

EQUIVALENCE CLOSURE IN THE TWO-VARIABLE GUARDED FRAGMENT*

EMANUEL KIEROŃSKI [†], IAN PRATT-HARTMANN [‡], AND LIDIA TENDERA [§]

Abstract. We consider the satisfiability and finite satisfiability problems for the extension of the two-variable guarded fragment in which an equivalence closure operator can be applied to two distinguished binary predicates. We show that the satisfiability and finite satisfiability problems for this logic are 2-EXPTIME-complete. This contrasts with an earlier result that the corresponding problems for the full two-variable logic with equivalence closures of two binary predicates are 2-NEXPTIME-complete.

Key words. computational complexity, decidability, guarded fragment

AMS subject classifications. 03B25, 03B70

1. Introduction. The two-variable fragment of first-order logic, FO^2 , and the two-variable guarded fragment, GF^2 , are widely investigated formalisms whose study is motivated by their close connections to modal, description and temporal logics. It is well-known that FO^2 enjoys the finite model property [13], and that its satisfiability (= finite satisfiability) problem is NExpTime-complete [4]. Since GF^2 is contained in FO^2 , it too has the finite model property; however, its satisfiability problem is slightly easier, namely EXPTIME-complete [5].

It is impossible, in FO^2 , to write a formula expressing the condition that a given binary relation is an equivalence (i.e. is reflexive, symmetric and transitive); and the question therefore arises as to whether such a facility could be added at reasonable computational cost. In a series of papers [10, 12, 9], various extensions of FO^2 were investigated in which certain distinguished binary predicates are declared to be equivalences, or in which an operation of equivalence closure can be applied to these predicates. (The equivalence closure of a binary relation is the smallest equivalence that includes it.) Denote by EQ_k^2 the extension of FO^2 in which k distinguished binary predicates are interpreted as equivalences, and by EC_k^2 the extension of FO^2 in which we can take the equivalence closure of any of k distinguished binary predicates. The following is known: (i) EQ_1^2 and EC_1^2 retain the finite model property, and their satisfiability (= finite satisfiability) problems are NEXPTIME-complete; (ii) EQ_2^2 and EC_2^2 lack the finite model property, and their satisfiability and finite satisfiability problems are 2-NEXPTIME-complete; (iii) when $k \geq 3$, the satisfiability and finite satisfiability problems for EQ_k^2 and EC_k^2 are undecidable.

Turning now to GF^2 , denote by GFEQ_k^2 the extension of GF^2 in which k distinguished binary predicates are interpreted as equivalences, and by GFEC_k^2 the extension of GF^2 in which we can take the equivalence closure of any of k distinguished binary predicates. It was shown in [7] that the satisfiability and finite satisfiability problems for GFEQ_1^2 are both NEXPTIME-hard, whence, by the above-mentioned results on EQ_1^2 , both these problems are NEXPTIME-complete. In addition, the undecidability results for EQ_k^2 and EC_k^2 ($k \geq 3$) referred to in the previous paragraph employ only guarded formulas, whence satisfiability and finite satisfiability for GFEQ_k^2

*This paper is supported by the Polish National Science Centre grant DEC-2013/09/B/ST6/01535 and EPSRC grant number EP/K017438/1.

[†]Institute of Computer Science, University of Wrocław, Poland

[‡]School of Computer Science, University of Manchester, United Kingdom

[§]Institute of Mathematics and Informatics, Opole University, Poland

and GFEC_k^2 are undecidable when $k \geq 3$. This leaves only the case $k = 2$. It is known that GFEQ_2^2 and GFEC_2^2 lack the finite model property (see, e.g., the example in Section 2 of [9], and observe that all formulas there are indeed guarded). The satisfiability problem for GFEQ_2^2 was shown to be 2-EXPTIME-complete in [7]. The corresponding finite satisfiability problem was not addressed there; and while the method employed to establish the 2-EXPTIME lower-bound can easily be adapted to the finite case, the proof of the upper bound, which was based on tree-unravelling methods, fails. In this paper we solve the remaining open problems by establishing a 2-EXPTIME upper bound on the satisfiability and finite satisfiability problems for GFEC_2^2 . It follows that these problems are 2-EXPTIME-complete, as is the finite satisfiability problem for GFEQ_2^2 .

To establish these results, we adopt the same strategy as that employed in [9] for the logics EQ_2^2 and EC_2^2 ; and we briefly review that strategy now. For concreteness, let φ be an EQ_2^2 -formula whose finite satisfiability we are trying to establish. We may assume that φ features two distinguished binary predicates, which must be interpreted as equivalences. Denote these equivalences (in some putative finite model) by r_1 and r_2 : note that their coarsest common refinement, $r_1 \cap r_2$, is also an equivalence. We call the equivalence classes of $r_1 \cap r_2$ *intersections*, and we show that the intersections arising in any finite model of φ can, without loss of generality, be assumed to have cardinality exponentially bounded as a function of the size of φ . In particular, every r_1 -class, and also every r_2 -class, is the union of some set of such “small intersections”; and any given r_1 -class and r_2 -class are either disjoint, or have exactly one common intersection. This decomposition into equivalence classes allows us to picture any finite model of φ as an edge-coloured, bipartite graph: the r_1 -classes are the left-hand vertices; the r_2 -classes are the right-hand vertices; and two vertices are joined by an edge just in case they share an intersection, with the colour of that edge being the isomorphism type of the intersection concerned. The formula φ imposes various constraints on the types of intersections that may arise, and on how intersections may be organized into r_1 - and r_2 classes. These constraints translate to conditions on the induced bipartite graph of equivalence classes, so that the original (finite) satisfiability problem becomes the problem of determining the existence of an edge-coloured bipartite graph satisfying certain conditions on the local configurations it realizes. These conditions, it transpires, can be expressed in purely numerical terms—specifically, as a system of linear inequalities over integer variables. Thus, the problem of determining the finite satisfiability of φ is translated into an instance of integer linear programming. The instance in question can be shown to be doubly exponential in the size of φ ; and using the well-known fact that integer linear programming feasibility is in NPTIME, the 2-NEXPTIME-upper bound for the finite satisfiability of EQ_2^2 follows. The main technical contribution of the present paper is a proof that, for the *guarded* sub-fragments, GFEQ_2^2 and GFEC_2^2 , we obtain systems of linear inequalities for which solutions may be sought not over integers, but over the reals. The 2-EXPTIME-upper bound for these logics then follows from the well-known fact that (real) linear programming feasibility is in PTIME. In addition, the need to obtain a 2-EXPTIME-upper bound—as opposed to a 2-NEXPTIME-upper as in the case of EC_2^2 —means that we no longer have the luxury of doubly exponentially many non-deterministic choices in the decision procedure. This fact necessitates various changes at a tactical level throughout the proof.

The plan of the paper is as follows. In Sec. 2, we define the logics GFEQ_2^2 and GFEC_2^2 , introducing a ‘Scott-type’ normal form for GFEC_2^2 that allows us to restrict

the nesting of quantifiers to depth two. In Sec. 3 we define a problem concerning the existence of certain bipartite graphs, called BGE^* , and show that both the finite and general versions of this problem are in PTIME. In Sec. 4, we prove some technical lemmas that allow us in Sec. 5 to reduce the (finite) satisfiability problem of a GFEC_2^2 -formula to (finite) BGE^* in deterministic doubly exponential time.

2. Preliminaries.

2.1. Logics, structures and types. We employ standard terminology and notation from model theory throughout this paper (see, e.g. [3]). In particular, we refer to structures using Gothic capital letters, and their domains using the corresponding Roman capitals. We denote by GF^2 the guarded two-variable fragment of first-order logic (with equality), without loss of generality restricting attention to signatures of unary and binary predicates. Formally, GF^2 is the intersection of FO^2 (i.e., the restriction of first-order logic in which only two variables, x and y are available) and the *guarded fragment*, GF [1]. GF is defined as the least set of formulas such that: (i) every atomic formula belongs to GF ; (ii) GF is closed under logical connectives $\neg, \vee, \wedge, \Rightarrow$; and (iii) quantifiers are appropriately relativised by atoms, i.e., denoting by \mathbf{x}, \mathbf{y} tuples of variables, if $\varphi(\mathbf{x}, \mathbf{y})$ is a formula of GF and $\alpha(\mathbf{x}, \mathbf{y})$ is an atomic formula containing all the free variables of φ , then the formulas $\forall \mathbf{y}(\alpha(\mathbf{x}, \mathbf{y}) \Rightarrow \varphi(\mathbf{x}, \mathbf{y}))$ and $\exists \mathbf{y}(\alpha(\mathbf{x}, \mathbf{y}) \wedge \varphi(\mathbf{x}, \mathbf{y}))$ belong to GF . In this context, the atom $\alpha(\mathbf{x}, \mathbf{y})$ is called a *guard*. Equalities $x=x$ or $x=y$ are also allowed as guards.

