length of φ .

This bound is essentially optimal, since even in GF without constants and equality one can construct a family of satisfiable formulas φ_i , each of them of length polynomial in i , but having only models of size at least 2^{2^i} . This is implicit in [7].

The finite model property of TGF (GFU) implies that its finite satisfiability problem is equal to its satisfiability problem and thus it is 2EXPTIME-complete in the absence of constants and N2EXPTIME-complete with constants as shown in [14].

4. Finite satisfiability of GF+TG and GFU+TG

Let us recall that in case of logics with transitive guards, we work with signatures containing no constants, however, in GF+TG we permit equality. Still, our decidability results for GFU+TG will be obtained in the absence of equality, since, as we said, already GFU with equality is undecidable.

For convenience, we first slightly enhance our normal form. Given a normal form GF+TG or GFU+TG formula as in (1) we split its $\forall\exists$ -conjuncts into those in which γ'_i is a non-transitive symbol and those in which it is transitive. Moreover, for the latter, we assume that the guard γ_i has only one variable. If this is not the case – that is we have a conjunct of the form

$$\forall \bar{x} \left(\gamma_i(x_1, \dots, x_k) \Rightarrow \exists y (\gamma'_i(x_j, y) \wedge \psi_i(x_j, y)) \right)$$

with $k > 1$ and γ'_i using a transitive symbol – we replace it by

$$\begin{aligned} \forall \bar{x} (\gamma_i(x_1, \dots, x_k) \Rightarrow G_i^j(x_j)) \\ \wedge \forall x_j (G_i^j(x_j) \Rightarrow \exists y (\gamma'_i(x_j, y) \wedge \psi_i(x_j, y))) \end{aligned}$$

where G_i^j is a fresh unary symbol.

Further, we assume that all the guards γ_i in the $\forall\exists$ -conjuncts are non-transitive. If this is not the case, that is we have a transitive guard γ_i , say of the form $T(x, y)$, then we replace it by $G(x, y)$, for a fresh, non-transitive symbol G , and append the \forall -conjunct $\forall xy (T(x, y) \Rightarrow G(x, y))$.

Finally, for convenience, we append to normal form formulas a conjunct saying that every guarded pair of elements is connected by Aux, where Aux is a fresh binary symbol. This auxiliary conjunct does not affect satisfiability of the formula.

So, we will assume that normal form formulas for GF+TG and GFU+TG are of the shape:

$$\begin{aligned}
& \bigwedge_h \forall \bar{x} \left(\gamma_h(\bar{x}) \Rightarrow \exists \bar{y} (\vartheta_h(\bar{x}, \bar{y}) \wedge \psi_h(\bar{x}, \bar{y})) \right) \\
& \wedge \bigwedge_i \forall x \left(\gamma_i(x) \Rightarrow \exists y (\theta_i(x, y) \wedge \psi_i(x, y)) \right) \\
& \wedge \bigwedge_j \forall \bar{x} \left(\gamma_j(\bar{x}) \Rightarrow \psi_j(\bar{x}) \right) \\
& \wedge \bigwedge_{P \in \sigma} \forall \bar{x} \left(P(\bar{x}) \Rightarrow \bigwedge_{1 \leq i, j \leq |\bar{x}|} \text{Aux}(x_i, x_j) \right) \quad (5)
\end{aligned}$$

where the γ_h and γ_i are non-transitive guards, γ_j is a guard (transitive or non-transitive), the ϑ_h are non-transitive guards and the θ_i are transitive guards. We recall that the transitive symbols appear in none of ψ_h , ψ_i and ψ_j . The conjuncts indexed by h will be called $\forall\exists^{ntr}$ -conjuncts, the conjuncts indexed by i will be called $\forall\exists^{tr}$ -conjuncts, the conjuncts indexed by j , together with the conjuncts speaking about Aux, will be called \forall -conjuncts.

4.1. GF+TG

Let us fix a finitely satisfiable normal form GF+TG formula φ over a purely relational signature σ , of the shape as in (5). Equalities are allowed in φ . Let \mathfrak{A} be a finite model of φ . We plan to construct a finite model $\mathfrak{A}' \models \varphi$ of size bounded doubly exponentially in $|\varphi|$. Let α be the set of 1-types realized in \mathfrak{A} . Let β be the set of non-degenerate guarded 2-types realized in \mathfrak{A} . For $\beta \in \beta$ let β^- denote the set of formulas obtained from β by removing $T(x_1, x_2)$, $T(x_2, x_1)$, $\neg T(x_1, x_2)$, $\neg T(x_2, x_1)$, for all transitive T , if they are present in β . Note that β^- still contains the literals speaking about $T(x_1, x_1)$ and $T(x_2, x_2)$. β^- will be called the *transitive-free reduction* of β . Also, let \mathfrak{A}^- denote the structure obtained from \mathfrak{A} by removing all facts $T[a, b]$ for a transitive T and $a \neq b$. That is, in \mathfrak{A}^- the only transitive facts may be of the form $T[a, a]$ for some $a \in A$.

Constructing \mathfrak{B}^* and \mathfrak{C}^* . We now construct two auxiliary formulas out of φ . Let

$$\begin{aligned}
\varphi_B &:= \bigwedge_h \forall \bar{x} \left(\gamma_h(\bar{x}) \Rightarrow \exists \bar{y} (\vartheta_h(\bar{x}, \bar{y}) \wedge \psi_h(\bar{x}, \bar{y})) \right) \\
& \wedge \bigwedge_j \forall \bar{x} \left(\gamma_j(\bar{x}) \Rightarrow \psi_j(\bar{x}) \right) \\
& \wedge \bigwedge_{P \in \sigma} \forall \bar{x} \left(P(\bar{x}) \Rightarrow \bigwedge_{1 \leq i, j \leq |\bar{x}|} \text{Aux}(x_i, x_j) \right) \\
& \wedge \bigwedge_{T_s} \forall xy \left(T_s(x, y) \Rightarrow x = y \right) \\
& \wedge \bigwedge_{\beta \in \beta} \exists xy \beta^-(x, y) \\
& \wedge \bigwedge_{\alpha \in \alpha} \exists x \alpha(x) \wedge \forall x \bigvee_{\alpha \in \alpha} \alpha(x)
\end{aligned}$$

That is, φ_B contains all the $\forall\exists^{ntr}$ -conjuncts of φ , all its \forall -conjuncts, plus the conjuncts saying that the transitive relations do not connect distinct elements, for every non-degenerate guarded 2-type realized in \mathfrak{A} its transitive-free reduction is realized (note that this is a guarded formula, since β^- remains guarded as it contains $\text{Aux}(x, y)$), and that a 1-type is realized iff it is realized in \mathfrak{A} . We remark that the conjuncts speaking about α and β may contain transitive atoms outside guards, but these may only be atoms of the form $T(x, x)$ or $T(y, y)$ for some transitive T . We can replace them by $P(x)$ or, resp., $P(y)$, for some fresh P , and add normal form conjuncts ensuring that $\forall x (P(x) \Leftrightarrow \exists y (T(x, y) \wedge x = y))$, obtaining this way formulas in which T is used only in guard positions. Moreover, as the restriction imposed by φ_B on the transitive relations makes their transitivity irrelevant, we will treat φ_B as a GF formula.

