Approximate counting by hashing in bounded arithmetic

Emil Jeřábek

Institute of Mathematics of the Academy of Sciences
Žitná 25, 115 67 Praha 1, Czech Republic, email: jerabek@math.cas.cz

November 10, 2008

Abstract

We show how to formalize approximate counting via hash functions in subsystems of bounded arithmetic, using variants of the weak pigeonhole principle. We discuss several applications, including a proof of the tournament principle, and an improvement on the known relationship of the collapse of the bounded arithmetic hierarchy to the collapse of the polynomial-time hierarchy.

1 Introduction

Counting the number of elements of a finite set is one of the most fundamental operations in discrete mathematics. However, exact counting is not available in weak systems of first-order arithmetic where exponentiation is not a total operation unless the polynomial hierarchy collapses, because of Toda's theorem [36]. This does not exclude the possibility of *approximate counting*, which is sufficient in many counting applications: we estimate the size of the set up to a negligible error (where the meaning of "negligible" depends on the context).

A popular way of simulating approximate counting arguments in bounded arithmetic is to apply variants of the weak pigeonhole principle, see e.g. [29, 30, 28]. A systematic approach was taken in [21]: we have proved in the theory $PV_1 + sWPHP(PV)$ (defined below in Section 2) that for any bounded set defined by a Boolean circuit, there exist suitable kind of surjective "counting functions" (also definable by circuits) which allow us to coherently define the approximate size of the set up to a polynomially small error. This framework admits smooth development of basic counting and probability arguments (including, e.g., the inclusion-exclusion principle, and the Chernoff bounds), and provides a suitable means to define and discuss randomized algorithms in bounded arithmetic. However, it also suffers from a significant defect: if X is a subset of $[0, a]$, we can only estimate the size of X up to an error polynomially small relative to a—the size of the ambient interval—rather than relative to the size of X itself. (One of the reasons being that the size of X is estimated by sampling it with a pseudorandom number generator.) Sufficiently "sparse" sets are thus indistinguishable from the empty set. This precludes more sophisticated combinatorial counting arguments (in particular, inductive arguments such as in the proof of the Ramsey theorem), and it is at odds with what usually goes by the name “approximate counting” in theoretical computer science.

Sipser’s Coding Lemma [34], which is an application of Carter–Wegman 2-universal families of hash functions [7], shows that the polynomial-time hierarchy is closed under a stronger form of approximate counting: if X is the finite set we want to count, and n is a parameter given in unary, we can find s such that $s \le |X| \le s(1+\varepsilon)$ for any $\varepsilon \le n^{-O(1)}$. Our aim is to show that Sipser’s definition makes a well-behaved concept of approximate counting in bounded arithmetic. We work in the theory $T^1_2 + sWPHP(PV_2)$ (i.e., the one as before, but relativized with an NP-oracle; it is a subtheory of T^3_2), or in the slightly weaker theory $T^1_2 + rWPHP(PV_2)$ (see Section 2). The key technical result (which can be thought of as formalization of the Coding Lemma in bounded arithmetic) states that Sipser-style approximate counting in terms of hash functions is (more or less) equivalent to the existence of certain surjective functions (Corollary 3.5 and Theorem 3.8). Armed with this “implementation-independent” view of hashing, we are able to prove basic properties of counting (Section 3), such as the size of a disjoint union is (approximately) the sum of sizes of the summands.

In Section 4 we mention some applications, intended as examples demonstrating how the methods developed in Section 3 may be used to formalize counting arguments in bounded arithmetic. We solve an open problem of Krajíček, Pudlák, and Takeuti [9, 23] by showing that $T^1_2 + rWPHP(PV_2)$ (hence also T^3_2) proves Erdős’s [13] tournament principle (Theorem 4.2). We also prove a generalization of the tournament principle (Theorem 4.3), which allows us to strengthen the results of [26, 3] showing that the collapse of the bounded arithmetic hierarchy implies collapse of the polynomial-time hierarchy (Theorem 4.6 and Corollary 4.7). We observe that approximate counting provides an approximate Euler characteristic (in the sense of Krajíček [25]) for models of $S_2(\alpha)$ (Theorem 4.10). We also include two applications from computational complexity: we formalize in bounded arithmetic Cai’s [5] result $S^P_2 \subseteq ZPP^{NP}$ (Theorem 4.11), and the existence of an AM-algorithm for graph nonisomorphism by Goldwasser and Sipser [15] (Theorem 4.12).

We remark that the “new” approximate counting method does not make the “old” counting of [21] superfluous: while the method in the present paper allows for better approximation (we can estimate the size of a set X up to an error which is a polynomially small fraction of $|X|$, rather than of the size of the ambient universe) which also agrees with the established usage of the term “approximate counting” in computer science, the price we pay is an increase in the complexity of the counting functions, which requires an increase of the strength of the base theory by one level of the bounded arithmetic hierarchy. To put it differently, $T^1_2 + sWPHP(PV_2)$ can count $P^{NP}/poly$-sets using the old method, but only $NP/poly$-sets using the new method. Moreover, some results from [21] are used in an essential way in Section 3.

Acknowledgement

I am indebted to Neil Thapen and Jan Krajíček for insightful discussions, and many useful suggestions, and to Pavel Pudlák for clarification of [26]. The research was supported by grant IAA1019401 of GA AV ČR, and grant 1M0545 of MŠMT ČR.

2 Preliminaries

We assume some degree of familiarity with first-order bounded arithmetic, however the basic definitions are summarized below. More background can be found in [23, 4, 17].