We denote by GFEC_k^2 the set of GF^2 -formulas over any signature $\tau = \tau_0 \cup \{r_1, \dots, r_k\} \cup \{r_1^\#, \dots, r_k^\#\}$, where τ_0 is an arbitrary set containing unary and binary predicates, and $r_1, \dots, r_k, r_1^\#, \dots, r_k^\#$ are distinguished binary predicates, not present in τ_0 . In the sequel, any signature τ is assumed to be of the above form (for some appropriate value of k). We denote by GFEC_k^2 the set of GFEC_k^2 -formulas in which the predicates $r_1^\#, \dots, r_k^\#$ do not occur.

The semantics for GFEC_k^2 are standard, subject to the restriction that $r_i^\#$ is always interpreted as the *equivalence closure* of r_i . More precisely: we consider only structures \mathfrak{A} in which, for all i ($1 \leq i \leq k$), $(r_i^\#)^\mathfrak{A}$ is the smallest reflexive, symmetric and transitive relation including $r_i^\mathfrak{A}$. Similarly, we require for GFEC_k^2 that r_i is always interpreted as an equivalence. Where a structure is clear from context, we may equivocate between predicates and their extensions, writing, for example, r_i and $r_i^\#$ in place of the technically correct $r_i^\mathfrak{A}$ and $(r_i^\#)^\mathfrak{A}$.

Let \mathfrak{A} be a structure over τ . We say that there is an r_i -edge between a and $a' \in A$ if $\mathfrak{A} \models r_i[a, a']$ or $\mathfrak{A} \models r_i[a', a]$. Distinct elements $a, a' \in A$ are r_i -connected if there exists a sequence $a = a_0, a_1, \dots, a_{k-1}, a_k = a'$ in A such that for all j ($0 \leq j < k$) there is an r_i -edge between a_j and a_{j+1} . Such a sequence is called an r_i -path from a to a' . Thus, $\mathfrak{A} \models r_i^\#[a, a']$ if and only if a and a' are r_i -connected. A subset B of A is called r_i -connected if every pair of distinct elements of B is r_i -connected. Maximal r_i -connected subsets of A are equivalence classes of $r_i^\#$, and are called $r_i^\#$ -classes.

We mostly work with the logic GFEC_2^2 . In any structure \mathfrak{A} , the relation $r_1^\# \cap r_2^\#$ is also an equivalence, and we refer to its equivalence classes, simply, as *intersections*. Thus, an intersection is a maximal set that is both r_1 - and r_2 -connected. When discussing induced substructures, a subtlety arises regarding the interpretation of the closure operations. If $B \subseteq A$, we take it that, in the structure \mathfrak{B} induced by B , the interpretation of $r_i^\#$ is given by simple restriction: $(r_i^\#)^\mathfrak{B} = (r_i^\#)^\mathfrak{A} \cap B^2$. This means that, while $(r_i^\#)^\mathfrak{B}$ is certainly an equivalence including $r_i^\mathfrak{B}$, it may not be the smallest, since, for some $a, a' \in B$, an r_i -path connecting a and a' in \mathfrak{A} may contain elements which are not members of B . (Such a situation may arise even when B

is an intersection.) To reduce notational clutter, we use the (possibly decorated) letter \mathfrak{A} to denote ‘full’ structures in which we are guaranteed that $(r_i^\#)^{\mathfrak{A}}$ is the equivalence closure of $r_i^{\mathfrak{A}}$. For structures denoted by other letters, $\mathfrak{B}, \mathfrak{C}, \dots$ (again, possibly decorated), no such guarantee applies. Typically, but not always, these latter structures will be induced substructures.

An (atomic) 1-*type* (over a given signature) is a maximal satisfiable set of atoms or negated atoms with free variable x . Similarly, an (atomic) 2-*type* is a maximal satisfiable set of atoms and negated atoms with free variables x, y . Note that the numbers of 1-types and 2-types are bounded exponentially in the size of the signature. We often identify a type with the conjunction of its elements.

For a given τ -structure \mathfrak{A} , we denote by $\text{tp}^{\mathfrak{A}}(a)$ the 1-type *realized* by a , i.e., the 1-type α such that $\mathfrak{A} \models \alpha[a]$. Similarly, for distinct $a, b \in A$, we denote by $\text{tp}^{\mathfrak{A}}(a, b)$ the 2-type *realized* by the pair a, b , i.e., the 2-type β such that $\mathfrak{A} \models \beta[a, b]$.

If φ is a formula, we write $\|\varphi\|$ to denote the number of symbols in φ .

2.2. Normal form and small intersections. In the context of FO^2 , it is often convenient to work with formulas in so-called Scott normal form. The following definition adapts this notion to the setting of GFEC_2^2 .

DEFINITION 2.1. *A normal guard is any of the formulas $r_1^\#(x, y) \wedge r_2^\#(x, y)$, $r_1^\#(x, y) \wedge \neg r_2^\#(x, y)$, $\neg r_1^\#(x, y) \wedge r_2^\#(x, y)$, or $\epsilon(x, y) \wedge \neg r_1^\#(x, y) \wedge \neg r_2^\#(x, y)$, where $\epsilon(x, y)$ is a τ_0 -atom. A GFEC_2^2 -sentence φ is in normal form if it is a conjunction of formulas of the forms:*

$$\begin{aligned} \exists x \psi(x) & \quad (\exists) \\ \forall x \psi(x) & \quad (\forall) \\ \forall xy (\eta(x, y) \rightarrow (x \neq y \rightarrow \psi(x, y))) & \quad (\forall\forall) \\ \forall x (\gamma(x) \rightarrow \exists y (\eta(x, y) \wedge x \neq y \wedge \psi(x, y))) & \quad (\forall\exists) \end{aligned}$$

where $\gamma(x)$ is a τ_0 -atom, $\eta(x, y)$ is a normal guard, and $\psi(x), \psi(x, y)$ are quantifier-free formulas not using $r_1^\#$ and $r_2^\#$, with free variables as indicated. Where a normal-form GFEC_2^2 -formula φ is given, we refer to these four types of formulas as its \forall -, \exists -, $\forall\forall$ - and $\forall\exists$ -conjuncts, respectively.

The forms (\forall) and (\exists) are not, strictly speaking, in GFEC_2^2 ; but they can easily be transformed into logically equivalent formulas which are. Thus, for example, $\forall x \psi(x)$ is equivalent to $\forall x ((x = x) \rightarrow \psi(x))$, and similarly for the case (\exists) .

For technical reasons we assume that τ_0 contains unary symbols $K_1^{\text{in}}, K_1^{\text{out}}, K_2^{\text{in}}, K_2^{\text{out}}$, and we will always assume, without loss of generality, that a normal form φ contains the following conjuncts (for $i = 1, 2$): $\forall x (K_i^{\text{in}}(x) \rightarrow \exists y (r_i(y, x) \wedge \neg r_{3-i}^\#(x, y)))$, $\forall x (K_i^{\text{out}}(x) \rightarrow \exists y (r_i(x, y) \wedge \neg r_{3-i}^\#(x, y)))$ and $\forall xy ((r_i(x, y) \wedge \neg r_{3-i}^\#(x, y)) \rightarrow (K_i^{\text{out}}(x) \wedge K_i^{\text{in}}(y)))$. This guarantees that in every model of φ , precisely the elements containing an r_i -edge leading into/outside their intersection are marked by $K_i^{\text{in/out}}$ predicates. It is easy to see that, for a GFEC_2^2 -formula not featuring the predicates $K_i^{\text{in/out}}$, adding these conjuncts results in a formula satisfiable over the same domains.

The following lemma justifies the usefulness of the introduced normal form.

LEMMA 2.2. *The (finite) satisfiability problem for GFEC_2^2 can be reduced to the (finite) satisfiability problem for disjunctions of an exponential number of linearly bounded GFEC_2^2 -formulas in normal form. This can be proved using some standard techniques. See [14] (Lemma 2) for the proof of a similar result, which easily adapts*

to the present case. Since we are going to show that (finite) satisfiability is in 2-EXPTIME, it is enough to consider formulas in normal form.

It is sometimes useful to group together various conjuncts of a normal-form formula.

DEFINITION 2.3. *If φ is as in Definition 2.1, let us write:*

$$\begin{aligned}\varphi_{univ} &:= \bigwedge \{ \psi \mid \psi \text{ a } \forall\text{-conjunct or } \forall\forall\text{-conjunct of } \varphi \} \\ \varphi_{12} &:= \varphi_{univ} \wedge \bigwedge \{ \psi \mid \psi \text{ a } \forall\exists\text{-conjunct of } \varphi \text{ with } \eta = r_1^\#(x, y) \wedge r_2^\#(x, y) \} \\ \varphi_1 &:= \varphi_{univ} \wedge \bigwedge \{ \psi \mid \psi \text{ a } \forall\exists\text{-conjunct of } \varphi \text{ with } \eta = r_1^\#(x, y) \wedge \neg r_2^\#(x, y) \} \\ \varphi_2 &:= \varphi_{univ} \wedge \bigwedge \{ \psi \mid \psi \text{ a } \forall\exists\text{-conjunct of } \varphi \text{ with } \eta = \neg r_1^\#(x, y) \wedge r_2^\#(x, y) \} \\ \varphi_{free} &:= \varphi_{univ} \wedge \bigwedge \{ \psi \mid \psi \text{ a } \forall\exists\text{-conjunct of } \varphi \text{ with } \eta = \epsilon(x, y) \wedge \neg r_1^\#(x, y) \wedge \neg r_2^\#(x, y) \}.\end{aligned}$$

FACT 1. *If φ is a normal form GFEC₂²-formula, and \mathfrak{A} a τ -structure, then $\mathfrak{A} \models \varphi$ if and only if the following three conditions hold:*

- (i) $\mathfrak{A} \models \bigwedge \{ \psi \mid \psi \text{ an } \exists\text{-conjunct of } \varphi \}$,
- (ii) for each intersection \mathfrak{I} of \mathfrak{A} , $\mathfrak{I} \models \varphi_{12}$,
- (iii) for each $r_i^\#$ -class \mathfrak{D} of \mathfrak{A} , $\mathfrak{D} \models \varphi_i$ (for $i = 1, 2$),
- (iv) $\mathfrak{A} \models \varphi_{free}$.