Further, let

$$\begin{aligned}
\varphi_C &:= \bigwedge_i \forall x \left(\gamma_i(x) \Rightarrow \exists y (\theta_i(x, y) \wedge \psi_i(x, y)) \right) \\
& \wedge \bigwedge_{j: \gamma_j \text{ transitive}} \forall \bar{x} \left(\gamma_j(\bar{x}) \Rightarrow \psi_j(\bar{x}) \right) \\
& \wedge \bigwedge_{P \in \sigma} \bigwedge_{\bar{x} \in S_P} \forall xy \left(P(\bar{x}) \Rightarrow \text{Aux}(x, y) \right) \\
& \wedge \forall xy \left(\text{Aux}(x, y) \Rightarrow (x \neq y \Rightarrow \bigvee_{\beta \in \beta} \beta^-(x, y)) \right) \\
& \wedge \bigwedge_{\alpha \in \alpha} \exists x \alpha(x) \wedge \forall x \bigvee_{\alpha \in \alpha} \alpha(x)
\end{aligned}$$

where S_P is the set of tuples of length equal to the arity of P built out of variables x and y , and containing at least one occurrence of each of them.

That is, φ_C contains all the $\forall\exists^{tr}$ -conjuncts and those \forall -conjuncts of φ that do not speak about Aux, plus the conjuncts saying that for every guarded tuple built out of two elements these two elements are connected by Aux, every guarded pair of distinct elements satisfies the transitive-free reduction of a guarded 2-type from \mathfrak{A} , and that a 1-type is realized iff it is realized in \mathfrak{A} . Note that the \forall -conjuncts we include here have transitive γ_j , so they use at most two variables; we may assume that they are x, y . The remaining conjuncts also use only variables x and y , so φ_C is a formula belonging to GF²+TG (again after the appropriate adjustments concerning the use of transitive relations in each of the α and β).

Note that both φ_B and φ_C are finitely satisfiable, as the former is satisfied in \mathfrak{A}^- and the latter in \mathfrak{A} . Treating φ_B as a GF formula we take its small finite model \mathfrak{B} as guaranteed by [2]. Similarly, treating φ_C as a GF²+TG formula we take its small finite model \mathfrak{C} , as guaranteed by [12]. We remark here that while [12] considers explicitly only signatures with relation symbols of arity 1 and 2, a routine inspection shows that the constructions there work smoothly even if symbols of arity greater than 2 are allowed, which is the case in our work. Indeed, the presence of relations of arity greater than 2 could be important in [12] only when 2-types are assigned

to pairs of elements. However, those 2-types are read off from a pattern model, and whether they contain higher arity relations or not is not relevant. Since φ_C is a two-variable formula, we may assume that \mathfrak{C} contains no fact with more than two distinct elements. We note that \mathfrak{C} happens to satisfy all the \forall -conjuncts of φ . The conjuncts with transitive γ_j are included explicitly in φ_C ; for those with non-transitive γ_j assume that $\mathfrak{C} \models \gamma_j[\bar{a}]$ for some tuple \bar{a} . By our assumption on \mathfrak{C} the tuple \bar{a} is built out of at most two elements. So, either it uses only one element and then its 1-type is realized in \mathfrak{A} (by the conjunct of φ_C speaking about α), or it uses two elements, and then the transitive-free reduction of their 2-type is the same as the reduction of some 2-type realized in \mathfrak{A} (by the conjunct speaking about β). Since $\mathfrak{A} \models \varphi$, and in particular $\mathfrak{A} \models \forall \bar{x}(\gamma_j(\bar{x}) \Rightarrow \psi_j(\bar{x}))$, and recalling that ψ_j does not contain any transitive relations, it follows that $\mathfrak{C} \models \psi_j[\bar{a}]$.

By the conjuncts of φ_B and φ_C speaking about α , we know that the sets of 1-types realized in \mathfrak{B} and \mathfrak{C} are equal (concretely, they are equal to α). We now construct models $\mathfrak{B}^* \models \varphi_B$ and $\mathfrak{C}^* \models \varphi_C$ such that the number of realizations of α in \mathfrak{B}^* is equal to the number of its realizations in \mathfrak{C}^* , for all α .

Let m be the maximal number of realizations of a 1-type α in \mathfrak{B} (over all $\alpha \in \alpha$). Let \mathfrak{C}^* be the disjoint union of m copies of \mathfrak{C} . By Lemma 2 we have that $\mathfrak{C}^* \models \varphi_C$. Note that for every α the number of realizations of α in \mathfrak{B} is less than or equal to the number of its realizations in \mathfrak{C}^* . To make these numbers equal we successively adjoin additional realizations of the appropriate 1-types to \mathfrak{B} . This is simple: to adjoin a realization b of a 1-type α we choose a pattern element a of type α in \mathfrak{B} and make the structure on $(B \setminus \{a\}) \cup \{b\}$ isomorphic to \mathfrak{B} ; we add no facts containing both a and b to the structure. After each such step the resulting structure is still a model of φ_B . This way we eventually get the desired \mathfrak{B}^* .

Constructing \mathfrak{D} . Let $K = |B^*| = |\mathfrak{C}^*|$. Let b_0, \dots, b_{K-1} be any enumeration of the elements of B^* and let $\alpha_i = \text{tp}^{\mathfrak{B}^*}(b_i)$, for all $0 \leq i < K$. We create a new structure \mathfrak{D} , with domain $D = \{0, \dots, K-1\} \times \{0, \dots, K-1\}$. We set $\text{tp}^{\mathfrak{D}}(k, \ell) := \alpha_{k+\ell \bmod K}$. Viewing in the natural way \mathfrak{D} as a square table, we see that from its every row and its every column one can construct bijections into \mathfrak{B}^* and \mathfrak{C}^* preserving the 1-types. Without modifying the 1-types, we can thus define the structure of \mathfrak{D} on every row and every column in such a way that they become isomorphic copies of \mathfrak{B}^* , and \mathfrak{C}^* , respectively. This completes the definition of \mathfrak{D} , that is, we add no further facts to it. Note that transitive relations cannot connect elements from different columns.

