Volume 72 (2026), pp. 135–156
Received: 8 April 2025
Revised: 24 July 2025
Accepted: 31 July 2025
DOI: 10.60866/CAM.257

ZML
Zeitschrift für Mathematische Logik
und Grundlagen der Mathematik

Limits of structures and total NP search problems

Ondřej Ježil

Faculty of Mathematics and Physics, Charles University, Sokolovská 83, Prague, 186 75, Czech Republic

Abstract. For an infinite class of finite graphs of unbounded size, we define a limit object, to be called a *wide limit*, relative to some computationally restricted class of functions. The limit object is a first-order Boolean-valued structure. The first-order properties of the wide limit then reflect how a computationally restricted viewer sees a generic member of the class. The construction uses arithmetic forcing with random variables [11]. We give sufficient conditions for universal and existential sentences to be valid in the limit, provide several examples, and prove that such a limit object can then be expanded to a model of weak arithmetic.

To illustrate the concept, we give an example in which the wide limit relates to total NP search problems. In particular, we take the wide limit of all maps from $\{0,\dots,k-1\}$ to $\{0,\dots,\lfloor k/2\rfloor-1\}$ to obtain a model of $\forall\mathrm{PV}_1(f)$ where the problem RetractionWeakPigeon is total, but WeakPigeon, the complete problem for PWPP, is not. Thus, we obtain a new proof of this unprovability and show it implies that WeakPigeon is not many-one reducible to RetractionWeakPigeon in the oracle setting.

1 Introduction

The notion of limits of finite structures is prevalent both in logic and in combinatorics. In logic the examples are the ultraproduct and the compactness theorem, which was used in [4] to prove the 0–1 law for structures over relational vocabularies. In combinatorics, the dense graph limit defined by Lovász & Szegedy [13] provided a framework to restate and find new proofs for results in extremal graph theory—for instance, Goodman's theorem relating the number of edges to the number of triangles in a graph. More notions of graph limits are discussed in [15]. Another recent use of limit objects for results of extremal combinatorics was by Razborov [17].

In this work we define a new construction of a limit object. Given a class of finite graphs $\mathcal{G}$, whose vertex sets are initial segments of $\mathbb{N}$, we can stratify it into the sequence of sets $\{\mathcal{G}_k\}_{k=1}^{\infty}$, where we define $\mathcal{G}_k = \{\omega \in \mathcal{G}\ ;\ \omega \text{ has } \{0,\dots,k-1\} \text{ as its vertex set}\}$. We are interested in the case when the cardinalities of $\mathcal{G}_k$ are unbounded and hence our intended limit is a limit of sets of finite graphs. For this reason we call such a sequence of sets of graphs a *wide sequence* and the limit object its *wide limit*. The qualification "wide" refers to the fact that we are interested in sequences of sets of graphs rather than sequences of individual graphs.

We will give a full definition in Section 3, but we describe some important properties here. For a class F of functions (typically chosen to be a class with some computational restrictions) we define the wide limit, denoted by $\lim_F \mathcal{G}_n$, where n is a technical parameter to be defined later. The wide limit $\lim_F \mathcal{G}_n$ is a Boolean-valued graph—its edge relation does not only permit the truth values **0** and **1** but also many other values from some infinite complete Boolean algebra $\mathcal{B}$.[1] This algebra is, in fact, also a measure algebra with a measure μ on

Correspondence email: ondrej.jezil@email.cz

[1] Generally, we can do this with any L-structures for some first-order language L. The limit object is then a Boolean-valued L-structure $\lim_F \mathcal{G}_n$. In this work we restrict ourselves to the language of graphs $L = \{E\}$ to simplify the presentation.

This article is licensed under the Creative Commons BY 4.0 license.
Copyright is retained by the author.

it, so to any statement formulated as a first-order sentence φ we can assign a real number $\mu([\![\varphi]\!]) \in [0,1]$, which measures how far the truth value of φ, denoted by $[\![\varphi]\!]$, is from the value $\mathbf{0}$. The key method we use is arithmetical forcing with random variables, developed by Krajíček [11], which allows us to construct models of (weak) arithmetical theories and, by restricting to the language of graphs, gives us Boolean-valued graphs. In these Boolean-valued graphs, existential statements that obtain the maximal truth value $\mathbf{1}$ (i.e., valid sentences) correspond to the ability of F to solve search problems over the class of graphs we are considering.

Our limit object can be expanded to the original model that Krajíček's method would otherwise construct. We prove (Theorem 24) that the truth values of first-order sentences concerning the object are preserved even when evaluated in the model of arithmetic relativized to the wide limit (under a mild condition on the family F).

As an application of this construction, we reprove a many-one separation of black-box total NP search problems. A total NP search problem, first introduced by Papadimitriou [16], is a problem that always has a solution of size polynomial in the input size and for which the correctness of a solution can be verified by a polynomial time machine. The qualifier *black-box* means that the input is given by a pair $(\mathcal{O}, x)$, where x is a binary string and $\mathcal{O}$ is a function oracle. The problems of our interest are the problem RetractionWeakPigeon, whose totality follows from the nonexistence of a retraction from $\{0,\dots,k-1\}$ and $\{0,\dots,\lfloor k/2\rfloor-1\}$, and the problem WeakPigeon, whose totality follows from the nonexistence of an injection between the same pair of sets.

We take a wide limit of all maps from the interval $\{0,\dots,k-1\}$ to the interval $\{0,\dots,\lfloor k/2\rfloor-1\}$ relative to trees of subexponential depth (in the length of k) querying images of elements to obtain a Boolean-valued graph of a function $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}$, which is, in a sense, elementarily equivalent to an injective map from the interval $\{0,\dots,n-1\}$ to $\{0,\dots,\lfloor n/2\rfloor-1\}$ for some nonstandard number n. We then expand the wide limit into a model of arithmetic $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$, where RetractionWeakPigeon is total, but $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}$ is an instance of WeakPigeon that has no solution.

There is already an established connection between complexity of search problems and logic (namely witnessing theorems in bounded arithmetic, see [7; 14]). The model we construct is a model of a relatively weak theory $\forall\mathrm{PV}_1(f)$ and also of open induction and open comprehension with parameters. The existence of this model gives a new proof of the result of Thapen [18, Theorem 4.10], that these principles along with the principle that RetractionWeakPigeon is total cannot prove that the problem WeakPigeon is total.

This paper has two main parts. The emphasis of the paper is on the first *conceptual* part, where we introduce the new notion of a wide limit. This part consists of Sections 2 to 4: In Section 2 we recall preliminary notions, most importantly nonstandard models of arithmetic. In Section 3 we give the definition of the wide limit, provide several examples, and show how density of finite substructures corresponds to validity of existential sentences in the wide limit (Theorem 16). In Section 4 we prove that, under some reasonable assumptions, the wide limit can be expanded to a model of two sorted arithmetic (Theorem 24).

The second part, consisting of Sections 5 and 6, is about our application of the new concept. In Section 5, we recall the definition of query TFNP and show that the wide limit of all black-box instances of WeakPigeon is a graph of an injective map from $\{0,\dots,n-1\}$ to $\{0,\dots,\lfloor n/2\rfloor-1\}$. Finally, in Section 6 we expand the wide limit to a model of two-sorted arithmetic, $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$, where we verify the following properties:

(1) the wide limit is an instance of WeakPigeon without a solution (Theorem 41),

(2) every instance of RetractionWeakPigeon has a solution (Theorem 49),

(3) open comprehension and open induction (Theorem 52),

(4) the theory $\forall\mathrm{PV}_1(f)$ (Theorem 53),

and also refute the existence of a many-one reduction from the problem WeakPigeon to the problem RetractionWeakPigeon (Theorem 51). This separation is not new and follows from known results (see Section 7 for more details), but our illustration gives the separation a clear semantic interpretation.

2 Preliminaries

By *digraphs*, we mean structures in a language with a single binary relation E that is antireflexive. By *graphs* we mean digraphs where E is symmetric. We will denote any particular (di)graph by ω, as it will be used, in some sense, as a sample of a discrete probability space. The edge relation of a particular graph ω will be denoted by E_ω.

In the rest of this section, we recall notions needed for Krajíček's forcing construction. A fundamental notion we use throughout the work is that of nonstandard models of (true) arithmetic. Let L_{all} be the language containing the names of all relations and functions on the natural numbers and let $\text{Th}_{L_{\text{all}}}(\mathbb{N})$ denote the set of true sentences in this language in the standard model $\mathbb{N}$. By classical results of logic, there exist L_{all}-structures that are not isomorphic to $\mathbb{N}$ in which all sentences from $\text{Th}_{L_{\text{all}}}(\mathbb{N})$ are valid. These are called *nonstandard models* (of $\text{Th}_{L_{\text{all}}}(\mathbb{N})$).

All nonstandard models of $\text{Th}_{L_{\text{all}}}(\mathbb{N})$ (and even much weaker theories) contain an isomorphic copy of $\mathbb{N}$ as an initial segment. Therefore, we can assume that all models we encounter satisfy $\mathbb{N} \subseteq \mathcal{M}$. After considering a concrete nonstandard model $\mathcal{M}$ (of $\text{Th}_{L_{\text{all}}}(\mathbb{N})$) we shall call the elements of $\mathcal{M}\backslash\mathbb{N}$ *nonstandard numbers*. These can be intuitively understood as infinite natural numbers. The key feature of those elements is that all functions and relations from L_{all} are defined even on nonstandard numbers. This includes functions for coding sequences and sets by numbers, therefore we can use notations like $a_0, \ldots, a_{n-1}$ even for a nonstandard number n. The notation then means that for each $i \in \mathcal{M}$ such that $i < n$ we have an object a_i coded by a number in $\mathcal{M}$ and that this whole sequence is coded by some number $\{a_i\}_{i=0}^{n-1} \in \mathcal{M}$. For a nonstandard number $S \in \mathcal{M}$ coding a set, we denote its nonstandard size (cardinality) by $|S|$. In the case where we talk about a binary string x, the notation $|x|$ denotes the length of x (which is nonstandard if x is), and if m is simply an element of $\mathcal{M}$ we denote its bit length by $|m|$. The symbol "∞" in sequences $\{a_i\}_{i=0}^{\infty}$ implies indexing over standard natural numbers, and in limits $\lim_{k\to\infty}(\ldots)$ has the standard meaning.

In the next section we will fix a nonstandard model $\mathcal{M}$ that has the model theoretic property of being $\aleph_1$-saturated. There is a self-contained construction of such a model in [11, Appendix], but we never explicitly use the $\aleph_1$-saturation. The reason we require this is that, without it, the Boolean algebra we construct might not be complete. A property, implied by our choice of language, that we do use explicitly is the following:

Property 1. For any sequence $\{a_i\}_{i=0}^{\infty}$ of standard natural numbers, there is a function symbol f in L_{all} such that for all $i \in \mathbb{N}$ we have $\mathcal{M} \models f(i) = a_i$. This gives rise to a sequence $\{f(i)\}_{i\in\mathcal{M}}$ that agrees with $\{a_i\}_{i=0}^{\infty}$ on elements with standard index, but also has elements with index i for any $i \in \mathcal{M}\backslash\mathbb{N}$. We call $\{f(i)\}_{i\in\mathcal{M}}$ the *nonstandard prolongation* of $\{a_i\}_{i=0}^{\infty}$ and for any $m \in \mathcal{M}$ we will denote $f(m)$ as simply a_m.

Using this property, we will now allow ourselves to freely use any elements of $\mathcal{M}$ as indices of sequences of standard numbers and generally of any sequences of standard finite objects, which can indeed be coded by standard natural numbers.

Any nonstandard model $\mathcal{M}$ can be extended to an ordered ring $\mathbb{Z}^{\mathcal{M}}$ by adding negative elements. This ring can then be extended to a fraction field $\mathbb{Q}^{\mathcal{M}}$. We shall call elements of $\mathbb{Q}^{\mathcal{M}}$ *$\mathcal{M}$-rationals*. The field $\mathbb{Q}^{\mathcal{M}}$ contains an isomorphic copy of $\mathbb{Q}$ as a substructure. We shall use the structure $\mathbb{Q}^{\mathcal{M}}$ analogously to how hyperreal numbers are used in nonstandard analysis. For more details about nonstandard analysis we recommend [5] to the interested reader. We call an element in $\mathbb{Q}^{\mathcal{M}}$ with absolute value greater than all $\frac{k}{1}, k \in \mathbb{N}$, *infinite*, otherwise we call it *finite*. We call elements in $\mathbb{Q}^{\mathcal{M}}$ with absolute value smaller than all $\frac{1}{k}, k \in \mathbb{N}$, *infinitesimal*. We will denote the set of finite $\mathcal{M}$-rationals as $\mathbb{Q}^{\mathcal{M}}_{\text{fin}}$ and one can check it forms an ordered ring.

Lemma 2 (The existence of a standard part, cf. Goldbring, [5, Theorem 1.9])**.** There is a function $\text{st}\colon \mathbb{Q}^{\mathcal{M}}_{\text{fin}} \to \mathbb{R}$ assigning to each finite $\mathcal{M}$-rational a real number. The function st is a ring homomorphism and the kernel of st is exactly the ideal of infinitesimal numbers.

When q is a finite $\mathcal{M}$-rational we call $\mathrm{st}(q)$ its *standard part*. The following result characterizes convergence of sequences of rational numbers using the $\mathcal{M}$-rational numbers $\mathbb{Q}^{\mathcal{M}}$.