Proof. Straightforward. \square

It was shown in [8] (Lemma 4.2) that, when considering (finite) satisfiability of EC₂²-formulas, one can restrict attention to models with exponentially bounded intersections. Since this result subsumes the case of GFEC₂², we have:

LEMMA 2.4. *Let φ be a satisfiable GFEC₂²-formula in normal form over a signature τ . Then there exists a model of φ in which the size of each intersection is bounded by $\mathfrak{f}(|\varphi|)$, for a fixed exponential function \mathfrak{f} .*

3. A problem concerning bipartite graphs.

3.1. Definition of BGE*. In the sequel, we denote by \mathbb{N} the set of non-negative integers, and by \mathbb{N}^* the set $\mathbb{N} \cup \{\aleph_0\}$, where $n < \aleph_0$ for all $n \in \mathbb{N}$. In [9], a family of problems was defined concerning the existence of edge-coloured bipartite graphs satisfying various collections of conditions. The simplest of these, BGE (*bipartite graph existence*), is shown to be in PTIME. This paper employs a variant of BGE, which we denote by BGE*.

Let Δ be a finite, non-empty set. A Δ -graph is a triple $H = (U, V, \mathbf{E}_\Delta)$, where U, V are disjoint sets of cardinality at most \aleph_0 , and \mathbf{E}_Δ is a collection of pairwise disjoint subsets $E_\delta \subseteq U \times V$, indexed by the elements of Δ . We call the elements of $W = U \cup V$ *vertices*, and the elements of E_δ , δ -edges. It helps to think of \mathbf{E}_Δ as the result of colouring the edges of the bipartite graph (U, V, E) , where $E = \bigcup_{\delta \in \Delta} E_\delta$, using the colours in Δ . For any $w \in W$, we define the function $\text{ord}_w^H : \Delta \rightarrow \mathbb{N}^*$, called the *order* of w , by

$$\begin{aligned}\text{ord}_u^H(\delta) &= |\{v \in W : (u, v) \in E_\delta\}| && (u \in U) \\ \text{ord}_v^H(\delta) &= |\{u \in W : (u, v) \in E_\delta\}| && (v \in V).\end{aligned}$$

Thus, ord_w^H tells us, for each colour δ , how many δ -edges w is incident to in H . Obviously, if H is finite, the values of ord_w^H all lie in \mathbb{N} . When constructing Δ -graphs,

it is sometimes more convenient to employ a slightly more general notion. We define a Δ -multigraph in the same way as a Δ -graph, except that the E_δ 's are now *multi*-sets, and are not required to be disjoint. Thus, in a Δ -multigraph $(U, V, \mathbf{E}_\Delta)$, a pair of nodes $u \in U$ and $v \in V$ may be joined by any number of edges of any colours. The order-functions are defined in the obvious way, recording the total number of edges of each colour to which the node in question is incident.

If M is a positive integer, we write \hat{M} to denote the set $\{=0, =1, \geq 0, \geq 1, \dots, \geq M\}$. We use the notation $=k$ and $\geq k$, where k is a variable, to denote elements of this set; formally, however, these elements have no internal structure. Note that $=k$ makes sense only for $k = 0$ or $k = 1$.

DEFINITION 3.1. *Let M be a positive integer and Δ a finite nonempty set. Given functions $f : \Delta \rightarrow \mathbb{N}^*$ and $p : \Delta \rightarrow \hat{M}$, we say that f realizes p if, for all $\delta \in \Delta$: $p(\delta) = =k$ implies $f(\delta) = k$, and $p(\delta) = \geq m$ implies $f(\delta) \geq m$.*

DEFINITION 3.2. *Let $p : \Delta \rightarrow \hat{M}$, where M and Δ are as in Definition 3.1. We denote by \bar{p} the function $\bar{p} : \Delta \rightarrow \mathbb{N}$ obtained by erasing decorations from the results of p ; that is: for each $\delta \in \Delta$, $\bar{p}(\delta) = k$ if and only if $p(\delta) = =k$ or $p(\delta) = \geq k$.*

A function $p : \Delta \rightarrow \hat{M}$, may be thought of as expressing a collection of conditions on functions $f : \Delta \rightarrow \mathbb{N}$. To say that $f : \Delta \rightarrow \mathbb{N}$ realizes p is simply to say that f satisfies the conditions in question. Note that, trivially, \bar{p} realises p . We now proceed to define the problem BGE^* .

DEFINITION 3.3. *A BGE^* -instance is a quintuple $\mathcal{P} = (\Delta, \Delta_0, M, P, Q)$, where Δ is a finite, non-empty set, $\Delta_0 \subseteq \Delta$, M is a positive integer, and P and Q are sets of functions $\Delta \rightarrow \hat{M}$. A solution of \mathcal{P} is a Δ -graph $H = (U, V, \mathbf{E}_\Delta)$ such that:*

$$\text{for all } \delta \in \Delta_0, E_\delta \text{ is non-empty}; \quad (\text{G1})$$

$$\text{for all } u \in U, \text{ord}_u^H \text{ realizes some function from } P; \quad (\text{G2})$$

$$\text{for all } v \in V, \text{ord}_v^H \text{ realizes some function from } Q. \quad (\text{G3})$$

The problem BGE^* is as follows:

GIVEN: a BGE^* -instance \mathcal{P} .

OUTPUT: Yes, if \mathcal{P} has a solution; No, otherwise.

The problem *finite* BGE^* is as follows:

GIVEN: a BGE^* -instance \mathcal{P} .

OUTPUT: Yes, if \mathcal{P} has a finite solution; No, otherwise.

That is: suppose we are given a set of colours Δ , a distinguished subset $\Delta_0 \subseteq \Delta$ and sets of functions P, Q mapping Δ to the set of decorated integers \hat{M} . We wish to know whether there exists a (finite) Δ -graph $(U, V, \mathbf{E}_\Delta)$ in which the vertices in U have only order-functions realising functions from P , the vertices in V have only order-functions realising functions in Q , and each of the colours in Δ_0 is represented by at least one edge.

In the next two subsections we show that BGE^* and finite BGE^* can be solved in polynomial time.

3.2. Complexity of finite BGE^* . By a *linear Diophantine clause* we mean a disjunction of linear Diophantine equations and inequalities. (We allow clauses to have just one disjunct). A *solution* of a system of linear Diophantine clauses is an

assignment of non-negative integers to its variables making all its clauses true. If such a solution exists, the system of linear Diophantine clauses in question is said to be *satisfiable*. We proceed to reduce finite BGE* to the satisfiability problem for systems of linear Diophantine clauses of a particular form.

To understand the reduction, let $\mathcal{P} = (\Delta, \Delta_0, M, P, Q)$ be a BGE*-instance, and let us suppose that \mathcal{P} has a finite solution $H = (U, V, \mathbf{E}_\Delta)$. Order the sets of functions P and Q arbitrarily. For every $p \in P$, let U_p be the set of vertices $u \in U$ such that p is the first function in P realized by ord_u^H , and, for every $q \in Q$, let V_q be the set of vertices $v \in V$ such that q is the first function in Q realized by ord_v^H . Now set

$$x_p = |U_p| \text{ for all } p \in P \quad y_q = |V_q| \text{ for all } q \in Q.$$

Employing the notation of Definition 3.2, it is obvious that, for all $\delta \in \Delta$:

$$\sum_{p \in P} \bar{p}(\delta) x_p \leq |E_\delta| \quad \sum_{q \in Q} \bar{q}(\delta) y_q \leq |E_\delta|. \quad (3.1)$$

Now define, for all $\delta \in \Delta$ and $p \in P$:

$$c_p^\delta = \begin{cases} 1 & \text{if } p(\delta) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

and similarly for c_q^δ ($q \in Q$). Since H is a solution of \mathcal{P} , E_δ is non-empty for all $\delta \in \Delta_0$. Hence, the following sets of inequalities hold:

$$\left\{ \sum_{p \in P} c_p^\delta x_p > 0 \mid \delta \in \Delta_0 \right\} \quad (3.2)$$

$$\left\{ \sum_{q \in Q} c_q^\delta y_q > 0 \mid \delta \in \Delta_0 \right\}. \quad (3.3)$$