Call a guarded tuple of elements of \mathfrak{D} *vertical (horizontal)* if it belongs to a single column (row) of the table. Observe that the tuples built out of a single element are both vertical and horizontal and every guarded tuple in \mathfrak{D} is either vertical or horizontal by the definition of \mathfrak{D} .

Note that \mathfrak{D} satisfies the $\forall\exists^{tr}$ -conjuncts of φ . Indeed, any element a satisfying some $\gamma_i[a]$ has the required witness in its column, since the relevant conjunct is a member of

φ_C . Also the \forall -conjuncts are satisfied: we have explained that they are satisfied in \mathfrak{C} , so it follows that they are also satisfied in \mathfrak{C}^* , and thus in every column; on the other hand φ_B includes them explicitly, so they are satisfied in every row. Concerning the $\forall\exists^{ntr}$ -conjuncts, note that any horizontal guarded tuple \bar{a} satisfying any of $\gamma_h[\bar{a}]$ has the required witnesses in its row (since its row satisfies φ_B).

The only problem is that some vertical guarded tuples may not have witnesses for some $\forall\exists^{ntr}$ -conjuncts. Note that every such tuple is built out of precisely two elements. Indeed, by our assumption about \mathfrak{C}^* there are no vertical guarded tuples containing three or more distinct elements there, and on the other hand any tuple built out of a single element has its witnesses in its row.

We will fix the above problem by taking an appropriate number of copies of \mathfrak{D} and adjoining every vertical guarded pair of elements to a row in some different copy of \mathfrak{D} . This will be done in a circular way, reminiscent of the small model construction for FO^2 from [8]. We emphasise that this process is much simpler than the U-saturation process from Section 3.2 since this time we do not need to deal with pairs of elements from different copies of our basic building block \mathfrak{D} . Let us turn to details.

Building a small model of φ . Assume that \mathfrak{D}_1 and \mathfrak{D}_2 are two copies of \mathfrak{D} . Consider a vertical guarded pair of distinct elements b_1, b_2 in \mathfrak{D}_1 . As the column of this pair is a model of φ_C we know that it satisfies $\text{Aux}[b_1, b_2]$ and thus also it satisfies $\beta^-[b_1, b_2]$ for some $\beta \in \beta$. (We recall that β^- does not mention any transitive connection between b_1, b_2 but such connections may be present in \mathfrak{D}_1 .) As any row \mathfrak{E} of \mathfrak{D}_2 is a model of φ_B it follows that this row contains a pair of distinct elements a_1, a_2 such that $\mathfrak{E} \models \beta^-(a_1, a_2)$. (Here, there are no transitive connections between a_1 and a_2 as ensured by φ_B .)

We now describe the procedure to which we will later refer by saying: *we connect the pair b_1, b_2 to the row \mathfrak{E} (using a_1, a_2 as a template)*. For any tuple \bar{a} containing at least one of b_1, b_2 and some elements of $E \setminus \{a_1, a_2\}$ and any non-transitive relation symbol $P \in \sigma$ of arity $|\bar{a}|$ we add the fact $P[\bar{a}]$ iff $\mathfrak{E} \models P[\mathfrak{h}(\bar{a})]$ where \mathfrak{h} is the function returning a_s for b_s ($s = 1, 2$) and a for all $a \in E$. In other words, we connect b_1, b_2 with the elements of $E \setminus \{a_1, a_2\}$ exactly as a_1, a_2 are connected with these elements in \mathfrak{E} . We add no other facts. (In particular in this procedure we add no transitive connections.) This way the guarded tuples containing any of b_1, b_2 (or both) and possibly some elements of E , have all the required witnesses for the $\forall\exists^{ntr}$ -conjuncts since the structure we have defined on $(E \setminus \{a_1, a_2\}) \cup \{b_1, b_2\}$ is isomorphic to \mathfrak{E} when the transitive relations are not taken into account (and recall that the $\forall\exists^{ntr}$ -conjuncts do not mention transitive relations at all).

Now, let the structure \mathfrak{A}' be the disjoint union of $3K$ copies of \mathfrak{D} . More specifically its domain is $A' = \{0, 1, 2\} \times \{0, \dots, K-1\} \times D$ and $\mathfrak{A}' \upharpoonright \{i\} \times \{j\} \times D$ is isomorphic to \mathfrak{D} for all i, j . Denote $A_i = \{i\} \times \{0, \dots, K-1\} \times D$, for $i = 0, 1, 2$.

For every vertical guarded pair of distinct elements b_1, b_2 in A_i chose a row in a copy of \mathfrak{D} contained in $A_{(i+1) \bmod 3}$ such that this row has not yet been used by any other pair from the column of b_1 and b_2 and connect b_1, b_2 to this row (using an appropriate template). As the number of pairs of elements in the row of b_1, b_2 is smaller than K^2 and there are K copies of \mathfrak{D} in each of the A_i , and each of them has K rows, we have sufficiently many rows to perform this step.

Using three sets A_i and applying the above described circular strategy guarantee that the process can be performed without conflicts: if an element b is connected to a row \mathfrak{E} then no element of \mathfrak{E} is ever connected to the row of b .

Now it should be clear that indeed the eventually obtained structure models φ .

Size of models and complexity. We now analyse the small model construction described in the previous paragraphs and obtain the following (optimal) bound on the size of minimal finite models for GF+TG. For further purposes in the second part of this theorem we formulate a more specific bound for normal form formulas.

Theorem 14. Every finitely satisfiable GF+TG formula without constants has a model of size bounded doubly exponentially in its length. For finitely satisfiable normal form formulas there are models of size bounded exponentially in the size of the signature and the number of their $\forall\exists$ -conjuncts, and doubly exponentially in the width of the signature.

Proof: Let us summarize the steps needed to produce a small model of an input finitely satisfiable GF+TG sentence φ_0 over a signature σ_0 . We convert it into a disjunction of normal form formulas over an extended signature σ as in Lemma 1, and choose its finitely satisfiable normal form disjunct φ and its finite model \mathfrak{A} . Let r be the number of symbols in σ and w its width. As by Lemma 1 $|\varphi|$ is polynomial in $|\varphi_0|$, both r and w are polynomial in $|\varphi_0|$. Perform our small model construction for φ and let $\varphi_B, \varphi_C, \mathfrak{B}, \mathfrak{C}, \mathfrak{B}^*, \mathfrak{C}^*, \mathfrak{D}$ and \mathfrak{A}' be as in this construction.

Recall that as \mathfrak{B} we take the small model for φ_B (which is a normal form formula over the signature σ), constructed as in [2]. By Thm. 4 the size of \mathfrak{B} is bounded exponentially in r and doubly exponentially in w , that is doubly exponentially in $|\varphi_0|$.