Theorem 3 (cf. Goldbring, [5, Theorem 3.2])**.** Let $\{a_i\}_{i=0}^{\infty}$ be a sequence of standard rational numbers and let $r \in \mathbb{R}$. Then the following are equivalent.

(1) $\lim_{i\to\infty} a_i = r$,
(2) for every nonstandard $t \in \mathcal{M}\backslash\mathbb{N}$ we have $\mathrm{st}(a_t) = r$.

This theorem shows that computing a standard part of an $\mathcal{M}$-rational obtained from a nonstandard prolongation of some standard sequence is equivalent to computing the limit of this sequence. It will be apparent that computations emerging in the analysis of wide limits can be completed in a purely standard setting, by computing limits of specific probabilities associated with the wide sequence we consider. In this work, we mostly present the computations with nonstandard parameters. This seems to be natural in the context of our limit object and, in some sense, provides a corresponding shift in perspective: instead of directly analyzing the sequence we start with, we perform calculations with the nonstandard numbers that enumerate elements at a concrete nonstandard index of the sequence. A reader interested in interpreting those computations in the standard setting can in most cases simply imagine the manipulations to be prefixed with "for all sufficiently large standard n".

It is important for arithmetical forcing with random variables to consider discrete probability spaces of nonstandard size. We shall always use the uniform distribution on the samples, although this is not necessary for the general construction. Thus, the probability of an event coded by an element $A \in \mathcal{M}$ is then just the $\mathcal{M}$-rational number $|A| / |S|$ where S is the set of all samples.

We conclude this section by restating classical results of real analysis using the nonstandard approach. Their validity follows from the validity of their standard counterparts and Theorem 3.

Theorem 4 (Bernoulli's inequality)**.** Let $y \in \mathcal{M}$, $x \in \mathbb{Q}^{\mathcal{M}}$ and $x \geq -1$, then $(1+x)^y \geq 1 + yx$.

Theorem 5 (Exponential equality)**.** Let $x \in \mathcal{M}\backslash\mathbb{N}$, then $\mathrm{st}\big(\big(1-\frac{1}{x}\big)^x\big) = \mathrm{e}^{-1}$.

3 Wide limits

We shall define a wide limit of every sequence of the following form.

Definition 6. A sequence of sets of graphs $\{\mathcal{G}_k\}_{k=1}^{\infty}$ is called *a wide sequence* if the following holds:

(1) every graph $\omega \in \mathcal{G}_k$ has the vertex set $\{0, \dots, k-1\}$,
(2) $\lim_{k\to\infty} |\mathcal{G}_k| = \infty$.

By abuse of notation we will simply talk about a wide sequence $\mathcal{G}_k$ instead of $\{\mathcal{G}_k\}_{k=1}^{\infty}$. Since a wide limit is a Boolean-valued graph, we need to construct a Boolean algebra in which the truth evaluation of statements shall take place. For the construction of the Boolean algebra, we will closely follow [11, Chapter 1], albeit with slight changes. Let us fix now, for the rest of this work, an $\aleph_1$-saturated model of $\mathrm{Th}_{L_{\mathrm{all}}}(\mathbb{N})$, that we will denote by $\mathcal{M}$, and along with it we fix a nonstandard number $n \in \mathcal{M}$.

Definition 7. We define $\mathcal{A} = \{A \subseteq \{0, \dots, n-1\} \,;\, A \in \mathcal{M}\}$. In words, $\mathcal{A}$ is the set of subsets of $\{0, \dots, n-1\}$ coded by an element in $\mathcal{M}$. This is a Boolean algebra and to each $A \in \mathcal{A}$ we assign an $\mathcal{M}$-rational $|A|/n$, which we call its *counting measure*.

Even though $\mathcal{A}$ is a Boolean algebra with a measure, it is not a σ-algebra. Indeed, $\mathcal{A}$ contains all singletons $\{k\}$ for k standard, but the countable union of those singletons, the set of standard natural numbers $\mathbb{N}$, is not

definable by overspill. However, if we had joins and meets of arbitrary subsets at our disposal it would allow us to interpret quantifiers in the Boolean-valued case. So, we now want to tweak this Boolean algebra.

Definition 8. Let $\mathcal{I}$ be the ideal of $\mathcal{A}$ consisting of elements with infinitesimal counting measure. We define $\mathcal{B} = \mathcal{A}/\mathcal{I}$. Each element in $\mathcal{B}$ is of the form $A/\mathcal{I}$, where $A \in \mathcal{A}$, and we define $\mu(A/\mathcal{I}) = \mathrm{st}(|A|/n)$. We will denote the maximal element of $\mathcal{B}$ by $\mathbf{1}$ and the minimal element by $\mathbf{0}$.

One can easily check that μ is well-defined since for all $A \in \mathcal{I}$ it holds that $\mathrm{st}(|A|/n) = 0$. The measure μ is called the *Loeb measure*. Relying on the $\aleph_1$-saturation of $\mathcal{M}$, the following then holds.

Lemma 9 (Krajíček, [11, Lemma 1.2.1])**.** $\mathcal{B}$ is a σ-algebra with a real valued measure μ. Moreover, $\mathcal{B}$ is a complete Boolean algebra.

It is important to note that $\mathbf{1} \in \mathcal{B}$ is the only element of $\mathcal{B}$ with measure $\mu(\mathbf{1}) = 1$ and similarly $\mathbf{0} \in \mathcal{B}$ is the only element with measure $\mu(\mathbf{0}) = 0$. Also, for $B, B' \in \mathcal{B}$ the inequality $B \leq B'$ implies $\mu(B) \leq \mu(B')$.

We now define precisely what we mean by the family of functions F relative to which we will be taking the wide limit. This is still a part of Krajíček's construction, we just modify it to make it compatible with our setup that starts with a wide sequence.

For every $k \in \mathbb{N}$ the set $\mathcal{G}_k$ is finite and thus can be coded by a standard number. Therefore, there is a nonstandard prolongation of this sequence, and we can consider the set coded by the nonstandard number $\mathcal{G}_n$ that matches the value of the function symbol in L_{all} describing the function $k \mapsto \mathcal{G}_k$ when $k = n$.

Definition 10. Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be a wide sequence. We say that F is a *family of random variables on* $\mathcal{G}_n$ if every $\alpha \in F$ is a function coded by a number in $\mathcal{M}$ with domain $\mathcal{G}_n$ and taking values in $\mathcal{M}$. We say $\alpha \in F$ is an F-vertex if for all $\omega \in \mathcal{G}_n$ it holds that $\alpha(\omega) \in \{0, \ldots, n-1\}$. The set of all F-vertices is denoted by $U(F)$.

If the wide sequence $\{\mathcal{G}_k\}_{k=1}^{\infty}$ is clear from context we just say F is a family of random variables. This is for now everything we need to recall from [11], and we can proceed to define the central object of our work.

Definition 11 (The wide limit)**.** Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be a wide sequence and let F be a family of random variables on $\mathcal{G}_n$. We define the *wide limit* $\lim_F \mathcal{G}_n$ as a $\mathcal{B}$-valued structure in the language consisting of a single binary relation symbol $\{E\}$ as follows. The universe of the wide limit is taken as the set of all F-vertices. We now inductively define the truth values for all $\{E\}$-sentences.

(1) $[\![\alpha = \beta]\!] = \{\omega \in \mathcal{G}_n \ ; \ \alpha(\omega) = \beta(\omega)\}/\mathcal{I}$,
(2) $[\![E(\alpha, \beta)]\!] = \{\omega \in \mathcal{G}_n \ ; \ E_\omega(\alpha(\omega), \beta(\omega))\}/\mathcal{I}$,
(3) $[\![-]\!]$ commutes with $\neg$, $\wedge$ and $\vee$,
(4) $[\![(\exists x)A(x)]\!] = \bigvee_{\alpha \in U(F)} [\![A(\alpha)]\!]$,
(5) $[\![(\forall x)A(x)]\!] = \bigwedge_{\alpha \in U(F)} [\![A(\alpha)]\!]$.

To stress in which Boolean-valued structure the truth evaluation $[\![-]\!]$ is taking place we will sometimes denote the evaluation by $\mathcal{C}_1[\![-]\!]$, $\mathcal{C}_2[\![-]\!]$ for Boolean-valued structures $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Furthermore, if $\mathcal{C}_1[\![\varphi]\!] = \mathbf{1}$ for some sentence φ we say φ is *valid* in $\mathcal{C}_1$.

Note that since $\mathcal{G}_n$ can be recovered from F as the domain of its elements, the wide limit only depends on F, strictly speaking. We keep $\mathcal{G}_n$ in the notation to cover the situation where we have a very general family of functions (e.g., the family of polynomial functions F_{PV}) that can be applied to every wide sequence. Thus, the notation $\lim_F \mathcal{G}_n$ means that F is restricted to those functions that take elements of $\mathcal{G}_n$ as an input even when F possibly contains other functions too.

The potential variability of the parameter n may also seem unnecessary—and indeed in this section it is—but in Section 6 we will assume that n is a power of two, which will allow us to more easily translate between the results about wide limits and reducibility of search problems.

3.1 Wide limits for shallow decision trees

Now we shall define the first nontrivial family of random variables relative to which we shall take wide limits of several sequences. The functions in the family will be computed by shallow decision trees. So the shape of the wide limit reflects what a tree of depth subexponential in $|n|$ can witness in the wide sequence with probability arbitrarily close to 1.

Definition 12. Let $\mathcal{T}_{\mathrm{rud}}$ be a family of labeled rooted binary trees in $\mathcal{M}$ of the following form.[2] At each vertex the tree is labeled by an element of $\{0,\dots,n-1\}\times\{0,\dots,n-1\}$ and the two outgoing edges incident to it are labeled as 0 and 1 respectively. The leaves are labeled by an element of $\mathcal{M}$. The depth of the tree is bounded by a number of a form $n^{1/t}$ (rounded to the nearest element of $\mathcal{M}$) for some $t\in\mathcal{M}\backslash\mathbb{N}$.

A *computation* of a $T\in\mathcal{T}_{\mathrm{rud}}$ on some $\omega\in\mathcal{G}_n$ is defined as follows. Start at the root and interpret each label (i,j) of the vertex as a question whether the pair (i,j) is in the edge set E_ω and follow a path through T reading 1 as a positive answer and 0 as a negative answer. The label of the leaf visited at the end of the path is the output of T on ω, denoted by $T(\omega)$. We define F_{rud} to be the set of all functions computed by a tree $T\in\mathcal{T}_{\mathrm{rud}}$.

In the following example, we consider a simple wide sequence of sets of graphs with exactly one edge.

Example 13. Let $\mathrm{EDGE}_k=\{(\{0,\dots,k-1\},E)\ ;\ |E|=1\}$. Since any $\omega\in\mathrm{EDGE}_k$ has only 1 edge in all potential $k\cdot(k-1)/2$ edges, it is not likely a shallow tree will find the edge. This is the idea behind the proof of the claim $\lim_{F_{\mathrm{rud}}}\mathrm{EDGE}_n[\![(\exists x)(\exists y)E(x,y)]\!]=\mathbf{0}$.

Let $\alpha,\beta\in U(F_{\mathrm{rud}})$, we proceed by proving that $[\![E(\alpha,\beta)]\!]=\mathbf{0}$, which is enough to prove the claim since $[\![(\exists x)(\exists y)E(x,y)]\!]=\bigvee_{\alpha\in U(F_{\mathrm{rud}})}\bigvee_{\beta\in U(F_{\mathrm{rud}})}[\![E(\alpha,\beta)]\!]=\bigvee_{\alpha\in U(F_{\mathrm{rud}})}\bigvee_{\beta\in U(F_{\mathrm{rud}})}\mathbf{0}=\mathbf{0}$. Let α and β be computed by $T\in\mathcal{T}_{\mathrm{rud}}$ and $S\in\mathcal{T}_{\mathrm{rud}}$ respectively. Let the depth of both T and S be at most $n^{1/t}$, where $t\in\mathcal{M}\backslash\mathbb{N}$. Walk down T from the root and always prolong the path along the edge labeled 0. On this path we have a set of at most $n^{1/t}$ different pairs of vertices as the labels of edges and a label of the leaf node ℓ_T.

We do the same for S, and we find another set of at most $n^{1/t}$ pairs of vertices and a label of the leaf ℓ_S. The labels ℓ_S and ℓ_T are then combined to one last pair $\{\ell_S,\ell_T\}$. Now we just need to compute the probability that none of these $2n^{1/t}+1$ pairs of vertices are in the edge set E_ω.

There are $\binom{n}{2}$ different graphs in EDGE_n and at least $\binom{n-4n^{1/t}-2}{2}$ graphs that fulfill our requirements, namely, those graphs whose sole edge is not incident with the vertices from the labels of the trees S and T. By Theorem 4, the probability is at least

$$\begin{aligned}\frac{\binom{n-4n^{1/t}-2}{2}}{\binom{n}{2}}&=\frac{(n-4n^{1/t}-2)(n-4n^{1/t}-3)}{n(n-1)}\\&\geq 1-\frac{8n^{1/t}+6}{n}.\end{aligned}$$

After taking the standard part of the last line we get $\mathrm{st}\left(1-\frac{8n^{1/t}+6}{n}\right)=1$. Therefore, $\mu([\![E(\alpha,\beta)]\!])=0$ and $[\![E(\alpha,\beta)]\!]=\mathbf{0}$.