Now define, for all $\delta \in \Delta$ and $p \in P$:

$$d_p^\delta = \begin{cases} 1 & \text{if } p(\delta) = \geq k \text{ for some } k \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

and similarly for d_q^δ ($q \in Q$). Evidently, if $\sum_{p \in P} d_p^\delta x_p = 0$, no edge in E_δ can be incident on more than one vertex of U , so that we in fact have $\sum_{p \in P} \bar{p}(\delta) x_p = |E_\delta|$. Likewise, if $\sum_{q \in Q} d_q^\delta y_q = 0$, no edge in E_δ can be incident on more than one vertex of V , so that we have $\sum_{q \in Q} \bar{q}(\delta) y_q = |E_\delta|$. It follows from (3.1) that the following sets of linear Diophantine clauses hold:

$$\left\{ \left(\sum_{p \in P} d_p^\delta x_p > 0 \right) \vee \left(\sum_{p \in P} \bar{p}(\delta) x_p \geq \sum_{q \in Q} \bar{q}(\delta) y_q \right) \mid \delta \in \Delta \right\} \quad (3.4)$$

$$\left\{ \left(\sum_{q \in Q} d_q^\delta y_q > 0 \right) \vee \left(\sum_{q \in Q} \bar{q}(\delta) y_q \geq \sum_{p \in P} \bar{p}(\delta) x_p \right) \mid \delta \in \Delta \right\}. \quad (3.5)$$

Regarding the x_p and y_q as variables ranging over \mathbb{N} , let $\mathcal{E}_{\mathcal{P}}$ be the system of linear Diophantine clauses (3.2)–(3.5). Notice that the various constants c_p^δ , c_q^δ , d_p^δ and d_q^δ depend only on \mathcal{P} and not on the supposed solution, H . Thus, we have shown that, if the BGE*-instance \mathcal{P} has a finite solution, then the system of linear Diophantine clauses $\mathcal{E}_{\mathcal{P}}$ has a solution over \mathbb{N} .

We now show that, conversely, if the system of linear Diophantine clauses $\mathcal{E}_{\mathcal{P}}$ has a solution over \mathbb{N} , then the BGE*-instance \mathcal{P} has a finite solution. Suppose, then, that the collections of integers $\{x_p\}_{p \in P}$ and $\{y_q\}_{q \in Q}$ satisfy (3.2)–(3.5). For each $p \in P$, we take a set of vertices U_p of cardinality x_p , and let U' be the disjoint union of the U_p ; similarly, for each $q \in Q$, we take a set of vertices V_q of cardinality y_q , and let V' be the disjoint union of the V_q . (We assume that U' and V' are disjoint.) Let us imagine each $u \in U_p$ to have $\bar{p}(\delta)$ ‘dangling’ δ -edges for all $\delta \in \Delta$; and likewise let us imagine each $v \in V_q$ to have $\bar{q}(\delta)$ ‘dangling’ δ -edges for all $\delta \in \Delta$. Our task is to match up these dangling edges so as to form a bipartite Δ -graph which is a solution of \mathcal{P} .

To make our task easier, we first construct a Δ -multigraph $G = (U', V', \mathbf{E}'_\Delta)$ with the required properties. Fix $\delta \in \Delta$, let $\delta(U')$ be the set of dangling δ -edges attached to the elements of U' , and let $\delta(V')$ be the set of dangling δ -edges attached to the elements of V' . Evidently,

$$|\delta(U')| = \sum_{p \in P} \bar{p}(\delta)x_p \quad |\delta(V')| = \sum_{q \in Q} \bar{q}(\delta)y_q. \quad (3.6)$$

Suppose first that $|\delta(U')| \leq |\delta(V')|$. Then we identify the elements of $\delta(U')$ with $|\delta(U')|$ elements of $\delta(V')$, thus dealing with all the dangling edges in $\delta(U')$ as well as $|\delta(U')|$ of the dangling edges in $\delta(V')$. If there are any remaining dangling edges in $\delta(V')$, we observe from (3.6) that $\sum_{p \in P} \bar{p}(\delta)x_p < \sum_{q \in Q} \bar{q}(\delta)y_q$, whence, from (3.4), $\sum_{p \in P} d_p^\delta x_p > 0$ —i.e. there exists $u \in U_p$, for some p such that $p(\delta) = \geq k$ for some $k \geq 0$. Pick any such u and attach all the dangling edges in $\delta(V')$ not yet accounted for to u . If, on the other hand, $|\delta(U')| \geq |\delta(V')|$, we apply the mirror-image construction and use (3.5). Either way, all dangling edges will eventually be linked up, and we obtain a Δ -coloured multigraph, G , with node-sets U' and V' . We see that, for each $p \in P$, and each $u \in U_p$, ord_u^G realizes p ; likewise, for each $q \in Q$, and each $v \in V_q$, ord_v^G realizes q . We make one small change to G before proceeding. Consider any $\delta \in \Delta_0$, and suppose there are no δ -edges in G . Then it is impossible that we have $x_p > 0$ for any $p \in P$ such that $\bar{p}(\delta) \geq 1$; and it is likewise impossible that we have $y_q > 0$ for any $q \in Q$ such that $\bar{q}(\delta) \geq 1$. From (3.2), then, we can find $p \in P$ such that $x_p > 0$ and $p(\delta) = \geq 0$; likewise, from (3.3), we can find $q \in Q$ such that $y_q > 0$ and $q(\delta) = \geq 0$. Pick $u \in U_p$ and $v \in V_q$, and add a δ -edge between u and v . This change clearly does not compromise the fact that ord_u^G realizes p or that ord_v^G realizes q . In this way, we can ensure that, for all $\delta \in \Delta_0$, G contains a δ -edge.

It remains to replace the Δ -multigraph G with a Δ -graph H in such a way that the realized order-functions are not disturbed. Let s be the maximum multiplicity of edges in H (i.e. the maximum number of any edges connecting any pair of vertices). Then H can be constructed by taking s replicas of G and appropriately rearranging the multiple edges. More precisely, let $U = \bigcup_{i=0}^{s-1} U_i$ and $V = \bigcup_{i=0}^{s-1} V_i$, where each U_i (V_i) is a fresh copy of U' (respectively, V'). For every $u \in U'$, denote by u_i the copy of u in U_i , and similarly for every $v \in V'$. Now execute the following process for all pairs $u \in U'$ and $v \in V'$ joined by at least one edge: let e_0, \dots, e_{r-1} ($0 < r \leq s$) be the collection of (coloured) edges joining u and v in G ; for every p ($0 \leq p \leq r-1$)

and i ($0 \leq i \leq s-1$), let H contain an edge with the same colour as e_p between u_i and $v_{i+p \bmod s}$. Since $r \leq s$, no pair of elements is joined by more than one edge, so that H is a Δ -graph. Moreover, for all $u \in U$, if u is a copy of some element $u' \in U_p$, then $\text{ord}_u^H = \text{ord}_{u'}^G = p$; similarly, if v is a copy of some element $v' \in V_q$, then $\text{ord}_v^H = \text{ord}_{v'}^G = q$. Thus, H is a solution of \mathcal{P} . We have proved:

LEMMA 3.4. *There is a polynomial-time reduction of finite BGE* to the satisfiability problem for sets of linear Diophantine clauses of the forms (3.2)–(3.5). We conclude:*

THEOREM 3.5. *Finite BGE* is in PTIME.*

Proof. By Lemma 3.4, it suffices to show that the satisfiability of systems of linear Diophantine clauses of the forms (3.2)–(3.5) can be solved in polynomial time.

Consider any system of r Diophantine linear inequalities of the form $\vec{a} \cdot \vec{z} \leq b$, in k variables; and let C be the maximum absolute value of any of the constants occurring in that system. It is well-known that, if there exists a solution (over \mathbb{N}), then there exists such a solution in which all values are bounded by $K = ((k+1)C)^r$ —that is, by an exponential function of the size of the system [2]. Evidently, therefore, given a system \mathcal{E} of linear Diophantine clauses of the forms (3.2)–(3.5), the same bound applies. Now let $R = kCK$, and replace any clause in \mathcal{E} of the form

$$(\vec{a} \cdot \vec{z} > 0) \vee (\vec{b} \cdot \vec{z} \geq \vec{c} \cdot \vec{z}) \quad (3.7)$$

by the corresponding linear equality

$$R\vec{a} \cdot \vec{z} + \vec{b} \cdot \vec{z} - \vec{c} \cdot \vec{z} \geq 0. \quad (3.8)$$

Let the resulting system of linear inequalities be \mathcal{E}' . We first observe that \mathcal{E}' entails \mathcal{E} . For if (3.8) holds, the corresponding instance of (3.7) clearly does too. Conversely, if \mathcal{E} has a solution, then so has \mathcal{E}' . For consider a solution of \mathcal{E} in which all entries are bounded by K , so that the expression $\vec{c} \cdot \vec{z}$ is at most $R = kCK$. If $\vec{a} \cdot \vec{z} = 0$, then (3.7) guarantees $\vec{b} \cdot \vec{z} - \vec{c} \cdot \vec{z} \geq 0$; on the other hand, if $\vec{a} \cdot \vec{z} \geq 1$, then $R\vec{a} \cdot \vec{z} - \vec{c} \cdot \vec{z} \geq 0$. Either way, (3.8) is satisfied, as required.