Concerning \mathfrak{C} , we take as it the small model for φ_C (which is a normal form formula over the signature σ), constructed by applying the small model construction from [12]. By Thm. 6, $|\mathfrak{C}|$ is bounded exponentially in the number of its $\forall\exists$ -conjuncts (which are actually taken from φ) and doubly exponentially the size of the signature. So, it is bounded doubly exponentially in $|\varphi_0|$.

Further, each of the structures $\mathfrak{B}^*, \mathfrak{C}^*$ has size at most $|B| \cdot |C|$, the structure \mathfrak{D} – at most $|B^*|^2$ and the final model \mathfrak{A}' – at most $3|B^*||D| = 3|B^*|^3 = 3|B|^3|C|^3$, which is still doubly exponential in $|\varphi_0|$.

The second part of the theorem, concerning normal form formulas follows easily from the information on the size of

\mathfrak{C} and \mathfrak{B} given above and from the observation that the final estimation on $|A'|$ is polynomial in $|B|$ and $|C|$. \square

Thm. 14 immediately yields a N2EXPTIME-upper bound on the complexity of the finite satisfiability problem for GF+TG: it suffices to guess a bounded size structure and verify that it is indeed a model of the input formula. To get the optimal 2EXPTIME-upper complexity bound, instead of guessing a model, we may construct φ_B and φ_C for various sets α and β and test their satisfiability. We first make two observations, Lemma 15 and Lemma 16, the second of which reduces the number of possible choices of β , which will be crucial for lowering the complexity.

Lemma 15. Let φ be a normal form GF+TG formula over a signature σ . Then φ is finitely satisfiable iff there are sets α, β of 1-types, and, resp., non-degenerate guarded 2-types over σ , such that (i) the formulas φ_B and φ_C constructed with such α and β have finite models, (ii) for every $\beta \in \beta$ any two-element structure of type β satisfies all the \forall -conjuncts of φ .

Proof: \Rightarrow Assume φ has a finite model \mathfrak{A} . As α and β take the set of 1-types and, resp., non-degenerate guarded 2-types realized in \mathfrak{A} . Then (i) holds since φ_B is satisfied in \mathfrak{A}^- and φ_C — in \mathfrak{A} (cf. the paragraph about the construction of \mathfrak{B}^* and \mathfrak{C}^*), and (ii) holds since the guarded types in β are taken from a model of φ . \Leftarrow Having α and β satisfying (i) and (ii) we can perform a finite model construction for φ exactly as we did in this section (with condition (ii) used to ensure that \mathfrak{C} respects all the \forall -conjuncts of φ). \square

Lemma 16. If φ_C has a finite model then it has one in which the number of realized 2-types is bounded polynomially in the number of 1-types, the number of the $\forall\exists$ -conjuncts of φ and the number of transitive relations.

Proof: Let \mathfrak{C} be a finite model of φ_C . Recall that φ_C is in GF²+TG and thus, as previously, we may assume that \mathfrak{C} contains no facts with more than two distinct elements. Additionally, we may also assume that every pair of distinct elements in \mathfrak{C} is connected by at most one transitive relation (this condition is ensured in the finite model construction in [12]; models satisfying this condition are called *ramified* there). Let β be the set of 2-types realized in \mathfrak{C} . For a 2-type β , by $\beta \upharpoonright x_1$ ($\beta \upharpoonright x_2$) we denote the subset of β consisting of those literals which use only variable x_1 (x_2); similarly by $\beta \upharpoonright \sigma_{tr}$ we denote the subset of β consisting of those literals which use a transitive symbol. Let β^{-1} be the result of switching the variables in β .

Let us introduce an equivalence relation \sim on β as follows: $\beta_1 \sim \beta_2$ iff the following conditions hold (i) $\beta_1 \upharpoonright x_1 = \beta_2 \upharpoonright x_1$, $\beta_1 \upharpoonright x_2 = \beta_2 \upharpoonright x_2$, $\beta_1 \upharpoonright \sigma_{tr} = \beta_2 \upharpoonright \sigma_{tr}$, (ii) for each conjunct $\forall x(\gamma_i(x) \Rightarrow \exists y(\theta_i(x, y) \wedge \psi_i(x, y)))$ it holds that $\beta_1 \models \theta_i(x, y) \wedge \psi_i(x, y)$ iff $\beta_2 \models \theta_i(x, y) \wedge \psi_i(x, y)$ and $\beta_1^{-1} \models \theta_i(x, y) \wedge \psi_i(x, y)$ iff $\beta_2^{-1} \models \theta_i(x, y) \wedge \psi_i(x, y)$. Observe that $\beta_1 \sim \beta_2$ iff $\beta_1^{-1} \sim \beta_2^{-1}$.

In every equivalence class of \sim we distinguish one of its members. We do this in such a way that if β is distinguished in its class then β^{-1} is also distinguished in its class. For

every pair of elements $a, b \in C$, if its type $\text{tp}^{\mathcal{C}}(a, b)$ is not distinguished in its class, change this type to the one which is distinguished there. The strategy of distinguishing always both β and β^{-1} allows us to perform this process without conflicts which could potentially arise when the types of pairs a, b and b, a are defined.

In so obtained structure \mathcal{C}' the interpretation of the transitive symbols remains unchanged (so they all remain transitive) and every element has precisely the same witnesses for every $\forall\exists$ -conjunct as it has in \mathcal{C} (even though it may be connected to them by different 2-types). The \forall -conjuncts are satisfied in \mathcal{C}' as all the types are imported from \mathcal{C} which is a model of φ_C . Thus, still $\mathcal{C}' \models \varphi_C$.

It is readily verified that the number of equivalence classes of \sim is bounded polynomially in the number of 1-types, in the number of the $\forall\exists$ -conjuncts, and in the number of transitive relations (the latter follows from the fact that \mathcal{C} is ramified). From this we get that \mathcal{C}' is as required. \square

Theorem 17. The finite satisfiability problem for GF+TG without constants is 2EXPTIME-complete. For normal form formulas it works in time polynomial in the size of the input formula, exponential in the number of the $\forall\exists$ -conjuncts, and doubly exponential in the size and width of the signature.

Proof: The lower bound is inherited from GF [7] or from GF²+TG [12]. Let us justify the upper bound. Let φ_0 be any GF+TG formula. Convert it into a disjunction of normal form formulas over a signature σ as in Lemma 1 and test satisfiability of each of its disjuncts. To do the latter, for a single normal form disjunct φ over a signature σ construct all possible sets of 1-types α , and all possible sets of guarded non-degenerate 2-types of cardinality bounded as in Lemma 16. For each pair of such α and β construct φ_B and φ_C , which are normal form formula, and test their finite satisfiability using the algorithm from Thm. 5, and, resp., Thm. 7. Our algorithm returns 'yes' iff for some normal form disjunct φ and some choice of α and β both the external algorithms return 'yes'.