3.2 Sufficient conditions for validity of universal and existential sentences

To understand wide limits we need to compute the truth values of sentences that describe properties whose complexity we are interested in. Generally, for sentences of arbitrary complexity, this can be hard. In this section, we prove sufficient conditions at least for the validity of universal and existential sentences.

We will start with the simpler condition for the validity of universal sentences. This is important, also because we would like to know that a wide limit of a wide sequence of graphs is also a graph; the statement

[2] The subscript "rud" stands for rudimentary. The name for the family is taken from [11].

that E is antireflexive and symmetric is valid in the wide limit, and this statement is indeed expressible as a universal sentence.

Theorem 14. Let $\mathcal{G}_k$ be a wide sequence and let F be any family of random variables. Let $\varphi(x_0, \ldots, x_{\ell-1})$ be an open $\{E\}$-formula and assume that $\lim_{k\to\infty} \Pr_{\omega\in\mathcal{G}_k}[\omega \models (\forall\overline{x})\varphi(\overline{x})] = 1$. Then $\lim_F \mathcal{G}_n[\![(\forall\overline{x})\varphi(\overline{x})]\!] = \mathbf{1}$.

Proof. By the assumption and Theorem 3 we get that $\mathrm{st}(\Pr_{\omega\in\mathcal{G}_n}[\omega \models (\forall\overline{x})\varphi(x)]) = 1$. Since φ is open, we have for every tuple of F-vertices $\overline{\alpha}$ that $[\![\varphi(\overline{\alpha})]\!] = \mathbf{1}$. Now $[\![(\forall x)\varphi(x)]\!] = \bigwedge_{\overline{\alpha}\in U(F)^\ell} [\![\varphi(\overline{\alpha})]\!] = \bigwedge_{\overline{\alpha}\in U(F)^\ell} \mathbf{1} = \mathbf{1}$. □

Corollary 15. Let $\mathcal{G}_k$ be a wide sequence and F any family of random variables, then $\lim_F \mathcal{G}_n$ is an $\{E\}$-structure in which both antireflexivity and symmetry of E is valid (i.e., $\lim_F$ is a Boolean-valued graph). Similarly, if $\mathcal{G}_k$ is a wide sequence of digraphs and F any family of random variables, then $\lim_F \mathcal{G}_n$ is an $\{E\}$-structure in which antireflexivity is valid.

Now, to give a sufficient condition for the validity of an existential sentence $(\exists\overline{x})\varphi(\overline{x})$, we use the auxiliary value of *density of* $\varphi(x_0, \ldots, x_{\ell-1})$ defined as the probability that a random graph $\omega \in \mathcal{G}_k$ and a random tuple $\overline{a} \in \{0, \ldots, k-1\}^\ell$ satisfy $\omega \models \varphi(\overline{a})$ and show that the limiting density gives a lower bound for the measure of $[\![(\exists\overline{x})\varphi(\overline{x})]\!]$.

Theorem 16. Let $\mathcal{G}_k$ be a wide sequence and let F be a family of random variables that contains all constant functions. Let $\varphi(x_0, \ldots, x_{\ell-1})$ be an open $\{E\}$-formula and let $p \in [0, 1]$. Assume that

$$\lim_{k\to\infty} \Pr_{\substack{\omega\in\mathcal{G}_k \\ \overline{a}}} [\omega \models \varphi(\overline{a})] \geq p,$$

where $\overline{a}$ is sampled uniformly over all elements of $\{0, \ldots, k-1\}^\ell$. Then $\mu(\lim_F \mathcal{G}_n[\![(\exists\overline{x})\varphi(\overline{x})]\!]) \geq p$. In particular, if $p = 1$ then $\lim_F \mathcal{G}_n[\![(\exists\overline{x})\varphi(\overline{x})]\!] = \mathbf{1}$.

Proof. Consider an array C indexed by $\omega \in \mathcal{G}_n$ and $\overline{a} \in \{0, \ldots, n-1\}^\ell$ such that

$$C_{\omega,\overline{a}} = \begin{cases} 1 & \text{if } \omega \models \varphi(\overline{a}), \\ 0 & \text{otherwise.} \end{cases}$$

By the assumption and induction in $\mathcal{M}$ we have that $\mathrm{st}\big(\frac{1}{n^\ell|\mathcal{G}_n|} \sum_{\omega\in\mathcal{G}_n} \sum_{\overline{a}} C_{\omega,\overline{a}}\big) \geq p$. We now claim that there exists a specific $\overline{b} \in \{0, \ldots, n-1\}^\ell$ such that $\mathrm{st}\big(\Pr_{\omega\in\mathcal{G}_n}\big[\omega \models \varphi(\overline{b})\big]\big) \geq p$. Assume for contradiction that the claim is false. Then $\frac{1}{|\mathcal{G}_n|n^\ell} \sum_{\omega\in\mathcal{G}_n} \sum_{\overline{a}} C_{\omega,\overline{\alpha}} = \frac{1}{n^\ell} \sum_{\overline{a}} \Pr_{\omega\in\mathcal{G}_n}[\omega \models \varphi(\overline{a})] \leq \Pr_{\omega\in\mathcal{G}_n}[\omega \models \varphi(\overline{a}_0)]$, where we pick $\overline{a}_0$ such that it maximizes $\Pr_{\omega\in\mathcal{G}_n}[\omega \models \varphi(\overline{a}_0)]$.[3] But after taking the standard part of the inequality we obtain that $\mathrm{st}\big(\frac{1}{n^\ell|\mathcal{G}_n|} \sum_{\omega\in\mathcal{G}_n} \sum_{\overline{a}} C_{\omega,\overline{a}}\big) \leq \mathrm{st}(\Pr_{\omega\in\mathcal{G}_n}[\omega \models \varphi(\overline{a}_0)]) < p$, which is a contradiction and so the claim is true. Let $\overline{\gamma}_b$ be a tuple of constant functions that is at every sample equal to $\overline{b}$. We have $[\![(\exists\overline{x})\varphi(\overline{x})]\!] = \bigvee_{\overline{\alpha}\in U(F)^\ell} [\![\varphi(\overline{\alpha})]\!] \geq [\![\varphi(\overline{\gamma}_b)]\!]$ and by taking μ of this inequality we finally obtain that $\mu([\![(\exists\overline{x})\varphi(\overline{x})]\!]) \geq p$. □

The following example demonstrates that Theorem 14 cannot be generalized to a similar hypothesis as Theorem 16.

Example 17. Let $\mathcal{G}_k$ consist of all undirected graphs on the vertex set $\{0, \ldots, k-1\}$ with exactly $\left\lceil \frac{k(k-1)}{2\log(k)} \right\rceil$ edges. One can see that

$$\lim_{k\to\infty} \Pr_{\substack{\omega\in\mathcal{G}_k \\ x,y}} [\omega \models \neg E(x, y)] = 1,$$

but in fact $\lim_{F_{\mathrm{rud}}} \mathcal{G}_n[\![(\forall x)(\forall y)\neg E(x, y)]\!] = \mathbf{0}$.

[3] This is possible because $\mathcal{M}$, being a model of $\mathrm{Th}(\mathbb{N})$, satisfies induction.

Let $t \in \mathcal{M}\backslash\mathbb{N}$ such that $n^{1/t}$ is not bounded above by a standard number. Let T be a tree that queries on all paths a fixed set of $n^{1/t}$ different potential edges. If we prove that any such set in $\mathcal{G}_n$ has to contain at least one edge with probability infinitesimally close to 1 then we can construct F_{rud}-vertices α and β using T such that $[\![E(\alpha,\beta)]\!] = \mathbf{1}$ by simply taking T and labeling each leaf on a path that finds an edge with either the lesser vertex (in the natural order of $\mathcal{M}$) to compute α, or with the greater vertex to compute β.

Let S be the set of potential edges queried by T and let $m = \binom{n}{2}$. Now we have

$$\begin{aligned}
\Pr_{\omega\in\mathcal{G}_n}[S \text{ contains no edge in } \omega] &= \frac{\left(m-n^{1/t}\right)!\left(m-\left\lceil\frac{m}{\log n}\right\rceil\right)!}{m!\left(m-\left\lceil\frac{m}{\log n}\right\rceil-n^{1/t}\right)!} \\
&= \prod_{i=0}^{n^{1/t}-1} \frac{m-\left\lceil\frac{m}{\log n}\right\rceil-i}{m-i} \\
&\leq \left(1-\frac{\left\lceil\frac{m}{\log n}\right\rceil}{m}\right)^{n^{1/t}} \\
&\leq \left(1-\frac{1}{2\log n}\right)^{n^{1/t}},
\end{aligned}$$

the standard part of which, using Theorem 5, is bounded above for all $k \in \mathbb{N}$ by $\mathrm{st}\left(\left(1-\frac{1}{2\log n}\right)^{k\cdot 2\log n}\right) = \mathrm{e}^{-k}$, which tends to 0 as $k \to \infty$.

For more examples, we point the interested reader to [9].

4 Expanding wide limits to two-sorted arithmetic

In this section, we will show that under reasonable assumptions one can embed the wide limit into the original models of Krajíček in such a way that first-order statements, after a suitable translation, keep their truth values.

4.1 Two-sorted structures of weak arithmetic

We will now recall the construction of two-sorted models of weak arithmetic $K(F,G)$ defined in [11, Chapter 5]. We will take the liberty to define them as an extension of the definition of a wide limit to obtain structures $K(\mathcal{G}_n,F,G)$.[4] Under the right conditions, these result in a structure in some sublanguage of L_{all} with two sorts: numbers and functions of bounded domain on numbers. This latter sort contains the wide limit as an object. These sorts are to be named *number sort* and *set sort*, as the bounded functions can be interpreted as sets (or more generally relations) and every such function f can be coded by a bounded set $\{(x,i)\ ;\ \text{the } i\text{-th bit of } f(x) \text{ is } 1\}$.

Definition 18. Let $L \subseteq L_{\mathrm{all}}$. This determines a *language* L^2, which we get by keeping the original variables as number sort variables, adding to L set sort variables $X, Y, \ldots$, whose intended interpretation are bounded functions, and the equality symbol for set sort variables (denoted the same as for the number sort). All set sort variables are treated as function symbols and can form terms with the number sort terms as arguments.

We will also use the function sort variables as relation symbols, and we define the atomic formula $X(x_0,\ldots,x_{k-1})$ to be evaluated the same as the formula $X(x_0,\ldots,x_{k-1}) \neq 0$.

[4] This notation is just making some parameters of the construction explicit, the models constructed can be obtained by the original method without first constructing the wide limit. Our contribution is in observing that the truth values of first-order sentences concerning the wide limit is preserved between the wide limit and the structure $K(\mathcal{G}_n,F,G)$.

We will now fix a wide sequence $\mathcal{G}_k$ and a family of random variables F on $\mathcal{G}_n$, which together determine a wide limit $\lim_F \mathcal{G}_n$.

Definition 19. We define $\mathcal{M}_n \subseteq \mathcal{M}$ to be the subset of $\mathcal{M}$ consisting of all numbers bounded above by $2^{n^{1/t}}$ for some $t \in \mathcal{M}\backslash\mathbb{N}$.

We define $L_n \subseteq L_{\text{all}}$ to contain all relation symbols from L_{all} and all functions from L_{all} for which their values on any tuple of elements of $\mathcal{M}_n$ is still in $\mathcal{M}_n$. We say F is L_n-closed if for every function symbol $f \in L_n$ of arity k and every $\alpha_0, \ldots, \alpha_{k-1} \in F$ we have that $f(\alpha_0, \ldots, \alpha_{k-1}) \in F$.

Note that $\mathcal{M}_n$ is a substructure of the L_n-reduct of $\mathcal{M}$.

Definition 20. We say that G is a *family of random functions* (on $\mathcal{G}_n$) if for every $\Theta \in G$ there is $k \in \mathbb{N}$ such that Θ assigns to each $\omega \in \mathcal{G}_n$ a k-ary function Θ_ω coded by an element in $\mathcal{M}$ that maps a set $\mathrm{dom}(\Theta_\omega) \subseteq \mathcal{M}_n$ into $\mathcal{M}_n$. Such Θ is then called k-ary. We say G is *F-compatible* if for every $\alpha_0, \ldots, \alpha_{k-1} \in F$, and a k-ary $\Theta \in G$ we have that the function $\Theta(\alpha_0, \ldots, \alpha_{k-1})$ is in F, where

$$\Theta(\alpha)(\omega) = \begin{cases} \Theta_\omega(\alpha_0(\omega), \ldots, \alpha_{k-1}(\omega)) & \text{if } (\alpha_0(\omega), \ldots, \alpha_{k-1}(\omega)) \in \mathrm{dom}(\Theta_\omega), \\ 0 & \text{otherwise.} \end{cases}$$

An example of a specific family of random functions will be provided in Section 6.

Definition 21. Let F be an L_n-closed family of random variables with values in $\mathcal{M}_n$. Let G be an F-compatible family of random functions. We define $K(\mathcal{G}_n, F, G)$ to be a $\mathcal{B}$-valued L_n^2-structure with the number sort of the universe as F and the set sort of the universe as G. The valuation of formulas is then given by the following inductive definition. We define the valuation only for terms consisting of variables substituted with elements of F, and because of the L_n-closedness and F-compatibility of G, we can evaluate more complex terms step-by-step, eventually reaching an element of F.