Hence, the problem of determining the satisfiability (over \mathbb{N}) of a system of linear Diophantine clauses of the forms (3.2)–(3.5) can be reduced in polynomial time to the problem of determining the corresponding problem for systems of linear Diophantine inequalities of the form $\vec{a} \cdot \vec{z} \geq b$. Evidently such a system has a solution over \mathbb{N} if and only if it has a solution over the non-negative rationals. The result then follows from the fact that linear programming feasibility is in PTIME [6]. \square

3.3. Complexity of BGE*. In this subsection we show that BGE* can be solved in polynomial time. Rather than introducing infinite values to the systems of equations considered in Sec 3.2, we proceed by reduction to the satisfiability problem for propositional Horn clauses. Recall, in this connection, that, if X_1, \dots, X_m, X are Boolean-valued variables, a *Horn clause* is an implication of either of the forms $X_1 \wedge \dots \wedge X_m \rightarrow X$ or $X_1 \wedge \dots \wedge X_m \rightarrow \perp$, interpreted in the usual way. It is well-known that the problem of determining the satisfiability of a collection of Horn clauses is in PTIME.

THEOREM 3.6. *BGE* is in PTIME.*

Proof. Let $\mathcal{P} = (\Delta, \Delta_0, M, P, Q)$ be an instance of BGE*. For $p \in P$, let X_p be a proposition letter, which we may informally read as “There are no vertices in U whose order-function realizes p .” Similarly, for $q \in Q$, let Y_q be a proposition letter,

which we may informally read as “There are no vertices in V whose order-function realizes q .” Consider the set Γ of propositional Horn-clauses

$$\left\{ \left(\bigwedge_{q \in Q: q(\delta) \neq 0} Y_q \right) \rightarrow X_p \mid p \in P, \delta \in \Delta \text{ s.t. } \bar{p}(\delta) > 0 \right\} \quad (3.9)$$

$$\left\{ \left(\bigwedge_{p \in P: p(\delta) \neq 0} X_p \right) \rightarrow Y_q \mid q \in Q, \delta \in \Delta \text{ s.t. } \bar{q}(\delta) > 0 \right\} \quad (3.10)$$

$$\left\{ \left(\bigwedge_{p \in P: p(\delta) \neq 0} X_p \right) \rightarrow \perp \mid \delta \in \Delta_0 \right\} \quad (3.11)$$

$$\left\{ \left(\bigwedge_{q \in Q: q(\delta) \neq 0} Y_q \right) \rightarrow \perp \mid \delta \in \Delta_0 \right\}. \quad (3.12)$$

Intuitively, (3.9) says “For all $\delta \in \Delta$, if no vertices in V are allowed to be incident on a δ -edge, then no vertices in U can be required to be incident on a δ -edge;” (3.10) expresses the mirror-image implication; (3.11) says “For all $\delta \in \Delta_0$, some vertices in U are allowed to be incident on some δ -edges;” and (3.12) expresses the same condition for vertices of V .

Suppose Γ is satisfiable. For each $p \in P$ such that X_p is false, take an infinite set U_p , and for each $q \in Q$ such that Y_q is false, take an infinite set V_q . Let $U = \bigcup_{p \in P} U_p$ and $V = \bigcup_{q \in Q} V_q$. For each $\delta \in \Delta$, for each $p \in P$, and each $u \in U_p$, attach $\bar{p}(\delta)$ ‘dangling’ δ -labelled edges to u ; and similarly for the elements of V , using the functions $q \in Q$. By (3.9), if a dangling δ -labelled edge is attached to some vertex of U , then there is $q \in Q$ with $q(\delta) \neq 0$ such that Y_q is false. We have two cases. Either we already have a dangling δ -labelled edge attached to some vertex of V , or no dangling δ -labelled edge is attached to any vertex of V , but there is $q \in Q$ with $q(\delta) \neq 0$ such that Y_q is false. In the latter case, we choose any such q , and for each $v \in V_q$, attach one dangling δ -edge. Using (3.10), we proceed analogously with U and V transposed. After this step, if there is a dangling δ -labelled edge attached to some vertex of U (hence infinitely many vertices of U), then there is a dangling δ -labelled edge attached to infinitely many vertices of V . These dangling edges can easily be matched up without clashes, thus forming an infinite Δ -graph. Finally, for every $\delta \in \Delta_0$ if there is no δ -edge attached to any vertex of U then using (3.11)-(3.12), we find $p \in P$ and $q \in Q$ such that $p(\delta) \neq 0$, X_p is false, $q(\delta) \neq 0$ and Y_q is false and we find $u \in U_p$ and $v \in V_q$ such that u and v are not connected by any edge and add a δ -edge from u to v . Hence \mathcal{P} is a positive instance of BGE^* .

Conversely, if \mathcal{P} is a positive instance of BGE^* , let $H = (U, V, \mathbf{E}_\Delta)$ be a solution. Now interpret the variables X_p and Y_q as indicated above. It is obvious that (3.9)–(3.12) hold. Thus, Γ is satisfiable. This completes the reduction. \square

4. Surgery on classes. In this section, we fix a normal form GFEC_2^2 -formula φ over signature $\tau = \tau_0 \cup \{r_1, r_2\} \cup \{r_1^\#, r_2^\#\}$.

Say that a τ -structure \mathcal{J} is a *pre-intersection* if for all $a, a' \in I$ we have $\mathcal{J} \models r_1^\#[a, a'] \wedge r_2^\#[a, a']$ (but we do not require $(r_i^\#)^{\mathcal{J}}$ to be the equivalence closure of $r_i^{\mathcal{J}}$). Obviously, if I is an intersection of \mathfrak{A} , then the induced substructure \mathcal{J} is a pre-intersection. We will use pre-intersections as building blocks to construct bigger

structures in which they will eventually become intersections. By the *type* of a pre-intersection, we mean its isomorphism type. Recalling Lemma 2.4, let Δ be the set of all types of pre-intersections \mathfrak{J} of size bounded by $f(|\tau|)$ such that $\mathfrak{J} \models \varphi_{12}$.

We write $\mathfrak{D} \approx \llbracket f \rrbracket_i$ if \mathfrak{D} is an r_i -connected structure built out of pre-intersections from Δ , and $f : \Delta \rightarrow \mathbb{N}^*$ is a function returning for each $\delta \in \Delta$ the number of realisations of δ in \mathfrak{D} . We will use such structures as $r_i^\#$ -classes in bigger models. Clearly, f does not fully determine \mathfrak{D} , but it carries information which is crucial for our purposes. In the case where $f : \Delta \rightarrow \mathbb{N}$, we define $|f| = \sum_{\delta \in \Delta} f(\delta)$.

For a given isomorphism type of pre-intersection δ and a 1-type α we say that δ *realizes* α if, for \mathfrak{D} of type δ , $\mathfrak{D} \models \exists x \alpha(x)$. For a given function $f : \Delta \rightarrow \mathbb{N}^*$ we say that f *realizes* α if δ realizes α for some $\delta \in \Delta$ such that $f(\delta) \geq 1$.

DEFINITION 4.1. *We say that a pre-intersection type δ is i -adjoinable to a function $f : \Delta \rightarrow \mathbb{N}^*$, and write $\delta \triangleright_i f$, if the following three conditions hold:*

- (i) every 1-type realized by δ is realized by f ,
- (ii) for any 1-types α realized by δ and α' realized by f (possibly $\alpha = \alpha'$) there exists a 2-type β such that $\beta(x, y) \models \varphi_{univ}$ and

$$\alpha(x) \cup \alpha'(y) \cup \{r_i^\#(x, y) \wedge \neg r_{3-i}^\#(x, y)\} \subseteq \beta(x, y),$$

- (iii) for any pre-intersection \mathfrak{J} of type δ , any of its r_i -connected components contains an element of 1-type α such that $K_i^{in}(x) \in \alpha$ or $K_i^{out}(x) \in \alpha$.

FACT 2. *Let $f : \Delta \rightarrow \mathbb{N}^*$, and assume that there is $\mathfrak{D} \models \varphi_i$ such that $\mathfrak{D} \approx \llbracket f \rrbracket_i$. If for some δ we have $f(\delta) \geq 2$ then $\delta \triangleright_i f$.*

Proof. Condition (i) follows from the fact that $f(\delta) > 0$. To see (ii) consider any 1-type α realized in δ . Assume that $\mathfrak{J}_1, \mathfrak{J}_2$ are pre-intersections of \mathfrak{D} of type δ promised by f . Let $a \in I_1$ be a realization of α . For any 1-type α' realized by f we can find a realization a' of α' such that $a' \notin I_1$. In particular, if α' is realized only in pre-intersections of type δ then a' can be found in I_2 . Now we choose β to be $\text{tp}^{\mathfrak{D}}(a, a')$. We have that $r_i^\#(x, y) \in \beta(x, y)$ since \mathfrak{D} is r_i -connected and $\neg r_{3-i}^\#(x, y) \in \beta(x, y)$ since a and a' are in different pre-intersections. Obviously $\beta(x, y) \models \varphi_{univ}$, since φ_{univ} is a fragment of φ_1 and $\mathfrak{D} \models \varphi_1$. Finally consider a realization \mathfrak{J} of δ in \mathfrak{D} and its r_i -connected component B . Since there are at least two pre-intersections in \mathfrak{D} and \mathfrak{D} is r_i -connected it follows that there must be at least one r_i -edge with one of its endpoints in B and the other in $D \setminus I$. Both endpoints in such an edge must be marked by K_i^{in} or by K_i^{out} (cf. a comment following Definition 2.1). \square

DEFINITION 4.2. *For functions $f, g : \Delta \rightarrow \mathbb{N}^*$ we say that g safely i -extends f , and write $g \succeq_i f$, if for all $\delta \in \Delta$, (i) $g(\delta) \geq f(\delta)$, and (ii) $g(\delta) > f(\delta)$ implies $\delta \triangleright_i f$.*

The following two lemmas will be used later. The first allows us to ‘inflate’ $r_i^\#$ -classes in a model of φ by adjoining pre-intersections. The second is much more involved: it allows us, under certain circumstances, to ‘deflate’ $r_i^\#$ -classes in a model of φ by removing intersections; we show that, after some minor housekeeping, the resulting structure is still a model of φ .