The correctness of the algorithm follows from Lemmas 1, 15 and 16.

Denote h the number of the $\forall\exists$ -conjuncts of φ , k the number of transitive relations, r the size of σ and w its width.

Recall that the number of disjuncts in the normal form of φ_0 is at most exponential in $|\varphi_0|$. Note that the number of 1-types is 2^r , so the number of possible choices for α is 2^{2^r} . Due to Lemma 16 we can restrict attention to sets β of size bounded polynomially in 2^r , h and k . Observing that the number of 2-types is $2^{O(r \cdot 2^w)}$ we see that the number of relevant choices of β is doubly exponential in r and w and singly exponential in h and k .

By Thm. 5 the first of the external procedures works in time polynomial in $|\varphi_B|$, exponential in r and doubly exponential in w . By Thm. 7 the second procedure works in time polynomial in $|\varphi_C|$, exponential in h and doubly exponential in r .

Regarding the size of φ_B and φ_C they are both bounded polynomially in $|\varphi|$, and in the size of α and β , that is they are doubly exponential in r and w and singly exponential in h and k .

Gathering the above, the claim for normal form formulas follows. The upper bound for arbitrary formulas follows from the fact that $|\varphi|$ is polynomial in $|\varphi_0|$, and thus all the parameters r, w, h are polynomial in $|\varphi_0|$. \square

4.2. GFU+TG

Assume that φ is a finitely satisfiable GFU+TG normal form formula without equality and let \mathfrak{A} be its finite (U-biquitous) model. We will explain how to construct a (U-biquitous) model \mathfrak{A}' of φ of size bounded doubly exponentially in $|\varphi|$. The whole construction is almost identical to the construction of a small model of a satisfiable GFU normal form formula in Section 3. There are only two, rather natural, differences: first, obviously, we use a different black box procedure; second, when constructing the ${}^+\mathfrak{C}_i^{k,\ell,m}$ structures we must properly handle the transitive relations.

So, we first construct the auxiliary formula φ^* exactly as in Section 3 (we only need the adjustment concerning transitive atoms of the form $T(x, x)$ used outside the guards in the $\alpha(x)$, similar to that for φ_B and φ_C), and treating it as a GF+TG formula we take its small (not necessarily U-biquitous) model \mathcal{C}_- as guaranteed by Thm. 14. We proceed as in Section 3, building the doubling \mathcal{C} of \mathcal{C}_- , a 5-fold copy \mathfrak{B} of \mathcal{C} , and a $5K \times 5K$ table \mathfrak{A}_0 of copies of \mathcal{C} , where $K = |C|$. All these structures are models of φ^* by Lemmas 3 and 2.

We employ the same notation as in the case of GFU. We choose the entry elements for the structures $\mathfrak{C}_0^{k,\ell,m}$ and proceed to the definition of the structures ${}^+\mathfrak{C}_i^{k,\ell,m}$. As previously, $\mathfrak{C}_i^{k,\ell,m}$ is the structure with domain $C_i^{k,\ell,m} \cup \{b_1, b_2\}$ for some fresh elements b_1, b_2 such that ${}^+\mathfrak{C}_i^{k,\ell,m} \upharpoonright C_i^{k,\ell,m} = \mathfrak{C}_i^{k,\ell,m}$. Concerning the connections involving the new elements b_1, b_2 , for each **non-transitive** $P \in \sigma$ and each tuple \bar{a} containing at least one of b_1, b_2 we set that ${}^+\mathfrak{C}_i^{k,\ell,m} \models P[\bar{a}]$ iff $\mathfrak{C}_0^{k,\ell,m} \models P[\mathfrak{h}(\bar{a})]$, where \mathfrak{h} is the function defined as $\mathfrak{h}(b_1) = e_1^{k,\ell,m}$, $\mathfrak{h}(b_2) = e_2^{k,\ell,m}$ and $\mathfrak{h}(a) = a$ for $a \in C_i^{k,\ell,m}$. That is, as previously, we copy the relations from $\mathfrak{C}_0^{k,\ell,m}$, but only those non-transitive. In particular the elements b_1, b_2 remain not connected by any transitive relation even if the entry elements are connected by some in $\mathfrak{C}_0^{k,\ell,m}$.

We then build successively the structures $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ exactly as in the case of GFU, obtaining finally a U-biquitous structure \mathfrak{A}_f which we take as the desired \mathfrak{A}' .

The correctness of the construction can be proved as in the case of GFU. The analogues of the Claims 9, 10 and 11 can be proved with literally no changes. Also, the proof of Claim 12 is almost the same. As we have emphasized, in our process we do not add any transitive connections; we also do not modify any 2-types containing any transitive connections. So, the \forall -conjuncts with transitive guards are satisfied in all the \mathfrak{A}_i . Note also that we never need new

witnesses for the $\forall\exists^{tr}$ -conjuncts, as such witnesses are required only for tuples built out of a single element, and such tuples have the required witnesses already in \mathfrak{A}_0 . Thus, we only need to take care of witnesses for the conjuncts not mentioning transitive relations, and for satisfaction for the \forall -conjuncts with non-transitive guards. For such conjuncts the fact that we do not copy from \mathfrak{A}_0 the complete types of tuples, but rather their transitive-free parts is not relevant.

Size of models and complexity. The following theorem follows from an analysis of the size of the \mathfrak{A}_i structures, similar to that in the case of GFU, and a use of the estimation on the size of \mathfrak{C}_- from the second part of Thm. 14.

Theorem 18. Every finitely satisfiable GFU+TG (TGF+TG) formula without constants has a model of size bounded doubly exponentially in its length.

Concerning the complexity, we first make the following observation.

Lemma 19. A normal form TGF+TG formula φ without constants is finitely satisfiable iff there exists a set of 1-types α such that the formula φ^* , treated as a GF+TG formula, has a finite (not necessarily U-biquitous) model.

Proof: \Rightarrow If \mathfrak{A} is a finite model of φ then we take as α the set of 1-types realized in \mathfrak{A} , and note that $\mathfrak{A} \models \varphi^*$. \Leftarrow Given a finite (not necessarily U-biquitous) model $\mathfrak{C}_- \models \varphi^*$ we construct a U-biquitous model of φ^* (and thus also of φ) as described above. \square

Finally we get:

Theorem 20. The finite satisfiability problem for GFU+TG (TGF+TG) without constants is 2EXPTIME-complete.

Proof: The lower bound is inherited from GF+TG. To justify the upper bound we design the following algorithm: For input φ_0 we convert it into normal form over a signature σ and for each of its disjuncts φ and each possible choice of α construct φ^* and test its finite satisfiability using the procedure for GF+TG. We answer 'yes' iff at least one of the tests is positive.