(1) $[\![\alpha = \beta]\!] = \{\omega \in \mathcal{G}_n \ ; \ \alpha(\omega) = \beta(\omega)\}/\mathcal{I}$, where $\alpha, \beta \in F$,
(2) $[\![R(\alpha_0, \ldots, \alpha_{k-1})]\!] = \{\omega \in \mathcal{G}_n \ ; \ \mathcal{M}_n \models R(\alpha_0(\omega), \ldots, \alpha_{k-1}(\omega))\}/\mathcal{I}$, where the elements $\alpha_0, \ldots, \alpha_{k-1}$ are from F and R is a relation symbol in L_n,
(3) $[\![\Theta = \Xi]\!] = \{\omega \in \mathcal{G}_n \ ; \ \Theta_\omega = \Xi_\omega\}/\mathcal{I}$, where $\Theta, \Xi \in G$
(4) $[\![-]\!]$ commutes with $\neg$, $\wedge$ and $\vee$,
(5) $[\![(\forall x)A(x)]\!] = \bigwedge_{\alpha \in F} [\![A(\alpha)]\!]$,
(6) $[\![(\exists x)A(x)]\!] = \bigvee_{\alpha \in F} [\![A(\alpha)]\!]$,
(7) $[\![(\forall X)A(X)]\!] = \bigwedge_{\Theta \in G} [\![A(\Theta)]\!]$,
(8) $[\![(\exists X)A(X)]\!] = \bigvee_{\Theta \in G} [\![A(\Theta)]\!]$.

An important detail of this definition is that now in the evaluation of the first-order quantifiers we take suprema and infima over the whole F, rather than just $U(F)$. This is to allow interpretation of increasing functions such as multiplication and addition in the model. Let us also note that, in general, it is possible that extensionality is not valid in $K(\mathcal{G}_n, F, G)$, as the formula $(\forall x)(\Theta(x) = \Xi(x))$ could be valid even if for every ω we had $\Theta_\omega \neq \Xi_\omega$.

4.2 Preservation of sentences concerning the wide limit

We will now prove (under a mild condition on F, cf. Definition 23) that valid first-order properties of $\lim_F \mathcal{G}_n$ remain valid for a set sort object in $K(\mathcal{G}_n, F, G)$ that encodes the wide limit $\lim_F \mathcal{G}_n$. This lets us construct models in which an object with the same first-order properties as the wide limit might be desired. Recall that every element $\Theta \in G$, where G is a family of random functions, actually determines a predicate that we evaluate the same as the formula $\Theta(\alpha_0, \ldots, \alpha_{k-1}) \neq 0$.

Definition 22. Let G be a family of random functions. We say that the edge relation of the wide limit $\lim_F \mathcal{G}_n$ is *represented in G by Γ* if $\Gamma \in G$ is binary and for all $\alpha, \beta \in U(F)$ we have that $K(\mathcal{G}_n, F, G)[\![\Gamma(\alpha,\beta)]\!] = \lim_F \mathcal{G}_n[\![E(\alpha,\beta)]\!]$.

Definition 23. We say a family of random variables F has *restrictable ranges* if for every $\alpha \in F$ and $m \in \mathcal{M}_n$ there is $\tilde{\alpha}_m \in F$ such that

$$\tilde{\alpha}_m(\omega) = \begin{cases} \alpha(\omega) & \text{if } \alpha(\omega) < m, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 24. Let φ be a $\{E\}$-sentence. Let F be L_n-closed and have restrictable ranges and let G be F-compatible. Let the edge relation of the wide limit $\lim_F \mathcal{G}_n$ be represented in G by Γ. We define $\tilde{\varphi}(\Gamma)$ to be the L_n^2-sentence obtained by replacing all the occurrences of the relation symbol E by Γ, keeping the structure of the logical connectives and replacing all quantifiers $(\forall x)(\dots)$ by $(\forall x)(x < n \to (\dots))$ and $(\exists x)(\dots)$ by $(\exists x)(x < n \wedge \dots)$. Then $\lim_F \mathcal{G}_n[\![\varphi]\!] = K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma)]\!]$.

Proof. We will prove that for all $\{E\}$-formulas $\varphi(\overline{x})$ and all $\overline{\alpha} \in U(F)$ we have that $\lim_F \mathcal{G}_n[\![\varphi(\overline{\alpha})]\!] = K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \overline{\alpha})]\!]$. We proceed by induction on the complexity of the formula. The case for atomic formulas follows from Definition 22 and the step for logical connectives is also clear, since $[\![-]\!]$ commutes with them. With the induction step for negation in hand, it is now enough to prove the induction step for the universal quantifier.

We now assume that the statement holds for a formula $\varphi(y, \overline{x})$. By the restrictability of ranges in F we get that for all $\beta \in F$ there is $\tilde{\beta}_n \in U(F)$ such that

$$\tilde{\beta}_n(\omega) = \begin{cases} \beta(\omega) & \text{if } \beta(\omega) < n, \\ 0 & \text{otherwise.} \end{cases}$$

For every $\beta \in F$ we have $K(\mathcal{G}_n, F, G)[\![\beta < n \to \beta = \tilde{\beta}_n]\!] = \mathbf{1}$ and by the validity of predicate logic we have $K(\mathcal{G}_n, F, G)[\![\beta < n \to (\tilde{\varphi}(\Gamma, \beta, \overline{\alpha}) \equiv \tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha}))]\!] = \mathbf{1}$, which together implies

$$K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha}) \to (\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha}))]\!] = \mathbf{1},$$

which can be rewritten as

$$K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!] \leq K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!]. \tag{1}$$

Moreover, for every $\gamma \in U(F)$ we have $\tilde{\gamma}_n = \gamma$ and thus every element of $U(F)$ is of the form $\tilde{\beta}_n$ for some β.

Claim 25. For all $\overline{\alpha} \in U(F)$ we have:

$$\bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!] = \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!].$$

Proof of Claim 25. Note that from Equation (1) we obtain

$$\bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!] \geq \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!].$$

For the second direction, using the fact that $U(F)$ contains exactly all $\tilde{\beta}_n$, we have

$$\begin{aligned} \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!] &\leq \bigwedge_{\beta \in U(F)} K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!] \\ &= \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\tilde{\beta}_n < n \to \tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!] \\ &= \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!], \end{aligned}$$

which proves Claim 25. ∎

With the claim established, we can finish the inductive step for the universal quantifier. Again using that $U(F)$ consists exactly of all $\tilde{\beta}_n$ and the induction hypothesis, we have that for all $\overline{\alpha} \in U(F)$

$$\begin{aligned} K(\mathcal{G}_n, F, G)[\![(\forall y < n)\tilde{\varphi}(\Gamma, y, \overline{\alpha})]\!] &= \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!] \\ &= \bigwedge_{\beta \in F} K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!] \\ &= \bigwedge_{\gamma \in U(F)} K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \gamma, \overline{\alpha})]\!] \\ &= \bigwedge_{\gamma \in U(F)} \lim_F \mathcal{G}_n[\![\varphi(\gamma, \overline{\alpha})]\!] \\ &= \lim_F \mathcal{G}_n[\![(\forall y)\varphi(y, \overline{\alpha})]\!]. \end{aligned}$$

□

5 Total NP search problems and wide limits

In this section we recall the class of search problems called Total NP Search problems, first defined in [16], and then take a wide limit of all instances of WeakPigeon and show that when it is intepreted as an instance, it has no solution.

5.1 Preliminaries on total NP search problems

Definition 26. A total NP search problem $P(x, y)$ is a relation on binary strings for which the following two conditions are satisfied.

(1) *Verifiability in polynomial time:* There exists a Turing machine M deciding whether $P(x, y)$ is time polynomial in the lengths of x and y.
(2) *Totality:* There is a $c \in \mathbb{N}$ such that for every $x \in \{0,1\}^*$ there is $y \in \{0,1\}^*$ such that $|y|$ is at most $(|x| + 2)^c$ and $P(x, y)$.

The class of all such search problems is denoted by TFNP.

There are many well-known TFNP problems, some of which were considered already in [16], such as Leaf, which intuitively amounts to the task "given a graph ω with an odd-degree vertex, find another," or Pigeon, intuitively described as "given a map $f\colon [2^k] \to [2^k - 1]$, find distinct x and y such that $f(x) = f(y)$," where $[m]$ denotes the set $\{0, \ldots, m-1\}$. The graph ω in the first problem is given by a circuit computing potential neighbors for each vertex, and the function f in the second problem is given by a circuit computing its values. This makes the problems nontrivial, as checking the whole ω or f may take exponential time in the size of the given circuit.

More relevant for our setting is the variant of the class called *black-box* TFNP, originally defined as "type 2 TFNP" in [1]. This variant allows the circuits in the input to be replaced by an oracle. For example, in the problem Leaf, we instead obtain oracle access to the neighbor-sets of G; or, in the problem Pigeon, to the values of the function f. We will start by defining a query tree, which is a natural computational model generalizing an oracle Turing machine. Query trees capture exactly the relationship between the values of the oracle and the computational states of the model. Usually such trees operate on relational oracles, see for example [6], but in our setting function oracles are more natural.[5] As in [6], we will allow problems whose solutions can be checked by non-uniform query trees in the definition of black-box TFNP.

[5] Moreover, function oracles with values polynomially bounded in the size of the input can be simulated by relational ones, which can be queried for the i-th bit of the given function value. Every query tree operating on a function oracle then can be transformed to a tree operating on a relational oracle with at most polynomially larger depth.

Definition 27. Let $k_q, k_a, k_o > 0$. A (k_q, k_a, k_o)-*query tree* T is a labeled rooted tree of the following shape:

(1) each non-leaf node is labeled by a binary string of length k_q,
(2) each non-leaf node has for every $w \in \{0,1\}^{k_a}$ an outgoing edge labeled by w,
(3) each leaf node is labeled by a binary string of length k_o.

The *depth* of a tree T is the length of the longest path starting at the root. If $\mathcal{O}$ is a function that maps $\{0,1\}^{k_q}$ to $\{0,1\}^{k_a}$, then the *computation* of T on $\mathcal{O}$ is the path obtained by starting at the root and continuing down the tree from each non-leaf node with some label ℓ to the next node along the edge labeled by $\mathcal{O}(\ell)$. The *output* of the computation $T(\mathcal{O})$ is then simply the label of the leaf met on the computation.

We will also define the *computation* of T on any $\mathcal{O}\colon \{0,1\}^{k_q} \to \{0,1\}^*$ as the computation of T on

$$\mathcal{O}'(x) = \begin{cases} x0^{k_a - |x|} & \text{if } |x| < k_a, \\ \text{the first } k_a \text{ bits of } x & \text{otherwise.} \end{cases}$$

In words, for functions whose output length is not neccessarily k_a we interpret the computation by either padding or truncating the outputs of $\mathcal{O}_{k_q}$.

For a function oracle $\mathcal{O}\colon \{0,1\}^* \to \{0,1\}^*$ and a number k we denote the restriction of $\mathcal{O}$ to $\{0,1\}^k$ as $\mathcal{O}_k$, and define the computation of a (k_q, k_a, k_o)-query tree T on $\mathcal{O}$ as the computation of T on $\mathcal{O}_{k_q}$.

Let us stress that we allow a (k_q, k_a, k_o)-query tree T to query a function oracle $\mathcal{O}$ only on strings of length k_q, as instances of query TFNP problems will be presented on a set of all strings of a given length, and thus the other values of the oracle are not relevant.

Definition 28. A *total query NP search problem* $P(\mathcal{O}, x, y)$ is a relation, where $\mathcal{O}$ is a function oracle, x and y are binary strings, along with two functions ℓ_a and ℓ_s such that the following three conditions are satisfied.

(1) *Conditions on lengths:* The functions ℓ_a and ℓ_s are polynomial time when the input and the output are given in unary, and thus are bounded by a polynomial. For every $x, y \in \{0,1\}^*$ we have that $P(\mathcal{O}, x, y)$ implies $|y| = \ell_s(|x|)$.
(2) *Verifiability in polynomial depth:* There is a polynomial p, and for any binary strings x and y there exists a $(|x|, \ell_a(|x|), 1)$-query tree $T_{x,y}$ of depth at most $p(|x|)$, such that for every $\mathcal{O}$ we have $T_{x,y}(\mathcal{O}) = 1$ if and only if $P(\mathcal{O}, x, y)$.
(3) *Totality:* For each $\mathcal{O}$ and $x \in \{0,1\}^*$ there is $y \in \{0,1\}^{\ell_s(|x|)}$ such that $P(\mathcal{O}, x, y)$.

Any pair $(\mathcal{O}, x)$ of an oracle and a binary string is called an *instance of* P, the string x is called the *size parameter*, and if $P(\mathcal{O}, x, y)$, then y is called a *solution* of the instance $(\mathcal{O}, x)$. The class of all such search problems is denoted by $\mathsf{TFNP}^{\mathrm{dt}}$.

We will be analyzing the following two $\mathsf{TFNP}^{\mathrm{dt}}$ problems. We will provide specific output lengths for the oracles in their definitions, but their instances could, in principle, be given by any oracle; this is done simply to emphasize how the query trees checking the validity of solutions would truncate or pad the output values of the oracle.