Buss' theories are formulated in the language $L = \langle 0, S, +, \times, \leq, \#, |x|, \lfloor \frac{x}{2} \rfloor \rangle$. The intended meaning of the symbols are the usual arithmetical operations on non-negative integers, and $|x| = \lceil \log_2(x+1) \rceil$, $x \# y = 2^{|x|\cdot|y|}$. *Bounded quantifiers* are introduced by

$$\begin{aligned} \exists x \leq t\, \varphi &\Leftrightarrow \exists x\, (x \leq t \wedge \varphi), \\ \forall x \leq t\, \varphi &\Leftrightarrow \forall x\, (x \leq t \to \varphi), \end{aligned}$$

where t is a term without an occurrence of the variable x. Such a quantifier is *sharply bounded*, if t has the form $|s|$ for some term s. A formula φ is bounded (sharply bounded) if all quantifiers in φ are bounded (sharply bounded). A formula is Σ^b_1 if it is constructed from sharply bounded formulas by means of $\wedge$, $\vee$, sharply bounded, and existential bounded quantifiers. In general, Σ^b_i-formulas consist of i alternating blocks of bounded quantifiers followed by a sharply bounded formula, where the first block is existential, and we ignore sharply bounded quantifiers which are allowed to appear anywhere in the quantifier prefix. Π^b_i-formulas are defined similarly, but the first block is universal; in other words, Π^b_i-formulas are negations of Σ^b_i-formulas. The class of Boolean combinations of Σ^b_i-formulas is denoted by $\mathcal{B}(\Sigma^b_i)$. Bounded formulas capture the polynomial-time hierarchy (PH). More precisely, for any $i \geq 1$ the class Σ^P_i coincides with sets of natural numbers definable by Σ^b_i-formulas in $\mathbb{N}$ (the standard model of arithmetic), and dually $\Pi^P_i = \Pi^b_i(\mathbb{N})$, in particular $NP = \Sigma^b_1(\mathbb{N})$.

The theory T^i_2 is axiomatized by a finite set of open axioms denoted by *BASIC*, which state elementary properties of the symbols of L, and the schema of *induction*

$$(IND) \qquad \varphi(0) \wedge \forall x < a\, (\varphi(x) \to \varphi(x+1)) \to \varphi(a)$$

for Σ^b_i-formulas φ. The theory S^i_2 is axiomatized over *BASIC* by the *polynomial induction* schema

$$(PIND) \qquad \varphi(0) \wedge \forall x \leq a\, (\varphi(\lfloor \tfrac{x}{2} \rfloor) \to \varphi(x)) \to \varphi(a)$$

for Σ^b_i-formulas φ. Alternatively, S^i_2 can be axiomatized over *BASIC* by the *length induction* schema

$$(LIND) \qquad \varphi(0) \wedge \forall x < |a|\, (\varphi(x) \to \varphi(x+1)) \to \varphi(|a|),$$

or by the *length maximization* schema

$$(LMAX) \qquad \varphi(|a|) \to \exists b \leq |a|\, (\varphi(b) \wedge \forall c < b\, \neg\varphi(c))$$

for Σ^b_i-formulas φ. We have $S^i_2 \subseteq T^i_2 \subseteq S^{i+1}_2$, the full bounded arithmetic is thus $S_2 = \bigcup_i S^i_2 = \bigcup_i T^i_2$. The theory S^{i+1}_2 is $\forall\Sigma^b_{i+1}$-conservative over T^i_2 by Buss' witnessing theorem [2] (in the case of $i = 0$ we need a minor adjustment of the language of T^0_2, see [19]).

PV is an equational theory introduced by Cook [11]. Its language contains function symbols for all polynomial-time algorithms, introduced inductively using limited recursion on notation (cf. Cobham [10]). It is axiomatized by defining equations of its function symbols, and a derivation rule similar to *PIND*. PV_1, also known as QPV, $T^0_2(\Box^p_1)$, or $\forall\Sigma^b_1(S^1_2)$, is a first-order variant of PV. It can be axiomatized by equations provable in PV together with the axioms $0 \neq 1$ and $\lfloor \frac{x}{2} \rfloor = 0 \to x = 0 \lor x = 1$, and it proves the *PIND* and *IND* schemata for sharply bounded formulas. We will also use the symbol PV to denote the set of function symbols of PV.

The theories PV_{i+1} for $i > 0$, introduced in [26], are defined similarly to PV_1, except that the basic functions of their language include the characteristic functions of all Σ^b_i-predicates, thus PV_{i+1}-functions correspond to $FP^{\Sigma^P_i}$ in the standard model. Again, we will also use PV_{i+1} to denote the class of PV_{i+1}-functions. The class of PV_{i+1}-predicates (which corresponds to Δ^P_{i+1} in the standard model) is denoted by Δ^b_{i+1}; it coincides with predicates provably $\Sigma^b_{i+1} \cap \Pi^b_{i+1}$ in either T^i_2 or S^{i+1}_2.

As PV_{i+1} is a conservative extension of T^i_2 by definitions, we will simply identify PV_{i+1} with T^i_2, and work freely with PV_{i+1}-functions in T^i_2. The theory $S^1_2(PV_{i+1})$, which is axiomatized by PV_{i+1} and $\Sigma^b_1(PV_{i+1})$-*PIND*, is a conservative extension of S^{i+1}_2 for the same reason, hence we will also identify $S^{i+1}_2 = S^1_2(PV_{i+1})$.

All these theories can be *relativized* in a straightforward way. We include a new predicate α in the language, and define $\Sigma^b_i(\alpha)$ as before, but allowing α to be used in atomic formulas. The theory $T^i_2(\alpha)$ consists of *BASIC* and $\Sigma^b_i(\alpha)$-*IND* (i.e., there are no axioms involving α apart from the induction axioms), and similarly $S^i_2(\alpha) = \mathit{BASIC} + \Sigma^b_i(\alpha)$-*PIND*. In the case of $PV(\alpha)$ and $PV_i(\alpha)$, we allow the characteristic function of α to appear in functions constructed by limited recursion on notation, so that function symbols of $PV(\alpha)$ correspond to polynomial-time oracle algorithms. More generally, if Γ is a set of formulas, we define $\Sigma^b_i(\Gamma)$, $T^i_2(\Gamma)$, $S^i_2(\Gamma)$, and $PV_i(\Gamma)$ by substituting Γ-formulas for α in $\Sigma^b_i(\alpha)$, $T^i_2(\alpha)$, $S^i_2(\alpha)$, and $PVi(\alpha)$. (Notice that $T^1_2(\Sigma^b_i) = T^{i+1}_2$, and so on.) The main point of relativization is that this kind of substitution preserves derivability. Hence, e.g., if we prove a statement about counting of $\Sigma^b_1(\alpha)$-sets in $T^1_2(\alpha) + \mathit{sWPHP}(PV_2(\alpha))$, it also applies to counting of Σ^b_i-sets in $T^i_2 + \mathit{sWPHP}(PV_{i+1})$ for every $i > 0$.