The correctness of this algorithm follows from Lemmas 1 and 19. Recall that the number of disjuncts in normal form of φ_0 is exponential in $|\varphi_0|$ and the number of choices of α is doubly exponential in $|\sigma|$ and thus also in $|\varphi_0|$. By Thm. 17 a single finsat test for φ^* takes time polynomial in $|\varphi^*|$ (exponential in $|\varphi_0|$), exponential in the number of the $\forall\exists$ -conjuncts and doubly exponential in the size and width of σ (which are polynomial in $|\varphi_0|$). So, overall, the algorithm is doubly exponential in $|\varphi_0|$. \square

5. Conclusion

Settling an open problem, we established the finite model property of the triguarded fragment (and consequently of guarded formulae preceded by a sequence $\exists^*\forall\exists^*$ of unguarded quantifiers [17]), even providing a doubly exponential upper bound on the model size. Using similar ideas, we settled open problems concerning the guarded and triguarded fragment extended by transitive

guards, providing tight complexity bounds for their finite satisfiability problem in the constant-free case.

While, by definition, GFU and GFU+TG disallow equality (and including unrestricted equality would lead to undecidability [14]), we note that adding equality statements of the form $x = c$ to GFU and equalities guarded by transitive guards to GFU+TG can be done at no computational cost and would not affect our constructions at all. The above additions nicely extend the expressive power of the logics. The first of them allows us, e.g., to express naturally the concept of nominals known from description or hybrid logics, while with the second we can say that a transitive relation is actually an equivalence (cf. [11], Section 5.1), which gives a chance to capture some scenarios from epistemic logics.

As a central open problem, it remains to clarify the decidability status of GF+TG and GFU+TG in the presence of constants. We assume the resulting fragments will still be decidable. Obviously a lower complexity bound for GFU+TG with constants, inherited from GFU, is N2EXPTIME and hence harder than the constant-free case (under standard complexity-theoretic assumptions).

Acknowledgments

E.K. is supported by Polish National Science Centre grant No 2016/21/B/ST6/01444. S.R. is supported by the European Research Council through the ERC Consolidator Grant 771779 (DeciGUT).

References

- [1] Hajnal Andr ka, Johan F. A. K. van Benthem, and Istv n N meti. Modal languages and bounded fragments of predicate logic. *J. of Philosophical Logic*, 27(3):217–274, 1998.
- [2] Vince B r ny, Georg Gottlob, and Martin Otto. Querying the guarded fragment. *Logical Methods in Computer Science*, 10(2), 2014.
- [3] Pierre Bourhis, Michael Morak, and Andreas Pieris. Making cross products and guarded ontology languages compatible. In *Proc. of IJCAI 2017*, 2017.
- [4] Ashok K. Chandra, Dexter Kozen, and Larry J. Stockmeyer. Alternation. *J. ACM*, 28(1):114–133, 1981.
- [5] Harald Ganzinger, Christoph Meyer, and Margus Veanes. The two-variable guarded fragment with transitive relations. In *14th Annual IEEE Symposium on Logic in Computer Science, LICS 1999*, pages 24–34, 1999.
- [6] Erich Gr del. On the restraining power of guards. *J. Symb. Log.*, 64(4):1719–1742, 1999.
- [7] Erich Gr del. On the restraining power of guards. *J. Symb. Log.*, 64(4):1719–1742, 1999.
- [8] Erich Gr del, Phokion Kolaitis, and Moshe Y. Vardi. On the decision problem for two-variable first-order logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.
- [9] Erich Gr del, Phokion G. Kolaitis, and Moshe Y. Vardi. On the decision problem for two-variable first-order logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.
- [10] Yevgeny Kazakov. *Saturation-Based Decision Procedures for Extensions of the Guarded Fragment*. PhD thesis, Universit t des Saarlandes, Saarbr cken, Germany, March 2006.

- [11] Emanuel Kieronski and Adam Malinowski. The triguarded fragment with transitivity. In *Logic for Programming, Artificial Intelligence and Reasoning 2020*, volume 73 of *EPiC*, pages 334–353. EasyChair, 2020.
- [12] Emanuel Kieronski and Lidia Tendera. Finite satisfiability of the two-variable guarded fragment with transitive guards and related variants. *ACM Trans. Comput. Logic*, 19(2):8:1–8:34, 2018.
- [13] Michael Mortimer. On languages with two variables. *Math. Log. Q.*, 21(1):135–140, 1975.
- [14] Sebastian Rudolph and Mantas Šimkus. The triguarded fragment of first-order logic. In *LPAR*, volume 57 of *EPiC Series in Computing*, pages 604–619, 2018.
- [15] Dana Scott. A decision method for validity of sentences in two variables. *Journal of Symbolic Logic*, 27(377):74, 1962.
- [16] Wiesław Szwaś and Lidia Tendera. The guarded fragment with transitive guards. *Annals of Pure and Applied Logic*, 128:227–276, 2004.
- [17] Balder ten Cate and Massimo Franceschet. Guarded fragments with constants. *Journal of Logic, Language and Information*, 14(3):281–288, 2005.

Appendix A.

Comments on the external procedures

Thm. 4. Inspecting the proof of Thm. 1.2 in [2], we see that the size of a minimal finite model of a normal form formula over a signature σ can be bounded by $|\mathfrak{J}|^{\mathcal{O}(w)}$ for some structure \mathfrak{J} whose size is bounded by the number of guarded atomic types over σ , where w is the width of σ .

The number of atomic k -types over σ is $2^{\mathcal{O}(r(k+u)^w)}$, where r is the number of relation symbols and u the number of constants in σ . Of course, every guarded type is a k -type for some $k \leq w$, so the number of guarded types is $w \cdot 2^{\mathcal{O}(r(w+u)^w)}$, which is $2^{\mathcal{O}(r(w+u)^w)}$. Thus the size of model is $2^{\mathcal{O}(r(w+u)^{\mathcal{O}(w^2)})}$. As each of r, w, u is bounded by $|\sigma|$ the claim follows.

Thm. 5. The algorithm from [7] (page 1731) is an alternating algorithm. The algorithm stores two guarded types and a counter counting up the the total number possible guarded types. That is it needs $2^{\mathcal{O}(r(w+u)^w)}$ space. Using the classical simulation of alternating Turing machines by deterministic ones from [4] we get an algorithm working in time $2^{\mathcal{O}(r(w+u)^w)} \cdot \mathcal{O}(n)$, where n is the length of the input formula.