Definition 29. The problem WeakPigeon is given as follows. Given x and a function oracle $\mathcal{O}$ with $\mathcal{O}_{|x|}\colon \{0,1\}^{|x|} \to \{0,1\}^{|x|-1}$, find distinct $x_1, x_2 \in \{0,1\}^{|x|}$ such that $\mathcal{O}(x_1) = \mathcal{O}(x_2)$.

This problem is total, because the size of the domain on a given length is larger than the size of the codomain. Let us now assume that the input of a $\mathsf{TFNP}^{\mathrm{dt}}$ problem can be given by multiple function oracles, since these oracles can be represented as a single oracle computing all their values in parallel.

Definition 30. The problem RetractionWeakPigeon is given as follows. Given x and function oracles $\mathcal{O}^f$ and $\mathcal{O}^g$ such that $\mathcal{O}^f_{|x|}\colon \{0,1\}^{|x|} \to \{0,1\}^{|x|-1}$ and $\mathcal{O}^g_{|x|-1}\colon \{0,1\}^{|x|-1} \to \{0,1\}^{|x|}$, find $x' \in \{0,1\}^{|x|}$ satisfying that $\mathcal{O}^g\big(\mathcal{O}^f(x')\big) \neq x'$.

The problem is total as a consequence of totality of WeakPigeon: if we have distinct x_1, x_2 such that $\mathcal{O}^f(x_1) = \mathcal{O}^f(x_2)$, then one of them already has to be a solution to RetractionWeakPigeon. This observation can be made precise using the notion of a many-one reduction, which we will now recall. Our notion of a many-one reduction is a non-uniform version of the original definition from [1] and is analogous to the one using decision trees in [6, Definition 4]. The intuition is that P being many-one reducible to Q means that we can use an instance $(\mathcal{O}, x)$ of P to compute an instance $(\mathcal{O}', z)$ of Q such that each solution to $(\mathcal{O}', z)$ can be converted to a solution of $(\mathcal{O}, x)$, where all computations are performed by trees of depth polynomial in $|x|$.

Definition 31. Let P and Q be $\mathsf{TFNP}^{\mathrm{dt}}$ problems such that the length functions for P are ℓ_{a}^P and ℓ_{s}^P and for Q they are ℓ_{a}^Q and ℓ_{s}^Q. We say that P is many-one reducible to Q, denoted by $P \leq_{\mathrm{m}} Q$, if the following conditions are met.

(1) There is a function r that accepts as input a pair $(\mathcal{O}, x)$ consisting of a function oracle and a binary string and outputs a binary string z. The function is computed for each x by a query tree T_x^r of depth polynomial in $|x|$ such that $r(\mathcal{O}, x) = T_x^r(\mathcal{O})$.[6]
(2) For every x and z, there are sequences $\left(T^{\mathrm{I}}_{x'}\right)_{x' \in \{0,1\}^{|z|}}$ of $\left(|x|, \ell_{\mathrm{a}}^P(|x|), \ell_{\mathrm{a}}^Q(|z|)\right)$-query trees and $\left(T^{\mathrm{s}}_{y'}\right)_{y' \in \{0,1\}^{\ell_{\mathrm{s}}^Q(|z|)}}$ of $\left(|x|, \ell_{\mathrm{a}}^P(|x|), \ell_{\mathrm{s}}^P(|x|)\right)$-query trees, and the depth of all of these trees is polynomial in $|x|$.
(3) For every instance $(\mathcal{O}, x)$ and $z = r(\mathcal{O}, x)$, consider the function oracle $\mathcal{O}'$ satisfying that $\mathcal{O}'_{|z|}(x')$ is given by the value $T^{\mathrm{I}}_{x'}(\mathcal{O})$, obtained from the trees given by the previous point. We require that $\forall y' \in \{0,1\}^{\ell_{\mathrm{s}}^Q(|z|)} : Q(\mathcal{O}', z, y') \implies P\big(\mathcal{O}, x, T^{\mathrm{s}}_{y'}(\mathcal{O})\big)$.

It is easy to check that RetractionWeakPigeon $\leq_{\mathrm{m}}$ WeakPigeon. Regarding the other direction, it follows from known results (see Section 7 for details) that WeakPigeon $\not\leq_{\mathrm{m}}$ RetractionWeakPigeon; we will give a new proof of this using wide limits in the remainder of Sections 5 and 6.

5.2 The wide limit of all instances of WeakPigeon

The following wide sequence essentially consists of all instances of WeakPigeon. We will show that, relative to trees of subexponential depth that are allowed to ask for neighbors of vertices, the wide limit will be a graph of an injective function.

Definition 32. Let $M_{k,\lfloor k/2 \rfloor}$ be a wide sequence consisting of all $\{E\}$-structures on $\{0, \ldots, k-1\}$, where E is a graph of a function from $\{0, \ldots, k-1\}$ to $\{0, \ldots, \lfloor k/2 \rfloor - 1\}$.

Note that, because a graph of a function is naturally a digraph, this is actually a wide sequence of digraphs. For a fixed $\omega \in M_{n,\lfloor n/2 \rfloor}$, the neighbor of $i \in \{0, \ldots, n-1\}$ in ω is simply the image of i in the function whose graph is ω.

Definition 33. We define $\mathcal{T}_{\mathrm{nb}}$ as the set of all labeled rooted trees of the following shape:

(1) each non-leaf node is labeled by some $v \in \{0, \ldots, n-1\}$,
(2) for each $u \in \{0, \ldots, \lfloor n/2 \rfloor - 1\}$ and node a there is an outgoing edge from a labeled u,
(3) each leaf is labeled by some $m \in \mathcal{M}_n$,
(4) the *depth* of the tree is defined as the maximal number of edges in a path from the root, and we require it is at most $n^{1/t}$ (rounded to the nearest element of $\mathcal{M}$) for some $t \in \mathcal{M} \backslash \mathbb{N}$.

The *computation* of such a tree in $\mathcal{T}_{\mathrm{nb}}$ on $\omega \in M_{n,\lfloor n/2 \rfloor}$ is defined as follows. We build a path by starting at the root and, for every node labeled by some v, we follow the output edge with the answer to the question "what is the neighbor of vertex v?" We continue analogously until we find a leaf. The label of the leaf is defined to be the *output* of the computation.

[6] The tree can be formally taken to be a $(|x|, \ell_{\mathrm{a}}^P(|x|), p(|x|)))$-query tree, where p is a function bounded by a polynomial and the output of the tree is then interpreted by ignoring leading zeroes to allow outputs of varying lengths.

We define F_{nb} to be the set of all functions on $M_{n,\lfloor n/2\rfloor}$ that are computed by some $T \in \mathcal{T}_{\mathrm{nb}}$. Note that the depth of the trees is subexponential in $|n|$.

Definition 34. We say a tree $T \in \mathcal{T}_{\mathrm{nb}}$ *fails* on $\omega \in M_{n,\lfloor n/2\rfloor}$ if, on the computation path of the tree T, the neighbors of all distinct vertices are distinct, and if $T(\omega) \in \{0,\dots,n-1\}$ then also the neighbor of $T(\omega)$ in ω is disjoint from the other neighbors.

Lemma 35. Let $T \in \mathcal{T}_{\mathrm{nb}}$, then $\mathrm{st}\big(\Pr_{\omega\in M_{n,\lfloor n/2\rfloor}}[T \text{ fails}]\big) = 1$.

Proof. By direct computation, we have that the probability of failure of tree of depth d is

$$\Pr_{\omega\in M_{n,\lfloor n/2\rfloor}}[T \text{ fails}] \geq \prod_{i=1}^{d}\left(1-\frac{i}{\lfloor n/2\rfloor}\right) \geq \left(1-\frac{d}{\lfloor n/2\rfloor}\right)^{d} \geq 1-\frac{d^2}{\lfloor n/2\rfloor},$$

where the last inequality follows from Theorem 4. Since the depth of any tree in $\mathcal{T}_{\mathrm{nb}}$ is bounded by $n^{1/t}$ for some $t \in \mathcal{M}\backslash\mathbb{N}$, we have that the lower bound is at least $1-((n^{2/t})/\lfloor n/2\rfloor)$, which is infinitesimally close to 1. □

Theorem 36. $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}[\![(\exists x)(\exists y)(\exists z)(x\neq y\wedge E(x,z)\wedge E(y,z))]\!] = \mathbf{0}$.

Proof. For contradiction assume that $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}[\![(\exists x)(\exists y)(\exists z)(x\neq y\wedge E(x,z)\wedge E(y,z))]\!] > \mathbf{0}$, then there are $\alpha,\beta,\gamma \in U(F_{\mathrm{nb}})$ such that there is $p>0$ and $\mu(\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}[\![(\alpha\neq\beta\wedge E(\alpha,\gamma)\wedge E(\beta,\gamma))]\!]) = p$, therefore, as the evaluated sentence is quantifier free, we have

$$\mathrm{st}\left(\Pr_{\omega\in M_{n,\lfloor n/2\rfloor}}[(\alpha(\omega)\neq\beta(\omega)\wedge E_\omega(\alpha(\omega),\gamma(\omega))\wedge E_\omega(\beta(\omega),\gamma(\omega)))]\right) = p. \tag{2}$$

Assume that $T_\alpha, T_\beta, T_\gamma \in \mathcal{T}_{\mathrm{nb}}$ are the trees computing α,β and γ. Let us consider the tree $T \in \mathcal{T}_{\mathrm{nb}}$ that determines the values $\alpha(\omega)$, $\beta(\omega)$ and $\gamma(\omega)$ on each path; such a tree can be obtained by taking T_γ and replacing every leaf by a copy of the tree T_α and extending its leaves by a copy of T_β through every possible labeled edge, effectively querying the value $\alpha(\omega)$ on each path. By Equation (2) the fraction of paths of T where the tree asks for neighbors of $\alpha(\omega)$ and $\beta(\omega)$ and obtains $\gamma(\omega)$ is infinitesimally close to p and thus the probability of failure of T is not infinitesimally close to 1. This is in contradiction with Lemma 35. □

6 A separation model

In this section, we will expand $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}$ to the model $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$, and show that RetractionWeakPigeon is total, but the problem WeakPigeon is not. We then show that this implies nonexistence of a many-one reduction from WeakPigeon to RetractionWeakPigeon.

6.1 Construction of the model

Let us start by assuming for the rest of this work that n is a power of 2; this will allow us to easily convert between sets of binary strings and numbers. In the models $\mathcal{M}_n$ we are working with, there is a pairing function $\langle i,j\rangle$ that codes pairs of numbers by a single number. Thus, we can represent functions of any finite arity by unary functions in $\mathcal{M}_n$. We use this to define the family G_{nb}.

Definition 37. We define G_{nb} to be the family of all random functions on $M_{n,\lfloor n/2\rfloor}$ that fulfill the following. For each unary $\Theta \in G_{\mathrm{nb}}$ there exists a tuple $(\gamma_0,\dots,\gamma_{m-1})$ coded by an element of $\mathcal{M}$ such that $\gamma_i \in F_{\mathrm{nb}}$ and, for $\alpha \in F_{\mathrm{nb}}$,

$$\Theta(\alpha)(\omega) = \begin{cases} \gamma_{\alpha(\omega)}(\omega) & \text{if } \alpha(\omega) < m, \\ 0 & \text{otherwise.} \end{cases}$$

For every $k \in \mathbb{N}, k > 1$ and k-ary $\Theta \in G_{\mathrm{nb}}$ there is a $(k-1)$-ary $\Theta' \in G_{\mathrm{nb}}$ such that for every $\alpha_0, \ldots, \alpha_{k-1} \in F_{\mathrm{nb}}$ we have $\Theta(\alpha)(\omega) = \Theta'(\alpha_0, \ldots, \langle \alpha_{k-2}, \alpha_{k-1} \rangle)(\omega)$.

One can also regard k-ary functions from G_{nb} as those computed by k-dimensional tuples (that is tuples indexed by k-tuples) of elements of F_{nb}. To further explain the formalism, if we have unary $\Theta \in G_{\mathrm{nb}}$ and some $\omega \in M_{n,\lfloor n/2 \rfloor}$, the function Θ_ω is the function determined by the tuple $(\gamma_0(\omega), \ldots, \gamma_{m-1}(\omega))$ of elements of $\mathcal{M}_n$ with the property that $i \mapsto \gamma_i(\omega)$ if $i < m$. Random function families defined using tuples of random variables are used frequently in [11] and, more importantly, they generalize functions computable by an oracle Turing machine. This allows us to obtain nonexistence of many-one reducibility to RetractionWeakPigeon in Theorem 51.

Definition 38. We define Γ, an element of G_{nb}, as the random function computed by the tuple $(\gamma_{\langle i,j \rangle})_{i,j=0}^{n-1}$, where $\gamma_{\langle i,j \rangle}$ is computed by a tree $T_{i,j} \in \mathcal{T}_{\mathrm{nb}}$ of depth 1 that queries the neighbor of i and outputs 1 if it is j, and otherwise outputs 0.

Lemma 39.

(1) F_{nb} has restrictable ranges,
(2) F_{nb} is L_n-closed,
(3) G_{nb} is F_{nb}-compatible,
(4) the edge relation of the wide limit $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2 \rfloor}$ is represented in G_{nb} by Γ.