The *choice* schema (aka bounded collection, or replacement) $BB\Gamma$ for a set of formulas Γ is defined by

$$\forall x < |a|\, \exists y \le b\, \varphi(x, y) \to \exists w\, \forall x < |a|\, \varphi(x, (w)_x),$$

where $\varphi \in \Gamma$, and $(w)_x$ denotes the xth member of the sequence encoded by w. $BB\Sigma^b_i(\alpha)$ is provable in $S^i_2(\alpha)$.

For any functions f, g, the surjective (also called dual), injective, and retraction pigeonhole principles are defined by

$$\begin{aligned} \mathit{sPHP}^a_b(f) &\Leftrightarrow \exists v < b\, \forall u < a\, f(u) \neq v, \\ \mathit{iPHP}^a_b(g) &\Leftrightarrow \exists v < b\, g(v) \ge a \lor \exists v < b\, \exists v' < v\, g(v) = g(v'), \\ \mathit{rPHP}^a_b(f, g) &\Leftrightarrow \exists v < b\, (g(v) \ge a \lor f(g(v)) \neq v). \end{aligned}$$

(Recall that a *retraction pair* is a pair of functions f, g such that $f \circ g = \mathrm{id}$; the function f is called a retraction, and g is its coretraction.) Note that the functions f, g may involve other parameters not explicitly shown. The *weak pigeonhole principles* are defined by

$$?\,WPHP(f) = \forall\bigl(x > 0 \to ?PHP^{x|y|}_{x(|y|+1)}(f)\bigr),$$
$$?\,WPHP(\Gamma) = \{?\,WPHP(f);\ f \in \Gamma\},$$

where $\forall$ denotes universal closure, Γ is a set of functions, and $? \in \{s, i, r\}$. In the case of $r\,WPHP(PV(\Gamma))$ and $i\,WPHP(PV(\Gamma))$, the principles thus introduced are equivalent to the more usual variants with bounds $?PHP^{x}_{2x}$ or $?PHP^{x}_{x^2}$ over $PV_1(\Gamma)$. This however does not hold for $s\,WPHP$ (we need $S^1_2(\Gamma)$ to prove the equivalence, see [20] for details), we thus need to state the principle in the strong form as above.

As $T^i_2(\alpha)$ proves that every $PV_{i+1}(\alpha)$-function is on a bounded domain computable by a polynomial-size circuit with a $\Sigma^b_i(\alpha)$-oracle, the schema $?\,WPHP(PV_{i+1}(\alpha))$ is over $T^i_2(\alpha)$ equivalent to its single instance where f is the evaluation function for $\Sigma^b_i(\alpha)$-oracle circuits, for any ?.

Clearly $r\,WPHP(f, g)$ follows from either $s\,WPHP(f)$ or $i\,WPHP(g)$. The weak pigeonhole principles $s\,WPHP(f)$ and $i\,WPHP(f)$ are provable in $T^2_2(f)$ [29, 23, 27] (in particular, $s\,WPHP(PV_2(\alpha))$ is contained in $T^3_2(\alpha)$), but no variant of $WPHP$ is provable in $S^2_2(f)$ [22, 32]. We will often use (for $i = 1$) the following connection between $r\,WPHP$ and $s\,WPHP$, which follows by relativization of [18, Cor. 1.15, 4.12].

Theorem 2.1 *For any $i \geq 0$, the theory $S^{i+1}_2(\alpha) + BB\Sigma^b_{i+2}(\alpha) + s\,WPHP(PV_{i+1}(\alpha))$ is $\forall\Sigma^b_{i+1}(\alpha)$-conservative over $T^i_2(\alpha) + r\,WPHP(PV_{i+1}(\alpha))$.*

We will often work with *bounded definable sets*, which are collections of numbers of the form

$$X = \{x < a;\ \varphi(x)\},$$

where φ is a formula. Bounded sets are *not* genuine objects in our arithmetical theories, but a figure of speech: $x \in X$ is an abbreviation for $x < a \wedge \varphi(x)$. We will write $X \in \Sigma^b_1(\alpha)$ if X is a bounded set defined by a $\Sigma^b_1(\alpha)$-formula. When used in a context which asks for a set, a number a is assumed to represent the integer interval $[0, a)$; thus, for example, $X \subseteq a$ means that all elements of X are less than a. We will use simple set-theoretic operations, whose meaning should be generally clear from the context; for example, if $X \subseteq a$ and $Y \subseteq b$, we may define

$$X \times Y = \{bx + y;\ x \in X, y \in Y\} \subseteq ab,$$
$$X \mathbin{\dot\cup} Y = X \cup \{y + a;\ y \in Y\} \subseteq a + b.$$

On the other hand, we will occasionally (especially in the applications) need to refer to "small" sets directly encoded by a number. They should be distinguishable from definable sets by the context; in particular, by the absence of a complexity measure (as in "a $\Sigma^b_1(\alpha)$-set"). If X is such a small set, we denote by $|X|$ its cardinality, defined in a natural way (e.g., as in Corollary 3.10). This notation should not be confused with the length (or logarithm) function $|a|$ from the basic language of bounded arithmetic.