LEMMA 4.3. *Let $f, g : \Delta \rightarrow \mathbb{N}^*$ be such that $g \succeq_i f$. If there exists $\mathfrak{D} \models \varphi_i$ such that $\mathfrak{D} \approx \llbracket f \rrbracket_i$, then there exists $\mathfrak{D}' \models \varphi_i$ such that $\mathfrak{D}' \approx \llbracket g \rrbracket_i$.*

Proof. Let $\mathfrak{D} \models \varphi_i$ and $\mathfrak{D} \approx \llbracket f \rrbracket_i$. We build \mathfrak{D}' out of \mathfrak{D} and $g(\delta) - f(\delta)$ pre-intersections of type δ for each δ (where $\aleph_0 - n = \aleph_0$ for all $n \in \mathbb{N}$). First let us define connections between \mathfrak{D} and each of the newly added pre-intersections \mathfrak{J} . Note that by Definition 4.2 the type of each of these new pre-intersections is i -adjoinable to f . For each element $a \in I$, using condition (i) of Definition 4.1, find an element $a' \in D$ realizing the same 1-type α . For each $b \in D$, $b \neq a'$ set $\text{tp}^{\mathfrak{D}'}(a, b) := \text{tp}^{\mathfrak{D}}(a', b)$.

Set $\text{tp}^{\mathfrak{D}'}(a, a') := \beta$, where β is a 2-type promised by condition (ii) of Definition 4.1 for $\alpha' = \alpha$. This guarantees that a has appropriate witnesses inside \mathfrak{D} and $\mathfrak{D}' \upharpoonright (D \cup I) \models \varphi_{univ}$, for, by Definition 2.3, φ_{univ} is a conjunct of φ_{12} and in this section every $\delta \in \Delta$ is a type of some pre-intersection satisfying φ_{12} . Moreover due to condition (iii) of Definition 4.1 \mathfrak{D}' is r_i -connected (as \mathfrak{D} is r_i -connected, and elements $a \in I$ satisfying K_i^{in} or K_i^{out} must have an r_i -edge to D).

It remains to define connections between newly added pre-intersections. Let $\mathfrak{J}, \mathfrak{J}'$ be two distinct newly added pre-intersections, and let $a \in I, a' \in I'$ be elements realizing 1-types α, α' , respectively. Condition (i) of Definition 4.1 guarantees that α' is realized by f . Set $\text{tp}^{\mathfrak{D}'}(a, a') := \beta$, where β is a 2-type promised by part (ii) of Definition 4.1. This ensures that $\mathfrak{D}' \models \varphi_{univ}$. Together with the earlier remarks we have that $\mathfrak{D}' \models \varphi_i$. \square

LEMMA 4.4. *Let f be a function $\Delta \rightarrow \mathbb{N}^*$ such that $\mathfrak{D} \approx \llbracket f \rrbracket_i$ for some $\mathfrak{D} \models \varphi_i$. Then there exists a function $g : \Delta \rightarrow \mathbb{N}$, such that $f \succeq_i g$ and $|g| \leq \mathfrak{g}(|\tau|)$ (for a fixed exponential function \mathfrak{g}) and a structure $\mathfrak{D}' \approx \llbracket g \rrbracket_i$ such that $\mathfrak{D}' \models \varphi_i$.*

Proof. We prove the lemma for $i = 1$. The proof for $i = 2$ is symmetric. Let m be the number of $\forall\exists$ -conjuncts in φ_1 . We proceed by first selecting a sub-model $\mathfrak{D}_0 \subseteq \mathfrak{D}$, and then modifying \mathfrak{D}_0 to produce the required structure \mathfrak{D}' . The selection of \mathfrak{D}_0 comprises four steps.

Step 1. For each 1-type α realized in \mathfrak{D} , mark m distinct pre-intersections containing a realisation of α (or all such pre-intersections if α is realized in fewer than m of them). Let B_0 be the union of all pre-intersections marked in this step. Note that, for any element whose connections with B_0 are not fixed, B_0 will be able to supply any witnesses required by the $\forall\exists$ -conjuncts in φ_1 .

Step 2. For each $a \in B_0$ and each conjunct $\forall x(\gamma(x) \rightarrow \exists y(r_1^\# [a, b] \wedge \neg r_2^\# [a, b] \wedge x \neq y \wedge \psi(x, y)))$ in φ_1 , if $\mathfrak{D} \models \gamma[a]$, then find a witness $b \in D$ such that $\mathfrak{D} \models r_1^\# [a, b] \wedge \neg r_2^\# [a, b] \wedge \psi[a, b]$, and mark the pre-intersection of b . Let B_1 be the union of all pre-intersections marked in this step.

Step 3. For each pair of elements $a, b \in B_0 \cup B_1$ which are not r_1 -connected in $\mathfrak{D} \upharpoonright (B_0 \cup B_1)$, if there is an r_1 -path of the form $a = c_1, c_2, \dots, c_l = b$ with $c_2, \dots, c_{l-1} \notin B_0 \cup B_1$, then choose one such path, call it P_{ab} , and mark the pre-intersections of c_2, \dots, c_{l-1} . (Obviously, an r_1 -path in \mathfrak{D} between a and b exists for all $a, b \in B_0 \cup B_1$, as \mathfrak{D} is r_1 -connected. Note that we consider only such paths in which all elements, except a and b , lie outside $B_0 \cup B_1$.) Additionally, call the two elements c_2 and c_{l-1} *peripheral connectors* (of P_{ab}). Let B_2 be the set of all pre-intersections marked in this step. Note that after this step any pair of elements from $B_0 \cup B_1$ is r_1 -connected in $\mathfrak{D} \upharpoonright (B_0 \cup B_1 \cup B_2)$. This does not mean however that $\mathfrak{D} \upharpoonright (B_0 \cup B_1 \cup B_2)$ is r_1 -connected (since some pre-intersections from B_2 may be not internally r_1 -connected).

Step 4. For any element a being a peripheral connector or belonging to $B_1 \setminus B_0$, and for each conjunct $\forall x(\gamma(x) \rightarrow \exists y(r_1^\# [a, b] \wedge \neg r_2^\# [a, b] \wedge x \neq y \wedge \psi(x, y)))$ from φ_1 , if $\mathfrak{D} \models \gamma[a]$, then find a witness $b \in D$ such that $\mathfrak{D} \models r_1^\# [a, b] \wedge \neg r_2^\# [a, b] \wedge \psi[a, b]$, and mark the pre-intersection of b . Let B_3 be the union of all pre-intersections marked in this step.

Let $\mathfrak{D}_0 = \mathfrak{D} \upharpoonright (B_0 \cup B_1 \cup B_2 \cup B_3)$. We now modify \mathfrak{D}_0 to yield the desired structure \mathfrak{D}' . At this moment recall that: (i) elements from B_0, B_1 and peripheral connectors from B_2 have the required witnesses in \mathfrak{D}_0 ; (ii) the sizes of B_0, B_1, B_3 and the number of peripheral connectors are bounded exponentially in $|\tau|$; (iii) all pairs of

elements from $B_0 \cup B_1$ are r_1 -connected in \mathfrak{D}_0 . What remains is to decrease the size of B_2 (which at this moment is unbounded), provide any remaining witnesses, and make the whole class r_1 -connected.

Consider a path P_{ab} of the form $a = c_1, c_2, \dots, c_{l-1}, c_l = b$ chosen in *Step 3*. Repeat the following procedure as long as possible: if for some i, j , $3 \leq i < j \leq l - 2$ such that the 1-types of c_i and c_j are identical, c_{i-1} does not belong to the pre-intersection of c_i , and c_{j-1} does not belong to the pre-intersection of c_j then remove elements c_i, \dots, c_{j-1} from P_{ab} and make the connection between c_{i-1} and c_j equal to the connection between c_{i-1} and c_i (note that this connection must contain an r_1 -edge; thus after this cut, P_{ab} remains an r_1 -path). Since the sizes of pre-intersections and the number of 1-types are bounded exponentially, the length of the final version of P_{ab} is also exponentially bounded. Perform the above process for all paths P_{ab} chosen in *Step 3*. Let B'_2 contain only pre-intersections of elements from the paths P_{ab} so shortened. The size of $B_0 \cup B_1 \cup B'_2 \cup B_3$ is now exponentially bounded.