Proof. (1), (2) Here we can proceed simply by relabeling the leaves of the trees computing the functions from F_{nb}.

(3) Assume that $\Theta \in G_{\mathrm{nb}}$ is computed by a tuple $(\gamma_0, \ldots, \gamma_{m-1})$ and that the depth of all trees computing γ_i is at most $n^{1/t}$ for some $t \in \mathcal{M}_n \backslash \mathbb{N}$. For all $\alpha \in F_{\mathrm{nb}}$, let the tree T be constructed from the tree that computes α by appending the tree γ_i to each leaf with label i, then T has depth at most $n^{1/t'}$ for some $t' \in \mathcal{M}_n \backslash \mathbb{N}$ and therefore $T \in \mathcal{T}_{\mathrm{nb}}$. The tree T computes $\Theta(\alpha)$ by the definition of G_{nb}. Hence, G_{nb} is F_{nb}-compatible.

(4) By the definition of Γ and $\mathcal{T}_{\mathrm{nb}}$, we have for every $\alpha, \beta \in U(F_{\mathrm{nb}})$ that $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![\Gamma(\alpha, \beta)]\!] = \lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2 \rfloor}[\![E(\alpha, \beta)]\!]$. □

The model that we will analyze in the rest of this section is $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$.

6.2 Non-totality of WeakPigeon

In this section, we will show that the formalization of the statement "the problem WeakPigeon is total" is not true in the model $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$. Since wide limits are defined on intervals of (nonstandard) numbers and $\mathsf{TFNP}^{\mathrm{dt}}$ problems are defined on sets of binary strings, let us describe how these sets correspond to each other in our formalized statement. The input oracles will be represented by elements of the set sort and each set of binary strings $\{0,1\}^t$, where $t \in \mathcal{M}_n$, will be identified with the interval $\{0, \ldots, 2^t - 1\}$. Since n is a power of two, the interval $\{0, \ldots, n-1\}$ reserved for the values of F_{nb}-vertices can be identified with $\{0,1\}^{|n|-1}$. We also obtain the bijection $\{0, \ldots, n-1\} \cong \{t \in \mathcal{M}_n \ ; \ |t| < |n|\}$.

We will now define a formula $\varphi_{\textsc{WeakPigeon}}(X, m, x_1, x_2)$ that is a formalization of the statement "the values x_1 and x_2 are a solution to the WeakPigeon instance (X, m)." It is more natural in the arithmetic setting to accept as instances functions of arbitrary range and then allow inputs mapped outside $R_m = \{s \in \mathcal{M}_n \ ; \ |s| < |m| - 1\}$ as solutions. This variant is many-one reducible to the original problem, as we can remap the inputs mapped outside R_m to a fixed value in R_m, and the original problem is many-one reducible to this variant by a trivial reduction.

Definition 40. Let $\varphi^0_{\textsc{WeakPigeon}}(X, m, x_1, x_2)$ be the following L^2_n-formula:

$$(|X(x_1)| \geq |m| - 1) \vee (|X(x_2)| \geq |m| - 1) \vee (x_1 \neq x_2 \wedge X(x_1) = X(x_2)).$$

Let $\varphi_{\textsc{WeakPigeon}}(X, m, x_1, x_2)$ be the following L^2_n-formula:

$$(m = 0) \vee \big(|x_1| < |m| \wedge |x_2| < |m| \wedge \varphi^0_{\textsc{WeakPigeon}}(X, m, x_1, x_2)\big).$$

Theorem 41. $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall X)(\forall m)(\exists x_1)(\exists x_2)\varphi_{\textsc{WeakPigeon}}(X, m, x_1, x_2)]\!] = \mathbf{0}$.

Proof. Let us fix $\Theta \in G_{\mathrm{nb}}$ that is computed by the tuple $(\vartheta_0, \dots, \vartheta_{n-1})$, where the element ϑ_i is computed by the depth one tree, whose root is labeled i and each leaf is labeled the same as its corresponding edge.

Claim 42. $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall x < n)(\forall y < n)((\Theta(x) = y) \equiv \Gamma(x, y))]\!] = \mathbf{1}$.

Proof of Claim 42. Since the sentence is universal, it is enough to check its validity on every $\omega \in M_{n,\lfloor n/2\rfloor}$, which in turn follows from the definitions of Θ and Γ. ■

Claim 43. $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall x < n)(\forall y < n)(\forall z < n)(x = y \vee \neg\Gamma(x, z) \vee \neg\Gamma(y, z))]\!] = \mathbf{1}$.

Proof of Claim 43. This follows from Theorem 36, which shows the validity of the corresponding sentence in the wide limit $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2\rfloor}$, Lemma 39 and Theorem 24. ■

Claim 44. $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall x < n)(\forall x_2 < n)(x_1 = x_2 \vee \Theta(x_1) \neq \Theta(x_2))]\!] = \mathbf{1}$.

Proof of Claim 44. By Claims 42 and 43 and validity of predicate logic. ■

Claim 45. $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall x < n)(|\Theta(x)| < |n| - 1)]\!] = \mathbf{1}$.

Proof of Claim 45. Again, by the universality we just need to check the validity at each $\omega \in M_{n,\lfloor n/2\rfloor}$, which follows, as Θ outputs neigbors of vertices in ω, which are by the definition of $M_{n,\lfloor n/2\rfloor}$ always in the range $\{0, \dots, \lfloor n/2\rfloor - 1\} = \{t \in M_n \ ; \ |t| < |n| - 1\}$. ■

Claims 44 and 45 show the validity of the negations of all disjuncts of the formula $\varphi^0_{\textsc{WeakPigeon}}(\Theta, n)$. The formula $(n = 0)$ is also obviously assigned the Boolean value $\mathbf{0}$. This implies that

$$K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\exists x_1 < n)(\exists x_2 < n)\varphi_{\textsc{WeakPigeon}}(\Theta, n, x_1, x_2)]\!] = \mathbf{0},$$

which implies the theorem by the Boolean evaluation of universal quantifiers. □

6.3 Totality of RetractionWeakPigeon

We start by defining a formula $\varphi_{\textsc{RetractionWeakPigeon}}(X, Y, m, x)$, which is a formalization of the statement "the value x is a solution to the instance (X, Y, m) of the problem RetractionWeakPigeon."

Definition 46. Let $\varphi^0_{\textsc{RetractionWeakPigeon}}(X, Y, m, x)$ be the following L^2_n-formula:

$$(|X(x)| \geq |m| - 1) \vee (Y(X(x)) \neq x).$$

Let $\varphi_{\textsc{RetractionWeakPigeon}}(X, Y, m, x)$ be the following L^2_n-formula:

$$(m = 0) \vee \big(|x| < |m| \wedge \varphi^0_{\textsc{RetractionWeakPigeon}}(X, Y, m, x)\big).$$

Definition 47. Let $\Theta, \Xi \in G_{\mathrm{nb}}$, $\mu \in F_{\mathrm{nb}}$ and $\omega \in M_{n,\lfloor n/2\rfloor}$. We say that a tree $T \in \mathcal{T}_{\mathrm{nb}}$ *fails for* (Θ, Ξ, μ) *on* ω if $\mu(\omega) \neq 0$ and either $T(\omega) \geq |\mu(\omega)|$ or $\Xi_\omega(\Theta_\omega(T(\omega)) = T(\omega)$ and $|\Theta_\omega(T(\omega))| < |\mu(\omega)| - 1)$.

Lemma 48. For every $\Theta, \Xi \in G_{\mathrm{nb}}$ and $\mu \in F_{\mathrm{nb}}$ there is a tree $T \in \mathcal{T}_{\mathrm{nb}}$ such that

$$\mathrm{st}\left(\Pr_{\omega \in M_{n,\lfloor n/2\rfloor}}[T \text{ fails for } (\Theta, \Xi, \mu) \text{ on } \omega]\right) = 0.$$

Proof. We can assume that Θ is computed by $(\vartheta_0, \dots, \vartheta_{r-1})$ and Ξ by $(\xi_0, \dots, \xi_{r-1})$ such that $\mu(\omega) < r$ for all $\omega \in M_{n,\lfloor n/2\rfloor}$, and that the depth of all the trees computing ϑ_i, ξ_i and μ is bounded by $n^{1/t}$ for some $t \in \mathcal{M}_n \backslash \mathbb{N}$.

For any Θ and Ξ, this can be achieved by padding any pair of tuples that are computed by the constant function 0 in F_{nb}.

We will recursively construct trees $T_d \in \mathcal{T}_{\mathrm{nb}}$ indexed by elements of $\mathcal{M}_n$ and then show that $T = T_d$ for a sufficiently large value of $d \in \mathcal{M}_n$ satisfies the statement of the lemma. We will be constructing the tree by prolonging paths of previously constructed trees, where, by a path, we mean a complete path of nodes from the root all the way to some leaf. We say a sample $\omega \in M_{n,\lfloor n/2 \rfloor}$ is *compatible* with a path p in a tree if the computation of the tree on ω is exactly p.

The initial tree T_0. The initial tree T_0 is the tree computing μ, whose leaves we relabel as follows. For every path p in T_0 such that all $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p give $\mu(\omega) = 0$, we have a non-failure of T_0 and we can keep the original label. On the other hand, for each path p in T_0 such that all $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p give $\mu(\omega) > 0$, we pick a label in the following way. Let p be such a path and m be the value $\mu(\omega)$ obtained for any $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p. For each $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p, we have

$$\Pr_{i \in \{0,\ldots,m-1\}}[|\Theta_\omega(i)| < |m| - 1 \text{ and } \Xi_\omega(\Theta_\omega(i)) = i] \leq \frac{2^{|m|-2}}{2^{|m|-1}} = \frac{1}{2},$$

because the image of $\Xi_\omega \restriction \{s \in M_n \,;\, |s| < |m| - 1\}$ has at most $2^{|m|-2}$ elements. This implies, by an averaging argument, that there is $i_p \in \{0, \ldots, m-1\}$ such that

$$\Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[\Theta_\omega(i_p) < \lfloor m/2 \rfloor \text{ and } \Xi_\omega(\Theta_\omega(i_p)) = i_p \mid \omega \text{ compatible with } p] \leq \frac{1}{2}.$$

We relabel each leaf at such a path p to i_p. This concludes the construction of T_0 and assures that $\Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[T_0 \text{ fails}] \leq \frac{1}{2}$ and that the depth of T_0 is at most $n^{1/t}$.

The recursive step. Let us now assume that $d > 0$ and that T_{d-1} was already constructed with the property that $\Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[T_{d-1} \text{ fails}] \leq \left(\frac{1}{2}\right)^d$ and that the depth of T_{d-1} is at most $2dn^{1/t}$. Moreover, the tree T_{d-1} on each path p determines a value $m = \mu(\omega)$ and if this path fails, it also determines two $(d-1)$-element sets $D \subseteq \{s \in M_n \,;\, |s| < |m|\}$ and $R \subseteq \{s \in M_n \,;\, |s| < |m| - 1\}$ such that $\Theta_\omega \restriction D$ and $\Xi_\omega \restriction R$ form a pair of inverse functions on every ω compatible with p. We also assume that the leaf of every path p is labeled by some element of $\{s \in M_n \,;\, |s| < |m|\} \backslash D$, which we will call ℓ_p. We will prolong T_{d-1} along each path p in two stages and relabel the leaves to obtain the tree T_d.

First prolongation. Let p be a path in T_{d-1}. If the tree does not fail on any sample $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p, we keep p as it is. Otherwise, we replace the leaf by the tree computing ϑ_{ℓ_p} and change the label of each new leaf from ℓ to the tuple (ℓ_p, ℓ). We call all such new paths *active* and we call the resulting tree T'_{d-1}. By the assumption on the tree computing ξ_{ℓ_p}, we get that the depth of T'_{d-1} is at most $n^{1/t} + 2dn^{1/t}$.

Second prolongation. Let p be an active path in T'_{d-1} with its leaf labeled (ℓ_p, ℓ'_p) and let $m = \mu(\omega)$ on any ω compatible with p. If $|\ell'_p| \geq |m| - 1$, then, after relabeling the leaf of p to the value ℓ_p, the tree does not fail on any $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p. If $|\ell'_p| < |m| - 1$, we replace the leaf by the tree $\xi_{\ell'_p}$ and change the label of each new leaf from ℓ to the triple (ℓ_p, ℓ'_p, ℓ). We will denote the resulting tree T''_{d-1} and again call its newly prolonged paths *active*. By the assumptions on the tree computing $\xi_{\ell'_p}$, we have that the depth of T''_{d-1} is at most $2n^{1/t} + 2dn^{1/t} = 2(d+1)n^{1/t}$.