Due to general absence of BB in our base theory, we will often need to work with a strengthened notion of surjectivity. We will call a function $f\colon X \to Y$ a *smooth surjection*, written as

$$f\colon X \twoheadrightarrow Y,$$

if for every sequence w of elements of Y, there exists a sequence v of elements of X, such that $\mathrm{lh}(v) = \mathrm{lh}(w)$, and $f(v_i) = w_i$ for every $i < \mathrm{lh}(v)$, where $\mathrm{lh}(v)$ denotes the length of the sequence v. (Note that the length of w is implicitly polynomially bounded as $\mathrm{lh}(w) \le |w|$, but we do not impose other restrictions on it.) We also extend the definition so that the empty *partial* function is considered a smooth surjection from any set X on the empty set Y. In many situations, a surjection is automatically smooth (in particular, we will often use (i) without explicit mention):

Observation 2.2 (*in* $T^1_2(\alpha)$) *Let* $X \in \Sigma^b_1(\alpha)$. *A surjective* $PV_2(\alpha)$*-function* $f\colon X \to Y$ *is smooth whenever at least one of the following holds:*

(*i*) f *has a* $PV_2(\alpha)$*-coretraction,*

(*ii*) f *has a* $\Sigma^b_1(\alpha)$*-graph,*

(*iii*) $BB\Sigma^b_2(\alpha)$,

(*iv*) f *is a composition of two smooth surjections.*

(Note in particular that all surjections are smooth in the standard model of arithmetic. Smoothness is only a technical condition needed to compensate for the lack of appropriate instances of BB.) We will write just $X \twoheadrightarrow Y$ if there exists a function $f\colon X \twoheadrightarrow Y$ (of suitable complexity, which should be clear from the context).

We will use the shorthand notation

$$x \in \mathrm{Log} \Leftrightarrow \exists y\, x = |y|.$$

We will also work with rational numbers, which are assumed to be represented by pairs of integers in a natural way. The expression $x^{-1} \in \mathrm{Log}$ is a shorthand notation meaning that x is a positive rational number, whose inverse is bounded above by an integer $n \in \mathrm{Log}$. The symbol $\mathbb{Q}_{\mathrm{Log}}$ denotes the set of rationals whose nominator and denominator belong to Log.

Many of our results take place *inside* formal theories like $T^1_2 + rWPHP(PV_2)$. If T is a theory, a parenthesized expression "in T" in the heading of a definition or theorem indicates that the definition is introduced in T, or that the theorem is formulated and proved inside T. However, we will slightly abuse this convention for reasons of compactness: when we write e.g. "for every PV_2-function f ..." in a formalized context, it is assumed that the quantification over PV_2-functions takes place in the metatheory, and only *parameters* of the function are quantified inside T. Formulas, definable sets, and other non-first-order objects are treated similarly.

In fact, in most cases the sets or functions thus quantified will only have a bounded domain. As already mentioned above, speaking of (say) $PV_2(\alpha)$-functions, or $\Sigma^b_1(\alpha)$-sets in such a context is equivalent to using circuits with a $\Sigma^b_1(\alpha)$-oracle, or non-deterministic Boolean circuits with an oracle α, respectively. We will, however, generally use the former expression, as we believe it is easier to read (even though the latter may be formally more correct). We point out that the reader should think about Σ^b_1-sets as corresponding to *NP/poly*, rather than just *NP* as is usual in bounded arithmetic.

We will also need some notation and results from [21]. For convenience, we state it in a relativized version (which is the one we will actually use); in particular, what we denote $\preceq_\varepsilon$ below is closer to what is denoted by $\preceq^1_\varepsilon$ in [21].

Let $X, Y \subseteq a$ be definable sets, and $\varepsilon \le 1$. We say that the size of X is approximately less than the size of Y with error ε, written as $X \preceq_\varepsilon Y$, if there exists a $PV_2(\alpha)$-function C, and $v \neq 0$, such that

$$C \colon v \times (Y \dot\cup \varepsilon a) \twoheadrightarrow v \times X.$$

The sets X and Y have approximately the same size with error ε, written as $X \approx_\varepsilon Y$, if $X \preceq_\varepsilon Y$ and $Y \preceq_\varepsilon X$. (We recall that we identify a number s with the interval $[0, s)$, thus as a special case, $X \approx_\varepsilon s$ means that the size of X is equal to s with error ε.)

If p is a rational, we also write

$$\Pr_{x<a}(\varphi(x)) \preceq_\varepsilon p \quad \text{iff} \quad \{x < a;\, \varphi(x)\} \preceq_\varepsilon pa,$$

and similarly for $\succeq$, $\approx$. We will often omit the mention of a when it is clear from context. For example, a sequence $\vec{A} = \langle A_i;\, i < k\rangle$ of $t \times n$ binary matrices is encoded by a number $x < 2^{ktn}$, hence we write $\Pr_{\vec{A}}(\varphi(\vec{A}))$ instead of $\Pr_{x<2^{ktn}}(\varphi(\text{the sequence of matrices encoded by } x))$.

Theorem 2.3 ([21, Thm. 2.7]) *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $X \subseteq a$, $X \in \Delta^b_2(\alpha)$, and $\varepsilon^{-1} \in \mathrm{Log}$. There exists a number $s \le a$ such that $X \approx_\varepsilon s$, moreover the surjections required by the definition of $\approx$ have $PV_2(\alpha)$-coretractions, and the numbers v from the definition belong to* Log.

The reader may find it helpful to familiarize her/himself with basic properties of $\preceq_\varepsilon$ from [21, §2].

We will occasionally use some results from [21] on definable randomized algorithms, in particular, AM. Recall that a promise problem is a pair $L = \langle L^+, L^-\rangle$ of disjoint sets of strings (a language $L \subseteq \Sigma^*$ is identified with the promise problem $\langle L, \Sigma^* \smallsetminus L\rangle$). A promise problem L is in *promise $AM(\alpha)$* (*$prAM(\alpha)$* for short), if there exists a probabilistic polynomial-time algorithm $A(x, y)$ with oracle α such that

$$A(x, y) \Rightarrow |y| \le p(|x|)$$

for some polynomial p, and

$$x \in L^+ \Rightarrow \Pr(\exists y\, A(x, y)) \ge 3/4,$$
$$x \in L^- \Rightarrow \Pr(\exists y\, A(x, y)) \le 1/4.$$

A language is in $AM(\alpha)$ if the corresponding promise problem is in $prAM(\alpha)$.