For any element a from $B'_2 \cup B_3$ which is not a peripheral connector, modify its connections to B_0 in such a way that witnesses for a required by the $\forall\exists$ -conjuncts of φ_i can be found in B_0 . This is possible due to our choice of B_0 in *Step 1*, since a may require at most m witnesses. We remark that the special treatment of peripheral connectors is important here: they differ from the remaining elements of $B_2 \cup B_3$ in that their connections to some elements from B_0 could earlier have been fixed to contain r_1 -edges. Note in addition that an element may be a peripheral connector of many chosen paths. Denote the resulting structure (over domain $B_0 \cup B_1 \cup B'_2 \cup B_3$) by \mathfrak{D}' .

We claim that \mathfrak{D}' has the required properties. Indeed, $|B_0 \cup B_1 \cup B'_2 \cup B_3|$ is exponentially bounded in $|\tau|$, as explained. Furthermore, all 2-types occurring in \mathfrak{D}' occur in \mathfrak{D} , and we took care to provide all required witnesses, whence $\mathfrak{D}' \models \varphi_1$. The r_1 -connectivity of \mathfrak{D}' now follows from the fact that r_1 -paths connect all pairs of elements from $B_0 \cup B_1$ and each of the remaining elements is r_1 -connected to some element in its pre-intersection satisfying K_1^{in} or K_1^{out} which further must have a witness in B_0 connected to it by a direct r_1 -edge (cf. *Remarks* after Definition 2.1).

Let $g : \Delta \rightarrow \mathbb{N}$ be the function such that $\mathfrak{D}' \approx \llbracket g \rrbracket_1$. It remains to see that $f \succeq_1 g$. Since we built \mathfrak{D}' using only pre-intersections from \mathfrak{D} and we did not change the connections inside pre-intersections it is clear that for all $\delta \in \Delta$ we have $f(\delta) \geq g(\delta)$. Consider δ such that $f(\delta) > g(\delta)$. We must show that $\delta \triangleright_1 g$. Let α be realized in δ . In *Step 1* we choose a pre-intersection containing α and make it a member of B_0 which later becomes a fragment of \mathfrak{D}' . Thus α is realized by g , and condition (i) of Definition 4.1 is satisfied. Consider now any α realized by δ and any α' realized by g . Let \mathfrak{J}_2 be a pre-intersection of \mathfrak{D}' which contains an element a' of type α' . By our construction \mathfrak{J}_2 is also a pre-intersection of \mathfrak{D} . Let \mathfrak{J}_1 be a pre-intersection of type δ in \mathfrak{D} different from \mathfrak{J}_2 (we can choose such \mathfrak{J}_1 even if \mathfrak{J}_2 is of type δ since we know that $f(\delta) > g(\delta)$). The type $\text{tp}^{\mathfrak{D}}(a, a')$ can be chosen as β from condition (ii). The argument for condition (iii) is similar to the one given in the proof of Fact 2. \square

Having obtained the function \mathbf{g} in Lemma 4.4, we can present a method of converting functions $f : \Delta \rightarrow \mathbb{N}^*$ to functions $f^{\triangleright i} : \Delta \rightarrow \hat{M}$.

DEFINITION 4.5. *Let $f : \Delta \rightarrow \{0, \dots, M\}$ be a function and $i \in \{1, 2\}$. Define $f^{\triangleright i} : \Delta \rightarrow \hat{M}$ as follows:*

$$f^{\triangleright i}(\delta) = \begin{cases} \geq f(\delta) & \text{if } \delta \triangleright_i f \\ = f(\delta) & \text{otherwise.} \end{cases}$$

Note that the values of $f^{\triangleright i}$ are indeed in \hat{M} : in particular, by Fact 2, $f^{\triangleright i}$ cannot return \bar{k} for $k > 1$.

FACT 3. *Let $M = \mathfrak{g}(|\tau|)$, $f : \Delta \rightarrow \{0, \dots, M\}$, $g : \Delta \rightarrow \mathbb{N}^*$, and $i \in \{1, 2\}$. If $g \succeq_i f$ then g realizes $f^{\triangleright i}$.*

Proof. Follows directly from definitions. \square

5. From logic to graphs. We now present a reduction of the (finite) satisfiability problem for normal-form GFEC₂²-formulas to the problem (finite) BGE*, running in time bounded by a doubly-exponential function of $\|\varphi\|$. The reduction employs some nondeterministic guesses, but all of them are of size exponential in $|\tau|$. This ensures that the eventual decision procedure runs in deterministic doubly-exponential time.

Let the normal-form GFEC₂²-formula φ , with signature τ , be given. Let the set of pre-intersections Δ be as defined in Sec. 4 and let $M = \mathfrak{g}(|\tau|)$, where \mathfrak{g} is the function given in Lemma 4.4. Note that $|\Delta|$ is doubly exponentially bounded as a function of $\|\varphi\|$. We proceed to construct a subset $\Delta_0 \subseteq \Delta$ and sets P, Q of functions mapping Δ to \hat{M} . Together with Δ and M , these will form a (finite) BGE*-instance giving us the desired reduction.

Reduction procedure:

- (i) Nondeterministically guess a set $\mathbf{\alpha}$ of 1-types over τ ; verify that for each $\alpha \in \mathbf{\alpha}$ and each conjunct of φ of the form $\forall x\psi(x)$, $\alpha(x) \models \psi(x)$.
- (ii) Verify that for each conjunct of φ of the form $\exists x\psi(x)$, there exists $\alpha \in \mathbf{\alpha}$ such that $\alpha(x) \models \psi(x)$.
- (iii) Verify that for each conjunct of φ_{free} of the form $\forall x(\gamma(x) \rightarrow \exists y(\eta(x, y) \wedge x \neq y \wedge \psi(x, y)))$ and each $\alpha \in \mathbf{\alpha}$ such that $\alpha(x) \models \gamma(x)$, there exists $\alpha' \in \mathbf{\alpha}$ and a 2-type $\beta \supseteq \alpha(x) \cup \alpha'(y) \cup \{-r_1^\#(x, y) \wedge \neg r_2^\#(x, y)\}$ such that:
 - $\beta(x, y) \models \eta(x, y) \wedge \psi(x, y)$;
 - for each conjunct of φ_{univ} of the form $\forall xy(\eta'(x, y) \rightarrow (x \neq y \rightarrow \psi(x, y)))$, $\beta(x, y) \models \eta'(x, y) \rightarrow \psi(x, y)$.
- (iv) For each $\alpha \in \mathbf{\alpha}$ guess a pre-intersection type $\delta_\alpha \in \Delta$ such that α is realized in δ_α and δ_α realizes only 1-types from $\mathbf{\alpha}$; let $\Delta_0 = \{\delta_\alpha : \alpha \in \mathbf{\alpha}\}$.
- (v) Start with $P = \emptyset$. Generate all $r_1^\#$ -classes \mathfrak{D} consisting of at most $M = \mathfrak{g}(|\tau|)$ pre-intersections of types from Δ realizing only 1-types from $\mathbf{\alpha}$; if $\mathfrak{D} \models \varphi_1$, then take f such that $\mathfrak{D} \approx \llbracket f \rrbracket_1$, and add to P the function $f^{\triangleright 1}$.
- (vi) Construct Q analogously, adding to Q the function $f^{\triangleright 2}$.

PROPOSITION 5.1. *φ is (finitely) satisfiable iff for some guesses of $\mathbf{\alpha}$ in (i) and Δ_0 in (iv), the run of the above reduction procedure produces a positive instance $(\Delta, \Delta_0, M, P, Q)$ of (finite) BGE*.*

Proof. \Rightarrow Let \mathfrak{A} be a (finite) model of φ with intersections bounded by $\mathfrak{f}(|\tau|)$ as guaranteed by Lemma 2.4, and recall that $M = \mathfrak{g}(|\tau|)$. In step (i) guess $\mathbf{\alpha}$ to be the set of 1-types realized in \mathfrak{A} . Steps (ii) and (iii) are then clearly successful. Also, by definition of Δ all pre-intersection types realized in \mathfrak{A} are members of Δ . In step (iv), for each $\alpha \in \mathbf{\alpha}$ choose δ_α to be a type of an arbitrary pre-intersection \mathfrak{J} from \mathfrak{A} containing a realisation of α ; this defines Δ_0 . We claim that $(\Delta, \Delta_0, M, P, Q)$, where P and Q are generated in steps (v) and (vi), has a (finite) solution. To construct a solution $H = (U, V, \mathbf{E}_\Delta)$ define U to be the set of $r_1^\#$ -classes and V to be the set of $r_2^\#$ -classes of \mathfrak{A} , and for each $\delta \in \Delta$ add $(\mathfrak{D}, \mathfrak{D}')$ to E_δ iff \mathfrak{D} and \mathfrak{D}' have a common intersection of type δ . Condition (G1) of Definition 3.3 is obviously satisfied. Consider condition (G2). Let $\mathfrak{D} \in U$. We know that $\mathfrak{D} \models \varphi_1$. Denote $f = \text{ord}_2^H$. We want

to see that f realizes a function from P . Note that $\mathfrak{D} \approx \llbracket f \rrbracket_1$. By Lemma 4.4 there exists g such that $f \succeq_1 g$, $|g| \leq \mathfrak{g}(|\tau|)$ and there is a structure $\mathfrak{D}' \models \varphi_1$ such that $\mathfrak{D}' \approx \llbracket g \rrbracket_1$. By step (v) of the reduction procedure P contains $g^{\triangleright 1}$. By Fact 3 we have that f realizes $g^{\triangleright 1}$. (G3) follows analogously.