Relabeling. We will now relabel the leaves of active paths of T''_{d-1}. Let p be an active path of T''_{d-1} and let its leaf be labeled $(\ell_p, \ell'_p, \ell''_p)$. If ℓ''_p is distinct from ℓ_p, we can relabel the leaf to the value ℓ_p. This makes the tree never fail along this path. Otherwise, $\ell_p \notin D$ and the tree so far has established a bijection between $D \cup \{\ell_p\}$ and $R \cup \{\ell'_p\}$ computed by Θ_ω and Ξ_ω on any $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p. This implies that for each $\omega \in M_{n,\lfloor n/2 \rfloor}$ compatible with p, we have

$$\Pr_i[|\Theta_\omega(i)| < |m| - 1 \text{ and } \Xi_\omega(\Theta_\omega(i)) = i] \leq \frac{2^{|m|-2} - d}{2^{|m|-1} - d} \leq \frac{2^{|m|-2}}{2^{|m|-1}} = \frac{1}{2},$$

where i is sampled uniformly in $\{s \in M_n \ ; \ |s| < |m|\} \setminus (D \cup \{\ell_p\})$, which, by an averaging argument, implies the existence of i_p such that $\Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[|\Theta_\omega(i_p)| < |m| - 1 \text{ and } \Xi_\omega(\Theta_\omega(i_p)) = i_p \mid \omega \text{ compatible with } p] \leq \frac{1}{2}$. We then relabel the leaf of p to i_p. After relabeling all active paths in this way, we obtain the tree T_d.

Properties of T_d. The depth of T_d is the same as of T''_{d-1}, which is at most $2(d+1)n^{1/t}$. During the relabeling, we have also established that the failing paths determine a bijection with one additional element added to both R and D of the corresponding path in T_{d-1} and that the leaves are properly labeled. It remains to analyze the probability of T_d failing. If a path p in T_{d-1} did not fail on some $\omega \in M_{n,\lfloor n/2 \rfloor}$, any of its extensions in T_d also did not fail on ω. Moreover, if p did fail, we found a new label for each of its extensions in T_d, which fails with probability at most $1/2$. Therefore, $\Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[T_d \text{ fails}] \leq \frac{1}{2} \cdot \Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[T_{d-1} \text{ fails}] \leq \left(\frac{1}{2}\right)^d$.

Conclusion. Let $d = n^{1/t}$, then the depth of T_d is at most $4n^{1/(t/2)}$. Moreover, $T = T_d$ satisfies the statement of the lemma, because $\Pr_{\omega \in M_{n,\lfloor n/2 \rfloor}}[T_d \text{ fails}] \leq \left(\frac{1}{2}\right)^{n^{1/t}+1}$, the standard part of which is 0. □

Theorem 49. $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall X)(\forall Y)(\forall m)(\exists x < m)\varphi_{\textsc{RetractionWeakPigeon}}(X, Y, m)]\!] = \mathbf{1}$.

Proof. Let $\Theta, \Xi \in G_{\mathrm{nb}}$ and $\mu \in F_{\mathrm{nb}}$, we will find $\gamma \in F_{\mathrm{nb}}$ such that

$$K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![\varphi_{\textsc{RetractionWeakPigeon}}(\Theta, \Xi, \mu, \gamma)]\!] = \mathbf{1},$$

which is enough to prove the theorem.

Let $\gamma \in \mathcal{T}_{\mathrm{nb}}$ be computed by the tree whose existence follows from Lemma 48 for Θ, Ξ and μ. The formula $\varphi_{\textsc{RetractionWeakPigeon}}(\Theta, \Xi, \mu, \gamma)$ is interpreted for each $\omega \in M_{n,\lfloor n/2 \rfloor}$ as the statement claiming the tree computing γ does not fail for (Θ, Ξ, μ) on ω. Since it is a quantifier-free formula, it can be evaluated in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$ by calculating its probability over the sample space, which, by Lemma 48, is infinitesimally close to 1. Therefore, the formula $\varphi_{\textsc{RetractionWeakPigeon}}(\Theta, \Xi, \mu, \gamma)$ is valid in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$. □

6.4 Nonreducibility

In the next theorem, we will be working both with nonstandard binary strings and nonstandard numbers. We will define two pairs of functions that convert between these two types of objects. The first pair consists of the Wrap function $\mathtt{w}$, which converts binary strings to nonzero numbers by appending a leading 1 and interpreting the string as a binary expansion of the resulting number. The inverse to $\mathtt{w}$ is the Unwrap function, denoted by $\mathtt{u}$. These are then used to convert between the size parameter of the formula formalizing totality, either $\varphi_{\textsc{WeakPigeon}}$ or $\varphi_{\textsc{RetractionWeakPigeon}}$, and the size parameter of the associated $\mathsf{TFNP}^{\mathrm{dt}}$ problem.

The second pair then provides an explicit bijection between the set $\{s \ ; \ |s| < |m|\}$ and $\{0,1\}^{|\mathtt{u}(m)|}$, or alternatively between $\{s \ ; \ |s| < |\mathtt{w}(x)|\}$ and $\{0,1\}^{|x|}$, depending on whether we start with a binary string x or an element $m \in \mathcal{M}_n$. The first function, Pad, denoted by $\mathtt{p}$ simply takes an input number s and outputs its binary expansion with enough leading zeros to obtain a string of required size—this required size is a second

Operation	Symbol	Domain	Range	Input	Output
Wrap	$\mathtt{w}$	binary strings	$\mathcal{M}_n \setminus \{0\}$	x	$2^{\lvert x\rvert} + \sum_{i<\lvert x\rvert} x_i \cdot 2^i$
Unwrap	$\mathtt{u}$	$\mathcal{M}_n \setminus \{0\}$	binary strings	s	$\mathtt{w}^{-1}(s)$
Pad	$\mathtt{p}$	$\mathcal{M}_n$	binary strings	s, ℓ	The binary expansion of s padded with leading zeros to the length of ℓ.
Unpad	$\mathtt{up}$	binary strings	$\mathcal{M}_n$	x	$\sum_{i<\ell} x_i \cdot 2^i$

Figure 1: Overview of operations.

argument of the function. However, we will never write this second argument explicitly, and its value will always be clear from the context. The function UNPAD, denoted by $\mathtt{up}$, is the inverse to $\mathtt{p}$ and simply takes a binary string and interprets it as a binary expansion of a number ignoring leading zeros.

We will begin by proving the following lemma. Its proof is straightforward and it essentially shows that the formulas $\varphi_{\text{WEAKPIGEON}}$ and $\varphi_{\text{RETRACTIONWEAKPIGEON}}$ properly formalize the fact that a given element is a solution to the respective search problem. We only state one direction for each of the problems; this simplifies the proof and exactly captures the properties needed for the proof of our separation.

Lemma 50.

(1) Let $\Delta \in G_{\text{nb}}$, $\nu \in F_{\text{nb}}$ and $\omega \in M_{n,\lfloor n/2 \rfloor}$. If $\nu(\omega) > 0$, there is an instance $I = (B(\Delta_\omega), \mathtt{u}(\nu(\omega)))$ of WEAKPIGEON, such that for every solution (x_1, x_2) to I, we have that

$$\mathcal{M}_n \models \varphi_{\text{WEAKPIGEON}}(\Delta_\omega, \nu(\omega), \mathtt{up}(x_1), \mathtt{up}(x_2)).$$

(2) Let $\Theta, \Xi \in G_{\text{nb}}$, $\mu \in F_{\text{nb}}$ and $\omega \in M_{n,\lfloor n/2 \rfloor}$. If
   (a) $\mu(\omega) > 0$,
   (b) the range of $\Theta_\omega \restriction \{s \ ; \ |s| < |\mu(\omega)|\}$ is a subset of $\{s \ ; \ |s| < |\mu(\omega)| - 1\}$,
   (c) the range $\Xi_\omega \restriction \{s \ ; \ |s| < |\mu(\omega)| - 1\}$ is a subset of $\{s \ ; \ |s| < |\mu(\omega)|\}$,

   then the instance $J = (\mathtt{p}(\Theta_\omega(\mathtt{up}(-))), \mathtt{p}(\Xi_\omega(\mathtt{up}(-))), \mathtt{u}(\mu(\omega)))$ of RETRACTIONWEAKPIGEON satisfies that for every $s \in \mathcal{M}_n$ satisfying $\mathcal{M}_n \models \varphi_{\text{RETRACTIONWEAKPIGEON}}(\Theta_\omega, \Xi_\omega, \mu(\omega), s)$, we have that $\mathtt{p}(s)$ is a solution to J.

Proof. (1) The function oracle $B(\Delta_\omega)$ can be defined as follows:

$$B(\Delta_\omega)(x') = \begin{cases} \mathtt{p}(\Delta_\omega(\mathtt{up}(x'))) & \text{if } |\Delta_\omega(\mathtt{up}(x'))| < |\nu(\omega)| - 1, \\ \mathtt{p}(0) & \text{otherwise.} \end{cases}$$

Any solution (x_1, x_2) to I is either translated to a pair of distinct numbers mapped to the same element by Δ_ω or one of them is mapped outside $\{s \ ; \ |s| < |\nu(\omega)| - 1\}$, hence $\mathcal{M}_n \models \varphi_{\text{WEAKPIGEON}}(\Delta_\omega, \nu(\omega), \mathtt{up}(x_1), \mathtt{up}(x_2))$.

(2) This follows straight from the assumptions and the definition of the formula $\varphi_{\text{RETRACTIONWEAKPIGEON}}$. □

Theorem 51. There is no many-one reduction from the problem WEAKPIGEON to the problem RETRACTIONWEAKPIGEON.

Proof. Assume that WEAKPIGEON is many-one reducible to RETRACTIONWEAKPIGEON. We will show that this implies $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\text{nb}}, G_{\text{nb}}\big)[\![(\forall X)(\forall m)(\exists x_1)(\exists x_2)\varphi_{\text{WEAKPIGEON}}(X, m, x_1, x_2)]\!] = \mathbf{1}$, which is in contradiction with Theorem 41.

The existence of the reduction implies that there is a function r from binary strings to binary strings computed by query trees $(T^r_x)_x$. We also assume the outputs of the trees have already been decoded to allow outputs of varying length.

Let $\Delta \in G_{\text{nb}}$, $\nu \in F_{\text{nb}}$. We will use the reduction to construct some $\Theta, \Xi, \Lambda^1, \Lambda^2 \in G_{\text{nb}}$ and $\mu \in F_{\text{nb}}$ such that the formula

$$\varphi_{\text{RETRACTIONWEAKPIGEON}}(\Theta, \Xi, \mu, \gamma) \to \varphi_{\text{WEAKPIGEON}}\big(\Delta, \nu, \Lambda^1(\gamma), \Lambda^2(\gamma)\big) \tag{3}$$

is valid in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\text{nb}}, G_{\text{nb}}\big)$ for any $\gamma \in F_{\text{nb}}$.

Let us assume Δ is computed by $(\delta_0, \ldots, \delta_{d-1})$, $d \in \mathcal{M}_n$, where each element is from F_{nb} and δ_i is computed by $T^\delta_i \in \mathcal{T}_{\text{nb}}$ for each $i < d$. First, we define μ as the function computed by the tree of ν, where we replace the each leaf with label ℓ by the tree $T^r_{\mathtt{u}(\ell)}$ and relabel the new leaves from $r(\mathtt{u}(\ell))$ to $\mathtt{w}(r(\mathtt{u}(\ell)))$. By Lemma 50, after the computation of the tree computing μ is determined, the tree computing ν is determined, thus a set of WEAKPIGEON instances of the form $(B(\Delta_\omega), \mathtt{u}(\nu(\omega)))$ is determined, where $\omega \in M_{n,\lfloor n/2 \rfloor}$ is compatible with

the computation of μ. By the existence of the reduction, after ν and μ are determined, we also obtain the sequences of query trees $(T^{\mathrm{I}}_{x'})_{x'\in\{0,1\}^{|r(\mathtt{u}(m))|}}$ and $(T^{\mathrm{s}}_{y'})_{y'\in\{0,1\}^{|r(\mathtt{u}(m))|}}$, where m is the value of ν.

Our goal now is to define sequences of $\mathcal{T}_{\mathrm{nb}}$ trees that will define Θ, Ξ, Λ_1 and Λ_2. To do this, we first have to convert the query trees to $\mathcal{T}_{\mathrm{nb}}$ trees. Assume that μ is determined, then ν is determined and we obtain the sequences $(T^{\mathrm{I}}_{x'})_{x'}$ and $(T^{\mathrm{s}}_{y'})_{y'}$. For each tree $T^{\mathrm{I}}_{x'}$ we define a $\mathcal{T}_{\mathrm{nb}}$ tree $S^{\mathrm{I}}_{x'}$ by simulating each query q with $T^{\delta}_{\mathtt{up}(q)}$. That is, layer by layer, starting at the parents of leaves, replace each non-leaf node labeled q with the tree computing $T^{\delta}_{\mathtt{up}(q)}$ and append the subtrees from the previous layer onto the leaves of the new copy of $T^{\delta}_{\mathtt{up}(q)}$ with the same label as the edge they were originally incident to. The depth of the resulting tree $S^{\mathrm{I}}_{x'}$ is bounded by the maximal depth of any $T^{\delta}_{\mathtt{up}(q)}$ times the depth of $T^{\mathrm{I}}_{x'}$. The depth of each $T^{\delta}_{\mathtt{up}(q)}$ is at most polynomial in $|m|$; from $m \in \mathcal{M}_n$ we get that $|m|$ is bounded by $n^{1/t}$ for some $t \in \mathcal{M}\backslash\mathbb{N}$, and thus $T^{\mathrm{I}}_{x'} \in \mathcal{T}_{\mathrm{nb}}$ implies that the depth of $S^{\mathrm{I}}_{x'}$ is also bounded by $n^{1/t}$ for some $t \in \mathcal{M}\backslash\mathbb{N}$. Therefore, the tree $S^{\mathrm{I}}_{x'}$ is in $\mathcal{T}_{\mathrm{nb}}$. An analogous observation about depth also holds for the remaining trees we will construct.