We formalize this definition in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$ as follows. Let β be a PV-function with values in $(0, 1/2)$. A pair $\langle\varphi, r\rangle$, where $\varphi(x, w)$ is a $\Sigma^b_1(\alpha)$-formula, and r is a PV-function, *β-defines a $prAM(\alpha)$ problem* $L = \langle L^+, L^-\rangle$ if $L^+ \supseteq L^+_{\varphi,r,\beta}$ and $L^- \supseteq L^-_{\varphi,r,\beta}$, where

$$\begin{aligned} x \in L^+_{\varphi,r,\beta} &\quad \text{iff} \quad \Pr_{w<r(x)}(\neg\varphi(x, w)) \preceq_0 \beta(x), \\ x \in L^-_{\varphi,r,\beta} &\quad \text{iff} \quad \Pr_{w<r(x)}(\varphi(x, w)) \preceq_0 \beta(x). \end{aligned}$$

The pair $\langle\varphi, r\rangle$ *β-defines an $AM(\alpha)$-language,* if $\forall x\, (x \in L^+_{\varphi,r,\beta} \vee x \in L^-_{\varphi,r,\beta})$. If unspecified, we take $\beta = 1/4$.

The definition is insensitive on the choice of β in the following sense: if t, s are PV-functions such that $t(x), s(x) > 0$ and $1/s(x) + 1/|t(x)| \le 1/2$, then $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$ proves that L is a definable $prAM(\alpha)$-problem iff it is a $(1/2 - 1/|t|)$-definable $prAM(\alpha)$-problem iff it is a $1/s$-definable $prAM(\alpha)$-problem [21, P. 4.3]. We could also use an asymmetric definition with different bounds for L^+ and L^-, but we will not write it down explicitly.

We will need the following statement, formalizing the result that $AM \subseteq NP/poly$.

Theorem 2.4 ([21, P. 4.5]) *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$)*
If L is a $1/4$-definable $prAM(\alpha)$-problem, and $n \in \mathrm{Log}$, then there exists a polynomial-size nondeterministic oracle circuit $C\colon 2^n \to 2$ such that

$$\begin{aligned} x \in L^+ &\Rightarrow C(x) = 1, \\ x \in L^- &\Rightarrow C(x) = 0 \end{aligned}$$

for every $x < 2^n$.

If L_0, L_1 are definable $prAM(\alpha)$-problems, it is easy to see that $L_0 \cap L_1 := \langle L^+_0 \cap L^+_1, L^-_0 \cup L^-_1\rangle$ and $L_0 \cup L_1 := \langle L^+_0 \cup L^+_1, L^-_0 \cap L^-_1\rangle$ are also definable $prAM(\alpha)$-problems. More importantly, definable $prAM(\alpha)$-problems are in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$ closed under bounded existential quantification [21, Thm. 4.4]: if q is a PV-function, and L is a definable $prAM(\alpha)$-problem, then so is $L^\exists := \langle L^{+\exists}, L^{-\forall}\rangle$, where

$$\begin{aligned} x \in L^{+\exists} &\quad \text{iff} \quad \exists y < q(x)\, \langle x, y\rangle \in L^+, \\ x \in L^{-\forall} &\quad \text{iff} \quad \forall y < q(x)\, \langle x, y\rangle \in L^-. \end{aligned}$$

3 The toolbox

We begin with a definition of approximate counting based on Sipser [34]. Rather than defining what is a size of a set, we introduce a predicate $X \precsim_\varepsilon s$ which means that the size of X is bounded above by s (approximately, with relative error ε). (This $\precsim$ should not be confused with $\preceq$.)

The basis of the construction is to use linear hash functions. The idea is as follows. Let $X \subseteq 2^n$, $s = |X|$, and choose a parameter t. We consider a random linear function $A\colon 2^n \to 2^t$

(which is given by a matrix, thus a polynomial-size object). Ideally, we would like A to be injective on X, which would witness that $s \le 2^t$. This is rather unlikely to happen unless t is really huge, but it is possible that A is injective at least on a sizable part of X. Let thus X' be the set of all elements $x \in X$ such that $A(x) \ne A(y)$ for all $y \in X$ different from x, so that A is injective on X'. Elements of X' are called *separated* by A. The probability that $A(x) = A(y)$ for $x \ne y$ is 2^{-t}, hence any $x \in X$ is not separated by A with probability at most $s2^{-t} \le 1/2$, as long as $2^t \ge 2s$. The expected size of X' is thus at least $s/2$. In order to cover all of X, we choose independently random linear functions $A_i \colon 2^n \to 2^t$ for $i < t$. The probability that $x \in X$ is not separated by A_i is at most $1/2$, hence the probability that it is not separated by any A_i, $i < t$, is at most 2^{-t}. The expected number of $x \in X$ not separated by any A_i is thus at most $s2^{-t} \le 1/2$, hence there exist matrices $A_0, \dots, A_{t-1}$ such that *every* $x \in X$ is separated by some A_i. However, the existence of such $\vec{A}$ does not conversely guarantee that $|X| \le s$. Each A_i injects a part of X to 2^t, hence we can inject X into $t2^t$. We may choose $2^t \le 4s$, hence we obtain only an injection of X to $4s \log 4s$.

This form of hashing thus directly distinguishes sets of size s from roughly $4s \log s$. We want to distinguish size s from $s(1+\varepsilon)$ for polynomially small ε; we achieve this by considering Cartesian powers. Instead of our set X, we apply the hashing to X^c for some c. We can distinguish its size s^c from $4s^c \log s^c = 4cs^c \log s$, and the latter is less than $(s(1+\varepsilon))^c$ as long as $(1+\varepsilon)^c \ge 4c \log s$. We have $(1+\varepsilon)^{c_1} \ge 2$ for $c_1 \ge \varepsilon^{-1}$, and $2^{c_2} \ge 4c \log s$ for c_2 about $\log\log s + \log c$, hence it suffices to take roughly $c = \Omega(\varepsilon^{-1}(\log\log s + \log \varepsilon^{-1}))$. We will actually use somewhat larger (but still polynomial in ε^{-1} and $\log s$) c for convenience to simplify some computations below.