\Leftarrow In the opposite direction consider a (finite) solution $H = (U, V, \mathbf{E}_\Delta)$ of an instance $(\Delta, \Delta_0, M, P, Q)$ of BGE* constructed by our procedure, with α being the guess in step (i). We show how to construct a (finite) model $\mathfrak{A}' \models \varphi$ from the graph H . However, first we construct an auxiliary structure \mathfrak{A} satisfying conditions (i)-(iii) of Fact 1, i.e., a structure satisfying all conjuncts of φ except possibly conjuncts from φ_{free} . Intersections of \mathfrak{A} correspond to the edges of H : for each $\delta \in \Delta$ and each $e \in E_\delta$, we put into \mathfrak{A} a pre-intersection I_e of type δ . Property (G1) of Definition 3.3 ensures that \mathfrak{A} satisfies all \exists -conjuncts, i.e. condition (i) of Fact 1 holds. Moreover, the fact that all intersections have types from Δ ensures condition (ii).

Consider now any vertex $u \in U$. Let \mathcal{U} be the set of all pre-intersections corresponding to the edges incident to u . Our task is to compose from them an $r_1^\#$ -class \mathfrak{D}_u satisfying φ_1 . Writing f for ord_u^H , let $p \in P$ be a function which is realized by f , guaranteed by (G2). First, we form from some subset of \mathcal{U} a class $\mathfrak{D} \approx \llbracket \bar{p} \rrbracket_1$ such that $\mathfrak{D} \models \varphi_1$. This is possible by the construction of P in step (v). Note that $f \succeq_1 \bar{p}$. Indeed, if for some $f(\delta) > \bar{p}(\delta)$ then $p(\delta) = \succeq k$ for some $k \in \mathbb{N}$ and by Definition 4.5 we have $\delta \triangleright_1 \bar{p}$. Thus, by Lemma 4.3, all the remaining intersections of \mathcal{U} can be joined to \mathfrak{D} , forming the desired \mathfrak{D}_u . We repeat this construction for all vertices in U . This ensures that every pre-intersection is r_1 -connected and that condition (iii) of Fact 1 holds for $i = 1$.

Similarly, from any vertex $v \in V$, we form a $r_2^\#$ -class consisting of all pre-intersections corresponding to edges incident on v , using (G3) and the construction of Q . This step ensures condition (iii) for $i = 2$ and makes every pre-intersection r_2 -connected. Thus, all pre-intersections become both r_1 - and r_2 -connected; moreover, no two pre-intersections can be connected to each other by both r_1 and r_2 (because no two edges of H can have common vertices in both U and V); hence, every pre-intersection becomes an intersection of \mathfrak{A} , as required.

At this point, we have specified the 2-type in \mathfrak{A} of any pair of elements which belong to the same $r_1^\#$ - or $r_2^\#$ -class. To complete the definition of \mathfrak{A} , for any pair of elements which do not share an $r_1^\#$ - nor $r_2^\#$ class join them by the unique compatible 2-type which does not contain any positive binary literals. Note that each 2-type in \mathfrak{A} either does not contain any positive binary literals or has been copied from a structure which satisfied φ_{univ} . This ensures that $\mathfrak{A} \models \varphi_{univ}$.

Now we take care of the $\forall\exists$ -conjuncts from φ_{free} using a circular witnessing pattern known from [4]. Let m be the number of formulas of type $\forall\exists$ in φ_{free} . We build \mathfrak{A}' out of $3m$ disjoint copies of \mathfrak{A} which are logically divided into three groups X_0, X_1, X_2 consisting of m copies each. This guarantees that in each group every 1-type from α is realized at least m times. Now, for every element $a \in X_i$ ($i = 0, 1, 2$) and every conjunct from φ_{free} of the form $\forall x(\gamma(x) \rightarrow \exists y(\eta(x, y) \wedge x \neq y \wedge \psi(x, y)))$, if the 1-type of a is α , and $\alpha(x) \models \gamma(x)$ then let α' and β be as guaranteed by step (iii) of the procedure. By step (iv) α' is realized in \mathfrak{A} . Take a fresh copy b of such a realisation from $X_{i+1 \bmod 3}$ and set $\text{tp}^{\mathfrak{A}'}(a, b) = \beta$. As in $X_{i+1 \bmod 3}$ each necessary 1-type is realized at least m times every element from X_i can choose its witnesses without conflicts. All 2-types in \mathfrak{A}' which have not been fixed yet are set in such a way that they contain no positive binary literals. One can see that \mathfrak{A}' satisfies all conditions of Fact 1, hence $\mathfrak{A}' \models \varphi$. \square

Recalling Lemma 2.2 and Theorems 3.5 and 3.6 we obtain the main result of our paper.

THEOREM 5.2. *The (finite) satisfiability problem for GFEC_2^2 is in 2-EXPTIME.*

Finally, we observe a model-theoretic difference between the logics EC_2^2 and GFEC_2^2 (which also holds between EQ_2^2 and GFEQ_2^2). Recall that all these logics allow one to construct infinity axioms. Moreover, in EC_2^2 or EQ_2^2 it is even possible to write a satisfiable formula all of whose models must contain infinite equivalence classes (see Example 5.1.1 in [10]). A closer inspection of our constructions shows that this is not the case in GFEC_2^2 (a similar observation for GFEQ_2^2 follows also from the proof of the upper bound in Theorem 5 from [7]). Indeed, in the model \mathfrak{A} obtained in the proof of \Leftarrow from Prop. 5.1, elements correspond to vertices, and intersections correspond to edges of the graph H , which we can assume is obtained as in the proof of Thm. 3.6. The maximal number of intersections in a class of \mathfrak{A} is thus equal to the maximal degree of a node of H . Each such node is incident to precisely $|f|$ edges for some function f , for which $f^{\triangleright 1}$ was added to P in step (v) or $f^{\triangleright 2}$ to Q in step (vi) of our reduction procedure. Note that $|f| \leq M$, and M is exponentially bounded in $\|\varphi\|$. Hence we can formulate the following *exponential classes property*.

COROLLARY 5.3. *Every satisfiable GFEC_2^2 formula φ has a model in which all $r_1^\#$ - and $r_2^\#$ -classes are bounded exponentially in $\|\varphi\|$.*

It is worth recalling that if we consider satisfiability over finite models then it is possible to enforce in GFEC_2^2 classes of doubly exponential size. It is even possible in a restricted fragment of GFEQ_2^2 in which equivalences are used only as guards (see Example 1 in [11]).

REFERENCES

- [1] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27:217–274, 1998.
- [2] I. Borosh, M. Flahive, and B. Treybig. Small solutions of linear Diophantine equations. *Discrete Mathematics*, 58(3):215–220, 1986.
- [3] C. C. Chang and H. J. Keisler. *Model Theory*. North-Holland, Amsterdam, 3rd edition, 1990.
- [4] E. Grädel, P. Kolaitis, and M. Vardi. On the decision problem for two-variable first-order logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.
- [5] Erich Grädel. On the restraining power of guards. *Journal of Symbolic Logic*, 64(4):1719–1742, 1999.
- [6] L. G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, 20:191–194, 1979.
- [7] E. Kieroński. Results on the guarded fragment with equivalence or transitive relations. In *Computer Science Logic*, volume 3634 of *LNCS*, pages 309–324. Springer, 2005.
- [8] E. Kieroński, J. Michaliszyn, I. Pratt-Hartmann, and L. Tendera. Two-variable first-order logic with equivalence closure. In *Logic in Computer Science*, pages 431–440. IEEE, 2012.
- [9] E. Kieroński, J. Michaliszyn, I. Pratt-Hartmann, and L. Tendera. Two-variable first-order logic with equivalence closure. *SIAM Journal of Computing*, 43(3):1012–1063, 2014.
- [10] E. Kieroński and M. Otto. Small substructures and decidability issues for first-order logic with two variables. *Journal of Symbolic Logic*, 77:729–765, 2012.
- [11] E. Kieroński and L. Tendera. On finite satisfiability of the guarded fragment with equivalence or transitive guards. In *Logic for Programming, Artificial Intelligence, and Reasoning*, volume 4790 of *LNAI*, pages 318–332. Springer Verlag, 2007.
- [12] E. Kieroński and L. Tendera. On finite satisfiability of two-variable first-order logic with equivalence relations. In *Logic in Computer Science*, pages 123–132. IEEE, 2009.
- [13] M. Mortimer. On languages with two variables. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 21:135–140, 1975.
- [14] W. Szwaś and L. Tendera. The guarded fragment with transitive guards. *Annals of Pure and Applied Logic*, 128:227–276, 2004.