The previous construction can also be done with $T^{\mathrm{s}}_{y'}$ to obtain the $\mathcal{T}_{\mathrm{nb}}$ tree $S^{\mathrm{s}}_{y'}$. Let $d' \in \mathcal{M}_n$ be a power of 2 larger than any possible value of μ. For each $i < d'$ we will construct the tree U^{I}_i as follows: start with the tree computing μ and to each path p with a leaf label m' satisfying $|i| < |m'|$ append the tree $S^{\mathrm{I}}_{\mathtt{p}(i)}$; the remaining paths are kept unchanged as they are not important. Analogously, we also define U^{s}_i to be the tree constructed from the tree computing μ by appending $S^{\mathrm{s}}_{\mathtt{p}(i)}$ onto each path labeled m' satisfying $|i| < |m'|$. Let us also assume that the labels of $U^{\mathrm{I}}_{x'}$ and $U^{\mathrm{s}}_{y'}$ are numbers obtained from the original labels by the function $\mathtt{up}$.

Since RetractionWeakPigeon is a problem presented by two function oracles, there is a way to compute either of the functions $\mathcal{O}^f$ and $\mathcal{O}^g$ using a single input oracle, which in our case is itself computed by the sequence $\big(T^{\mathrm{I}}_{x'}\big)_{x'}$. We will define Θ, Ξ, Λ_1 and Λ_2 by specifying the elements of the tuples they are computed by, which are denoted by the respective lowercase letters. We define ϑ_i to be computed by the tree U^{I}_i, except we relabel its leaves to only keep the value of $\mathcal{O}^f$ and similarly let ξ_i be also computed by U^{I}_i except we relabel it to only keep the value of $\mathcal{O}^g$. We also define λ^1_i to be computed by U^{s}_i, relabeled to keep only the value for the first part of the solution, and λ^2_i to be computed by U^{s}_i, relabeled to keep only the second part of the solution.

To prove the validity of (3), which is a quantifier-free sentence, we can simply check the validity over every $\omega \in M_{n,\lfloor n/2\rfloor}$. After ω is fixed, the tuples $J = (\mathtt{p}(\Theta_\omega(\mathtt{up}(-))), \mathtt{p}(\Xi_\omega(\mathtt{up}(-))), \mathtt{u}(\mu(\omega)))$ and $I = (B(\Delta_\omega), \mathtt{u}(\nu(\omega)))$ are instances of the problems RetractionWeakPigeon and WeakPigeon respectively. Assume that the antecedent of (3) is satisfied, therefore $\mathcal{M}_n \models \varphi_{\textsc{RetractionWeakPigeon}}(\Theta_\omega, \Xi_\omega, \mu(\omega), \gamma(\omega))$.

Notice that the instance J satisfies the assumptions (a), (b) and (c) in Lemma 50 (2) by construction. This shows that $\mathtt{p}(\gamma(\omega))$ is a solution to J. By the existence of the reduction, we get that $\mathtt{p}\big(\lambda^1_{\gamma(\omega)}(\omega)\big)$ and $\mathtt{p}\big(\lambda^2_{\gamma(\omega)}(\omega)\big)$ form a solution to I. Finally, Lemma 50 (1) implies $\mathcal{M}_n \models \varphi_{\textsc{WeakPigeon}}\big(\Delta(\omega), \nu(\omega), \Lambda^1_\omega(\gamma(\omega)), \Lambda^2_\omega(\gamma(\omega))\big)$, which concludes the proof of (3).

Using the fact that for any $\Delta \in G_{\mathrm{nb}}$, $\nu \in F_{\mathrm{nb}}$ we can find $\Theta, \Xi, \Lambda^1, \Lambda^2 \in G_{\mathrm{nb}}$ and $\mu \in F_{\mathrm{nb}}$ that make (3) valid, Theorem 49, and the validity of predicate logic, we obtain

$$K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)[\![(\forall X)(\forall m)(\exists x_1)(\exists x_2)\varphi_{\textsc{WeakPigeon}}(X, m, x_1, x_2)]\!] = \mathbf{1}. \qquad \square$$

6.5 Further properties of the constructed model

In the rest of this section we state which additional principles hold in the model.

Theorem 52. Let $\varphi(x)$ be an open L^2_n-formula with parameters from F_{nb} and G_{nb}. Then for every $m \in \mathcal{M}_n$ the open comprehension principle $(\exists X)(\forall y < m)(X(y) \equiv \varphi(y))$ and the open induction principle $\neg\varphi(0) \vee \varphi(m) \vee (\exists x < m)(\varphi(x) \wedge \neg\varphi(x+1))$ are both valid in $K\big(M_{n,\lfloor n/2\rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$, where we interpret m as the constant function always outputting m in F_{nb}.

Proof. This can be proven completely analogously to [11, Lemma 20.2.5]. We will therefore only describe the basic idea. Let $m \in \mathcal{M}_n$ and let $\varphi(x)$ be an open L^2_n formula with a single free variable with parameters

from F_{nb} and G_{nb}. For each number $i \in \mathcal{M}_n, i < m$, the validity of $\varphi(i)$ for a given ω can be decided by a $\mathcal{T}_{\mathrm{nb}}$ tree T_i, and the depth of all such trees is bounded above by $n^{1/t}, t \in \mathcal{M}_n \backslash \mathbb{N}$. A tuple of elements of F_{nb} each computed by the corresponding tree then determines an element of G_{nb} that satisfies the comprehension principle for $\varphi(x)$ and m. The induction principle follows directly from the comprehension principle by an application of a binary search procedure. □

Let us give a remark about the logical strength of the separation we gave. The language L_n contains names for all the functions on $\mathbb{N}$ that are polynomially bounded (the length of the output is bounded by a polynomial in the length of the input). This allows us to interpret the language $L_{\mathrm{PV}(f)}$, containing the names for every polynomial time algorithm with oracle access to some polynomially bounded function f, in the structure $\mathcal{M}_n$ and therefore in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$. We define $\forall\mathrm{PV}_1(f)$ to be the theory consisting of all universal $L_{\mathrm{PV}(f)}$-formulas true for any interpretation of f. The function f may represent a function oracle, hence symbols from $L_{\mathrm{PV}(f)}$ represent (the functions computed by) the polynomial time oracle machines with access to f, and $\forall\mathrm{PV}_1(f)$ is simply the set of all universal sentences true for such functions for any oracle.

Theorem 53. Let the function symbol f be interpreted in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$ by any $\Theta \in G_{\mathrm{nb}}$. Then, all the axioms of $\forall\mathrm{PV}_1(f)$ are valid in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$.

Proof. The theory $\forall\mathrm{PV}_1(f)$ consists of all true universal statements for any interpretation of f. Therefore, for any $\omega \in M_{n,\lfloor n/2 \rfloor}$, any axiom of $\forall\mathrm{PV}_1(f)$ is true in $\mathcal{M}_n$, after f is interpreted as Θ_ω, under any substitution of the universal quantifiers. Thus, any axiom of $\forall\mathrm{PV}_1(f)$ is also valid in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$. □

This theorem is relevant for interpretation of our construction from the point of view of bounded arithmetic. The theory $\forall\mathrm{PV}_1(f)$ is the true universal extension of the relativized version of the theory PV_1. The theory PV_1 was first defined in [12] as a first-order extension of Cook's equational theory PV [2].

7 Concluding remarks

In this paper, we have developed some basic theory around the new concept of a limit object, the *wide limit*. In the author's view, it provides a more elementary counterpart to Krajíček's structures $K(F, G)$, into which they can be expanded while preserving the values of sentences restricted to their universe (Theorem 24). In particular, even though the ideas that are needed to prove our many-one separation (Theorem 51) could be collected in a direct proof of this separation, the wide limit provides a concrete semantic interpretation of this separation: an observer who queries the images of a random map $\omega \in M_{n,\lfloor n/2 \rfloor}$, only polynomially many times will observe with high probability a map without a collision. This map is embodied by our limit object $\lim_{F_{\mathrm{nb}}} M_{n,\lfloor n/2 \rfloor}$.

The problem RetractionWeakPigeon and its variations have been investigated in the context of TFNP (rPHP^{2n}_{n}, rWPHP in [8; 14], "lossy code" in [10]) and its relation to WeakPigeon follows from well-known results. A nonreducibility of WeakPigeon to a problem stronger than RetractionWeakPigeon, namely RetractionPigeon, follows from a lower bound on the Sherali–Adams proof system in [3] and established connections between propositional proofs and total NP search problems (see [6]).

The main reason why the problem RetractionWeakPigeon was total in the constructed model, was that it had many solutions that could be checked in a constant number of queries. It would be interesting to see how this can be generalized for other total search problems.

Problem 54. Characterize which TFNP problems are total in $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$.

In [14], the totality of certain total search problems is relatively unprovable even while preserving quite a lot of induction. In our model $K\big(M_{n,\lfloor n/2 \rfloor}, F_{\mathrm{nb}}, G_{\mathrm{nb}}\big)$, only $\forall\mathrm{PV}_1(f)$ was verified. It would be interesting to investigate further how much induction can be verified in the models emerging as expansions of wide limits.

In [11], several models of the theory V_1^0 are constructed. However, these are obtained by expanding the family F and using a suitable switching lemma to obtain a form of quantifier elimination, which in our case makes the wide limit interpretation unnatural, because for this construction a more complex sample space is used.

A general direction of further research that would be particularly interesting, would be to characterize valid sentences of some concrete wide limit without the direct construction, and thus to prove upper and lower bounds for the family F of functions being considered.

Acknowledgements. This work is based on the author's Master's thesis [9] which was completed under the supervision of Jan Krajíček. The author thanks Eitetsu Ken for comments on a draft of this paper.

This work has been supported by Charles University Research Center program No. UNCE/24/SCI/022, the project SVV-2025-260837 and by the GA UK project No. 246223.

References

[1] P. Beame, S. Cook, J. Edmonds, R. Impagliazzo & T. Pitassi. The relative complexity of NP search problems. In: *STOC '95: Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, 1995, pp. 303–314. DOI: 10.1145/225058.225147.

[2] S. A. Cook. Feasibly constructive proofs and the propositional calculus (preliminary version). In: *STOC '75: Proceedings of the Seventh Annual ACM Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, 1975, pp. 83–97. DOI: 10.1145/800116.803756.

[3] S. Dantchev, B. Martin & M. Rhodes. Tight rank lower bounds for the Sherali–Adams proof system. *Theor. Comput. Sci.* 410:21–23 (2009), pp. 2054–2063. DOI: 10.1016/j.tcs.2009.01.002.

[4] R. Fagin. Probabilities on finite models. *J. Symb. Log.* 41:1 (1976), pp. 50–58. DOI: 10.2307/2272945.

[5] I. Goldbring. *Lecture notes on nonstandard analysis*. 2014.

[6] M. Göös, A. Hollender, S. Jain, G. Maystre, W. Pires, R. Robere & R. Tao. Separations in proof complexity and TFNP. *J. ACM* 71:4 (2024), pp. 1–45. DOI: 10.1145/3663758.

[7] J. Hanika. *Search Problems and Bounded Arithmetic*. PhD thesis. Charles University, Prague, 2004.

[8] E. Jeřábek. Approximate counting by hashing in bounded arithmetic. *J. Symb. Log.* 74:3 (2009), pp. 829–860. DOI: 10.2178/jsl/1245158087.

[9] O. Ježil. *Pseudofinite Structures and Limits*. Master's thesis. Charles University, Prague, 2022.

[10] O. Korten. *Efficient low-space simulations from the failure of the weak pigeonhole principle*. Technical report. 2022. Electronic Colloquium on Computational Complexity: TR22-025.

[11] J. Krajíček. *Forcing With Random Variables and Proof Complexity*. London Mathematical Society Lecture Note Series, Vol. 382. Cambridge University Press, 2011. DOI: 10.1017/CBO9781139107211.

[12] J. Krajíček, P. Pudlák & G. Takeuti. Bounded arithmetic and the polynomial hierarchy. *Ann. Pure Appl. Log.* 52:1–2 (1991), pp. 143–153. DOI: 10.1016/0168-0072(91)90043-l.

[13] L. Lovász & B. Szegedy. Limits of dense graph sequences. *J. Comb. Theory*. Series B 96:6 (2006), pp. 933–957. DOI: 10.1016/j.jctb.2006.05.002.

[14] M. Müller. Typical forcings, NP search problems and an extension of a theorem of Riis. *Ann. Pure Appl. Log.* 172:4, 102930 (2021). DOI: https://doi.org/10.1016/j.apal.2020.102930.

[15] J. Nešetřil & P. Ossona de Mendez. A model theory approach to structural limits. *Commentationes Math. Univ. Carolinae* 53:4 (2012), pp. 581–603.

[16] C. H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *J. Comput. Syst. Sci.* 48:3 (1994), pp. 498–532. DOI: https://doi.org/10.1016/S0022-0000(05)80063-7.

[17] A. A. Razborov. Flag algebras. *J. Symb. Log.* 72:4 (2007), pp. 1239–1282. DOI: 10.2178/jsl/1203350785.

[18] N. Thapen. Structures interpretable in models of bounded arithmetic. *Ann. Pure Appl. Log.* 136:3 (2005), pp. 247–266. DOI: https://doi.org/10.1016/j.apal.2005.04.005.