Definition 3.1 Let $X \subseteq 2^n$, and $x < 2^n$. A matrix $A \in 2^{t \times n}$ (i.e., a t-by-n matrix over $GF(2)$) *separates* x from X if $Ax \ne Ay$ for all $y \in X \smallsetminus \{x\}$ (where we view elements of 2^n as column vectors over $GF(2)$). A sequence $\vec{A} = \langle A_i;\, i < k \rangle$ of matrices *isolates* X, written as

$$\vec{A} \colon X \looparrowright 2^t,$$

if every $x \in X$ is separated from X by some A_i. Let $\varepsilon^{-1} \in \mathrm{Log}$. If $s > 0$, we write

$$X \precsim_\varepsilon s$$

if there exist $\langle A_i;\, i < t \rangle$ such that $\vec{A} \colon X^c \looparrowright 2^t$, where $c = 12|S|\lceil \varepsilon^{-1} \rceil^2$, and $t = |S^c| + 1$ for some $0 < S \le s$. We also define $X \precsim_\varepsilon 0$ iff X is empty. We write $X \precsim s$ if $X \precsim_\varepsilon s$ for every $\varepsilon^{-1} \in \mathrm{Log}$.

Remark 3.2 If X is $\Sigma^b_1(\alpha)$, then the properties "A separates x from X", and "$\vec{A}$ isolates X" are $\Pi^b_1(\alpha)$, hence $X \precsim_\varepsilon s$ is $\Sigma^b_2(\alpha)$.

The definition makes $\precsim$ monotone: if $Y \subseteq X \precsim_\varepsilon s \le t$, then $Y \precsim_\varepsilon t$.

If $X \subseteq 2^n$, $n < m$, $\vec{A} \in 2^{t \times m}$, and $\vec{B} \in 2^{t \times n}$ is the sequence of restrictions of A_i's to the first n columns, then $\vec{A}$ isolates X iff $\vec{B}$ does. The definition of $X \precsim_\varepsilon s$ thus does not depend on the choice of n.

A moment's reflection will persuade the reader that it is next to impossible to work directly with the hash functions. For example, if $\vec{A} \colon X \looparrowright 2^t$, and $\vec{B} \colon Y \looparrowright 2^t$, there is apparently no way of constructing $\vec{C}$ such that, say, $\vec{C}\colon X \cup Y \looparrowright 2^{t+1}$. In the real world, this is no problem as we have a well-behaved preexisting notion of cardinality, and we merely observe that the hashes agree with it. Obviously, this does not work in bounded arithmetic if we want to use the hashes to define (approximate) cardinality in the first place. We get around the problem by showing that $X \precsim_\varepsilon s$ is, up to ε, equivalent to the existence of suitable surjections from a power of s to a corresponding power of X; these surjections will be much easier to handle. The key result is Theorem 3.4 (the other direction will be much simpler), which is essentially a formalization of Sipser's Coding Lemma in bounded arithmetic.

Lemma 3.3 *(in T^0_2) If $c \in \mathrm{Log}$, there exists a PV-bijection*

$$f\colon \dot{\bigcup_{i \le c}} \binom{c}{i} \times X^i \times Y^{c-i} \simeq (X \mathbin{\dot\cup} Y)^c$$

with a PV-inverse.

Proof: Let $u < \binom{c}{i}$, $\langle x_j;\ j < i\rangle \in X^i$, and $\langle y_j;\ j < c-i\rangle \in Y^{c-i}$. We can enumerate subsets of c of size i by $\binom{c}{i}$, let thus $U \subseteq c$ be the uth set. Let $\langle \pi_j;\ j < i\rangle$ be an increasing enumeration of U, and $\langle \varrho_j;\ j < c-i\rangle$ an increasing enumeration of $c \smallsetminus U$. We define $f(u, \vec{x}, \vec{y}) = \vec{z}$, where $z_{\pi_j} = x_j$, $z_{\varrho_j} = y_j$. It is easy to see that f is a bijection. □

Theorem 3.4 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $d, r > 0$, $d \in \mathrm{Log}$, and $f\colon rs^d \twoheadrightarrow r \times X^d$, where X is $\Sigma^b_1(\alpha)$, and f is $PV_2(\alpha)$. Then there exists $\langle A_i;\ i < t\rangle$ such that $\vec{A}\colon X \looparrowright 2^t$, where $t = |s| + 1$, and moreover,*

$$\Pr_{\vec{A}}\bigl(\vec{A} \textit{ does not isolate } X\bigr) \preceq_0 2/3.$$

Proof: Let B be the set of sequences $\langle A_i;\ i < t\rangle$, $A_i \in 2^{t \times n}$, such that $\vec{A}$ does not isolate X. We define a PV-function

$$g_0\colon (2^n \smallsetminus \{0\}) \times 2^{n-1} \to 2^n$$

as follows. Let $i < n$ be the index of the first set bit of x. We decompose $w = w_0 \frown w_1$, where $w_0 < 2^i$ and $w_1 < 2^{n-i-1}$, and we put $g_0(x, w) = w_0 \frown b \frown w_1$, where $b = x^{\mathrm{T}}(w_0 \frown 0 \frown w_1)$. Clearly, $g_0(x, \cdot)$ is a surjection of 2^{n-1} onto $\{a \in 2^n;\ a^{\mathrm{T}} x = 0\}$ whenever $x \neq 0$. Then we can define a PV-function

$$g\colon (2^n \smallsetminus \{0\}) \times 2^{(n-1)t} \to 2^{t \times n}$$

so that $g(x, \langle w_0, \dots, w_{t-1}\rangle)$ is the matrix A such that the jth row of A is $g_0(x, w_j)^{\mathrm{T}}$ for every $j < t$. It follows that

$$g(x, \cdot)\colon 2^{(n-1)t} \twoheadrightarrow \{A \in 2^{t \times n};\ Ax = 0\}$$

for every $x \neq 0$.