Thm. 6. Let φ be a normal form formula in $\text{GF}^2 + \text{TG}$ over a signature σ . Denote L the number of 1-types over σ ($L = 2^{|\sigma|}$), h the number of the $\forall\exists$ -conjuncts of φ and k the number of transitive relations in σ .

In [12] an important role is played by the parameter M_φ . It is defined on page 14 as $M_\varphi = 3L|\varphi|^3$, but a closer inspection of the proof of Lemma 3.5 (iii), where the role of M_φ is revealed, shows that it is sufficient to take $M_\varphi = 3Lh^3$.

Looking at page 27 of [12] we see that the domain of the small finite model for φ which is constructed there is of size bounded by

$$(2L + 1) \cdot 4 \cdot h \cdot k \cdot T \cdot F$$

where T is the number of the so-called *enriched- M_φ -counting types* and F is a bound on the size of small models for the ‘symmetric’ part of φ .

T is bounded by $2^L \times 2^L \times L^{M_\varphi+1}$, as each enriched- M_φ -counting type is determined by two subsets \mathcal{A} and \mathcal{B} of 1-types and a function which for each 1-type returns a number from the range $\{0, \dots, M_\varphi\}$ (see Def. 6.1). Substituting $3Lh$ for M_φ we see that T is bounded exponentially in L (so, doubly exponentially in $|\sigma|$) and h .

Concerning F , it is actually a bound on the size of the structure constructed in the proof of Lemma 5.7, that is $3h(XL)^k$, where X is the maximal value of a variable in some minimal solution for the system of linear inequalities constructed on page 20. By Lemma 5.6 $X \leq N \cdot (NM_\varphi)^{2N+1}$, for N being the number of inequalities in the system, which is bounded by $3L$. Summarizing, F is not greater than $3h(3L^2 \cdot (9L^2h^3)^{6L+1})^k$.

Gathering the above estimations we get that the size of the finite model constructed is exponential in the number of 1-types (and thus doubly exponential in the size of the signature), in h , and in k . Taking into account that $k \leq |\sigma|$ the claim follows.

Thm. 7. Again, the algorithm from [12] (page 30) is alternating. Assume that the notation is as in the previous paragraph. What the algorithm stores is:

- a counter counting up to kT ,
- a collection of at most kL enriched- M_φ -counting types

That is, the total space required is polynomial in the length of the input formula and exponential in L and h . Simulating alternating Turing machines by deterministic ones as previously we get an algorithm for normal form formulas working in time bounded polynomially in the size of the input and exponentially in the number of 1-types (doubly exponentially in the size of the signature) and the $\forall\exists$ -conjuncts.

Appendix B.

Details on the FMP for GFU with constants

Given a structure \mathfrak{A} interpreting a signature σ consisting of relation symbols and constants we call the subset $\hat{A} \subseteq A$ consisting of all the elements interpreting the constants of σ the *named part* of \mathfrak{A} . The *unnamed part* is defined as $\check{A} := A \setminus \hat{A}$. A 1-type is *named* if contains $x_1 = c$ for some constant c and *unnamed* otherwise.

We first redefine the notions of disjoint unions and doublings of structures. Let $(\mathfrak{A}_i)_{i \in \mathcal{I}}$ be a family of σ -structures having disjoint unnamed parts and sharing the same named part; call such a family *harmonized*. Their *harmonized union* is the structure \mathfrak{A} with domain $A = \hat{A}_t \cup \bigcup_{i \in \mathcal{I}} \check{A}_i$, where t is chosen as any element of \mathcal{I} , such that for all i we have that $\mathfrak{A} \upharpoonright \hat{A}_t \cup \check{A}_i$ is isomorphic to \mathfrak{A}_i , and for any tuple \bar{a} containing elements from at least two different \check{A}_i and any relation symbol $P \in \sigma$ of arity $|\bar{a}|$ we have $\mathfrak{A} \models \neg P(\bar{a})$.

Lemma 21. Let φ be a GF normal form formula over a signature consisting of relation symbols and constants.

The harmonized union of any harmonized family of models of φ is also its model.

Let \mathfrak{A}_- be a σ -structure. Its *harmonized doubling* is the structure \mathfrak{A} with domain $A := \hat{A}_- \times \{0\} \cup \hat{A}_- \times \{0, 1\}$ in which for each $P \in \sigma$ we have $\mathfrak{A} \models P[(a_1, \ell_1), \dots, (a_k, \ell_k)]$ iff $\mathfrak{A}_- \models P[a_1, \dots, a_k]$ for all $a_i \in \hat{A}_-$ and $\ell_i \in \{0, 1\}$ if $a_i \in \hat{A}_-$ and $\ell_i = 0$ if $a_i \in \hat{A}_-$.

Lemma 22. Let φ be a normal form GF or GFU formula which does not use equality outside guards and let \mathfrak{A}_- be a model of φ . Then its harmonized doubling \mathfrak{A} is still a model of φ .

Let us now fix a GFU sentence φ in normal form, without equality over a signature σ consisting of relation symbols and constants and let \mathfrak{A} be a U-biquitous model of φ . Our goal is to build a finite U-biquitous model \mathfrak{A}' of φ .

B.1. Preparing building blocks

We construct φ^* precisely as in the case without constants. We remark that this time each $\alpha \in \mathbf{\alpha}$ fully specifies the substructure consisting of the given element and the named part of the structure.

It is clear that φ^* , treated as a GF-formula, is satisfiable. In fact, \mathfrak{A} is its model. Thus, by the finite model property for GF, it also has a finite (not necessarily U-biquitous) model. We take such a finite model $\mathfrak{C}_- \models \varphi^*$, and let \mathfrak{C} be its harmonized doubling. As φ^* does not use equality, by Lemma 22 we have that $\mathfrak{C} \models \varphi^*$.

Recalling the definition of indistinguishable elements we adapt Claim 8 to our current setting.

Claim 23. For any pair of **unnamed** 1-types $\alpha, \alpha' \in \mathbf{\alpha}$ there is a pair of their distinct realizations a, a' in \mathfrak{C} such that $\mathfrak{C} \models U[a, a'] \wedge U[a', a]$. Moreover, if $\alpha = \alpha'$, then we even find indistinguishable a, a' with that property.

As previously we build yet another model $\mathfrak{B} \models \varphi^*$, as the harmonized union of five copies of \mathfrak{C} . Letting $K = |\check{C}|$, we assume that the unnamed part of \mathfrak{B} is $B := \{1, \dots, 5K\}$; and that for $m = 0, \dots, 4$ the structure on $\hat{B} \cup \{mK + 1, \dots, mK + K\}$ is isomorphic to \mathfrak{C} .