We define a $PV_2(\alpha)$-function

$$h\colon r^{t+1}\bigl(s^{t+1} 2^{(n-1)t^2}\bigr)^d \to r^{t+1}\bigl(2^{t \times n}\bigr)^{td}$$

as follows. We interpret the input to h as sequence consisting of u, $\langle v_i;\ i < t\rangle$, and $\langle w_{i,j};\ i < t, j < d\rangle$, where $u, v_i < rs^d$, $w_{i,j} < 2^{(n-1)t}$. We compute $f(u) = \langle p, x_j;\ j < d\rangle \in r \times X^d$, and in a similar way, $f(v_i) = \langle q_i, y_{i,j};\ j < d\rangle$. For each $i < t$ and $j < d$, we define $A_{i,j} = g(x_j + y_{i,j}, w_{i,j})$ (where $+$ is vector addition) if $x_j \neq y_{i,j}$, and $A_{i,j} = 0$ otherwise. We let $\langle p, q_i, A_{i,j};\ i < t, j < d\rangle$ be the output of h.

We claim that h is a surjection of $r^{t+1}\big(s^{t+1}2^{(n-1)t^2}\big)^d$ onto $r^{t+1} \times B^d$. Indeed, consider a sequence $\langle p, q_i, A_{i,j};\ i < t, j < d\rangle \in r^{t+1} \times B^d$. For each $j < d$, there exists an $x_j \in X$ which is not separated from X by $\langle A_{i,j};\ i < t\rangle$; we can collect them to a sequence $\langle x_j;\ j < d\rangle$ by $BB\Sigma^b_1(\alpha) \subseteq T^1_2(\alpha)$. Likewise, there exists a sequence $\langle y_{i,j};\ i < t, j < d\rangle$ of witnesses to the non-separation of x_j by $A_{i,j}$, i.e., $y_{i,j} \neq x_j$, and $A_{i,j}x_j = A_{i,j}y_{i,j}$. The latter is equivalent to $A_{i,j}(x_j + y_{i,j}) = 0$, and as $x_j + y_{i,j} \neq 0$, we can use the properties of g to find a sequence $\langle w_{i,j};\ i < t, j < d\rangle$ such that $g(x_j + y_{i,j}, w_{i,j}) = A_{i,j}$. As f is surjective, we can find a $u < rs^d$ such that $f(u) = \langle p, x_j;\ j < d\rangle$. We construct suitable v_i in a similar way, using smoothness of f. Then $h(u, \vec v, \vec w) = \langle p, \vec q, \vec A\rangle$.

As $s \leq 2^{t-1}$, we have

$$r^{t+1}\big(s^{t+1}2^{(n-1)t^2}\big)^d \leq 2^{-d}r^{t+1}2^{nt^2d},$$

thus $sWPHP(PV_2(\alpha))$ implies that h is not onto $r^{t+1} \times (2^{t\times n})^{td}$, hence $B \neq (2^{t\times n})^t$. Any $\vec A \in (2^{t\times n})^t \smallsetminus B$ isolates X.

As B is $\Sigma^b_1(\alpha)$, and we assume $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$, there exists a b such that $B \approx_{1/20} b$ by Theorem 2.3. By definition, there exists $0 < e \in \mathrm{Log}$, and a $PV_2(\alpha)$-surjection

$$e \times \left(B \mathbin{\dot\cup} \tfrac{1}{20}2^{nt^2}\right) \twoheadrightarrow eb.$$

For any $k \in \mathrm{Log}$, we can thus construct a chain of surjections

$$\begin{aligned}
&r^{(t+1)\lceil k/d\rceil}e^k\sum_{i=0}^{k}\binom{k}{i}2^{(nt^2-1)d\lceil i/d\rceil}\left(\tfrac{1}{20}2^{nt^2}\right)^{k-i} = \\
&\qquad = e^k\sum_{i=0}^{k}\binom{k}{i}\left(r^{t+1}2^{(nt^2-1)d}\right)^{\lceil i/d\rceil}\left(r^{t+1}\right)^{\lceil k/d\rceil-\lceil i/d\rceil}\left(\tfrac{1}{20}2^{nt^2}\right)^{k-i} \twoheadrightarrow \\
&\qquad\qquad \twoheadrightarrow e^k\dot{\bigcup_{i\leq k}}\binom{k}{i}r^{(t+1)\lceil k/d\rceil}B^i\left(\tfrac{1}{20}2^{nt^2}\right)^{k-i} \simeq \\
&\qquad\qquad \simeq r^{(t+1)\lceil k/d\rceil}\dot{\bigcup_{i\leq k}}\binom{k}{i}(eB)^i\left(\tfrac{1}{20}e2^{nt^2}\right)^{k-i} \twoheadrightarrow \\
&\qquad\qquad\qquad \twoheadrightarrow r^{(t+1)\lceil k/d\rceil}\left(e\left(B \mathbin{\dot\cup} \tfrac{1}{20}2^{nt^2}\right)\right)^k \twoheadrightarrow r^{(t+1)\lceil k/d\rceil}e^kb^k,
\end{aligned}$$

where the surjection from the second to the third line is constructed using $h\colon r^{t+1}2^{(nt^2-1)d} \twoheadrightarrow r^{t+1}B^d$, and the last but one surjection follows from Lemma 3.3. We have

$$\begin{aligned}
\sum_{i=0}^{k}\binom{k}{i}2^{(nt^2-1)d\lceil i/d\rceil}\left(\tfrac{1}{20}2^{nt^2}\right)^{k-i} &\leq 2^{nt^2(k+d)}\sum_{i=0}^{k}\binom{k}{i}2^{-i}20^{-(k-i)} \\
&= 2^{nt^2(k+d)}(11/20)^k = \left(\tfrac{11}{20}2^{nt^2}\right)^k 2^{nt^2d} \leq \tfrac{1}{2}\left(\tfrac{12}{20}2^{nt^2}\right)^k
\end{aligned}$$

as long as $k \geq 8(nt^2d + 1)$, hence $b \leq (12/20)2^{nt^2}$ by $sWPHP(PV_2(\alpha))$. As $B \preceq_{1/20} b$, we have $B \preceq_0 b + (1/20)2^{nt^2} < (2/3)2^{nt^2}$. □