B.2. U-saturation

We now build a finite sequence of finite structures $\mathfrak{A}_0, \mathfrak{A}_1, \dots, \mathfrak{A}_f$, each of them being a model of φ^* and the last of them being a desired U-biquitous model \mathfrak{A}' of φ^* (and thus also of φ).

The domains of all these structures will be identical.

$$A_i = \hat{B} \cup (\check{B} \times \{1, \dots, 5K\} \times \{1, \dots, 5K\}).$$

We will view each of the \mathfrak{A}_i as a $5K \times 5K$ table containing unnamed parts plus the shared named part.

The initial structure \mathfrak{A}_0 is defined as the harmonized union of $(5K)^2$ copies of \mathfrak{B} . Namely, for each $k, \ell \in \{1, \dots, 5K\}$ we make $\mathfrak{A}_0 \upharpoonright \hat{B} \cup (\check{B} \times \{k\} \times \{\ell\})$ isomorphic to \mathfrak{B} (via the isomorphism working as the identity on \hat{B}

and as the natural projection $(b, k, \ell) \mapsto b$ on the unnamed elements). By Lemma 21 we have that $\mathfrak{A}_0 \models \varphi^*$.

Some notation. We adapt our notation. For each k, ℓ we denote by $\mathfrak{B}_i^{k, \ell}$ the structure consisting of the common named part and the unnamed part in the cell (k, ℓ) of \mathfrak{A}_i , that is the structure $\mathfrak{A}_i \upharpoonright \hat{B} \cup (\check{B} \times \{k\} \times \{\ell\})$. We recall that $\mathfrak{B}_0^{k, \ell}$ is isomorphic to \mathfrak{B} . Further, for $m = 0, \dots, 4$, we denote by $\mathfrak{C}_i^{k, \ell, m}$ the structure $\mathfrak{B}_i^{k, \ell} \upharpoonright \hat{B} \cup \{mK + 1, \dots, mK + K\} \times \{k\} \times \{\ell\}$. We recall that each $\mathfrak{C}_0^{k, \ell, m}$ is isomorphic to \mathfrak{C} .

Entry elements and their use. This time only members of the unnamed parts are entry elements. For any $1 \leq k, \ell \leq 5K$, let $\alpha^k = \text{tp}^{\mathfrak{B}}(k)$ and $\alpha^\ell = \text{tp}^{\mathfrak{B}}(\ell)$. Note that α^k and α^ℓ are unnamed.

For each such pair k, ℓ we now choose a pair of *entry elements* for each of the five structures with unnamed parts in the cell (k, ℓ) of \mathfrak{A}_0 , that is for the structures $\mathfrak{C}_0^{k, \ell, m}$ ($m = 0, 1, \dots, 4$).

By Claim 23, there are distinct elements $e_1, e_2 \in \mathfrak{C}$ such that $\mathfrak{C} \models \alpha^k[e_1] \wedge \alpha^\ell[e_2] \wedge U[e_1, e_2] \wedge U[e_2, e_1]$ and if $\alpha^k = \alpha^\ell$ then e_1 and e_2 are indistinguishable in \mathfrak{C} . We choose the entry elements $e_1^{k, \ell, m}, e_2^{k, \ell, m}$ to $\mathfrak{C}_0^{k, \ell, m}$ to be the corresponding copies of e_1 and e_2 in each of $\mathfrak{C}_0^{k, \ell, m}$.

By $+\mathfrak{C}_i^{k, \ell, m}$ we denote the structure with domain $C_i^{k, \ell, m} \cup \{b_1, b_2\}$ for some fresh unnamed elements b_1, b_2 such that $+\mathfrak{C}_i^{k, \ell, m} \upharpoonright C_i^{k, \ell, m} = \mathfrak{C}_i^{k, \ell, m}$ and for each $P \in \sigma$ and each tuple \bar{a} containing at least one of b_1, b_2 we have $+\mathfrak{C}_i^{k, \ell, m} \models P[\bar{a}]$ iff $\mathfrak{C}_0^{k, \ell, m} \models P[\mathfrak{h}(\bar{a})]$, where \mathfrak{h} is the function defined as $\mathfrak{h}(b_1) = e_1^{k, \ell, m}$, $\mathfrak{h}(b_2) = e_2^{k, \ell, m}$ and $\mathfrak{h}(a) = a$ for $a \in C_i^{k, \ell, m}$. In particular $+\mathfrak{C}_i^{k, \ell, m} \models \alpha^k[b_1] \wedge \alpha^\ell[b_2] \wedge U[b_1, b_2] \wedge U[b_2, b_1]$.

From \mathfrak{A}_i to \mathfrak{A}_{i+1} . Assume now that the structure \mathfrak{A}_i has been defined, for some $i \geq 0$, $\mathfrak{A}_i \models \varphi^*$. If \mathfrak{A}_i is U-biquitous then we are done. Otherwise let b_1, b_2 be a pair of elements in A_i such that $\mathfrak{A}_i \models \neg U[b_1, b_2]$. Note that b_1, b_2 must be unnamed. Indeed, as in the case without constants, our process does not modify the types of the guarded tuples. In particular, in each of the \mathfrak{A}_i the 1-types are retained from \mathfrak{A}_0 , where they are copied from the model \mathfrak{C} of φ^* . By the conjunct (2) of φ^* all the 1-types in \mathfrak{A}_0 belong to $\mathbf{\alpha}$, the set of 1-types realized in the U-biquitous structure \mathfrak{A} . Thus, if, say, b_1 would interpret a constant c then, as the 1-type of b_2 must contain $U(x, c) \wedge U(c, x)$ (since this 1-type belongs to $\mathbf{\alpha}$), we would have that $\mathfrak{A}_i \models U[b_1, b_2] \wedge U[b_2, b_1]$.

For $s = 1, 2$ let k_s, l_s, n_s be such that b_s is the n_s -th element of the unnamed part of $\mathfrak{B}_i^{k_s, l_s}$. Let us choose $t \in \{0, \dots, 4\}$ such that $\mathfrak{C}_i^{n_1, n_2, t}$ does not contain the k_1 -th, l_1 -th, k_2 -th or l_2 -th element of the unnamed part of $\mathfrak{B}_i^{n_1, n_2}$. Such a t must exist by the pigeon hole principle. We make the structure $\mathfrak{A}_{i+1} \upharpoonright C_i^{n_1, n_2, t} \cup \{b_1, b_2\}$ isomorphic to $+\mathfrak{C}_i^{n_1, n_2, t}$. The rest of the structure \mathfrak{A}_i remains untouched.

Correctness. The correctness of the construction can be proved in the same vain as in the case without constants. In particular Claims 9, 10, 11 and 12 remain true with literally no changes and their proofs require only routine adjustments concerning the division of the domains into

named/unnamed parts. That is the whole process eventually ends in U-biquitous model of φ^* .