Corollary 3.5 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $d, r > 0$, $d, \varepsilon^{-1} \in \mathrm{Log}$, and $f\colon rs^d \twoheadrightarrow r \times X^d$, where X is $\Sigma^b_1(\alpha)$, and f is $PV_2(\alpha)$. Then $X \precsim s$, and moreover,*

$$\Pr_{\vec{A}}\bigl(\vec{A} \textit{ does not isolate } X^c\bigr) \preceq_0 2/3,$$

where $c = 12|s|\lceil\varepsilon^{-1}\rceil^2$, $t = |s^c| + 1$, $X \subseteq 2^n$, and $\vec{A} = \langle A_i;\, i < t\rangle$ is a sequence of matrices $A_i \in 2^{t\times n}$ as in Definition 3.1.

Remark 3.6 If we assume that f has a $PV_2(\alpha)$-coretraction, the existence of $\vec{A}$ in Theorem 3.4 and Corollary 3.5 is provable even in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$, as the statement becomes $\forall\Sigma^b_2(\alpha)$. This is quite typical behaviour. To avoid unnecessary cluttering of the text, we will only indicate provability in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$ below if it applies to an unmodified statement of a theorem, or if it does not directly follow from Theorem 2.1.

Lemma 3.7 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $s, \varepsilon^{-1}, c \in \mathrm{Log}$, $c > 0$, $X \in \Sigma^b_1(\alpha)$. If there exists a $PV_2(\alpha)$-surjection $f\colon \lfloor(s + 1 - \varepsilon)^c\rfloor \twoheadrightarrow X^c$, then there exists a surjection $s \twoheadrightarrow X$ (encoded by a sequence, hence PV-definable).*

Proof: Let $k \le s + 1$ be maximal such that there exists a sequence of length k of pairwise distinct elements of X (by $\Sigma^b_1(\alpha)$-*LMAX* $\subseteq T^1_2(\alpha)$). If $k \le s$, we have $s \twoheadrightarrow X$. Otherwise $X \twoheadrightarrow s + 1$, which implies

$$(s+1)^c\left(1 - \frac{\varepsilon}{s+1}\right) \ge (s + 1 - \varepsilon)^c \twoheadrightarrow X^c \twoheadrightarrow (s+1)^c,$$

contradicting *sWPHP*. □

Theorem 3.8 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) If X is $\Sigma^b_1(\alpha)$, and $X \precsim_\varepsilon s$, there exists a $PV_2(\alpha)$-function f such that $f\colon \lfloor s(1+\varepsilon)\rfloor^c \twoheadrightarrow X^c$ (with a $PV_2(\alpha)$-coretraction), where $0 < c \le 12|s|\lceil\varepsilon^{-1}\rceil^2$.*

Proof: W.l.o.g. $s = S$ in the notation of Definition 3.1. The case $s \le 1$ is left to the reader, we assume $s \ge 2$. Let $a = 4|s|\lceil\varepsilon^{-1}\rceil$, $c = 3\lceil\varepsilon^{-1}\rceil a$, and fix $x_0 \in X^c$, and $\vec{A}\colon X^c \looparrowright 2^t$, where $t = |s^c| + 1$. We define a mapping $f\colon t \times 2^t \to X^c$ by

$$f(i, u) = \begin{cases} x & \text{if } x \in X^c,\ A_i x = u \text{ and } A_i \text{ separates } x \text{ from } X^c, \\ x_0 & \text{otherwise.} \end{cases}$$

The definition of $\looparrowright$ ensures that f is onto; it has a $PV_2(\alpha)$-coretraction defined by

$$g(x) = \langle i, A_i x\rangle, \qquad i = \min\{i < t;\ A_i \text{ separates } x \text{ from } X^c\}.$$

The function f is itself $PV_2(\alpha)$, as it is computable by the following algorithm: if

$$\neg\exists x \in X^c\, A_i x = u \vee \exists x, x' \in X^c\, (A_i x = A_i x' = u \wedge x \ne x')$$

(these are $\Sigma^b_1(\alpha)$ oracle calls), output x_0. Otherwise there exists a unique x satisfying the $\Sigma^b_1(\alpha)$-condition $x \in X^c \wedge A_i x = u$, and we can find it by binary search.

As $a \geq 4|s| \geq 8$, we have $2^a \geq 3a^2+4$. Moreover $(1+\varepsilon/3)^{\lceil 3/\varepsilon\rceil} \geq 2$, hence

$$\begin{aligned} t2^t &\leq 4s^c(|s^c|+1) \leq 4s^c(c|s|+1) = s^c(12|s|\lceil\varepsilon^{-1}\rceil a+4) \\ &\qquad\qquad \leq s^c(3a^2+4) \leq s^c 2^a \leq s^c(1+\varepsilon/3)^{\lceil 3\varepsilon^{-1}\rceil a} \leq (s(1+\varepsilon/3))^c. \end{aligned}$$

If $s \geq 3/(2\varepsilon)$, we obtain

$$(s(1+\varepsilon/3))^c \leq (s(1+\varepsilon)-1)^c \leq \lfloor s(1+\varepsilon)\rfloor^c.$$

If $s \leq 3/(2\varepsilon)$, we have $s(1+\varepsilon/3) \leq s+1/2$, and $s \in \mathrm{Log}$, hence $s \twoheadrightarrow X$ by Lemma 3.7 (which works in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$, as our surjection has a $PV_2(\alpha)$-coretraction). □

The proof of Theorem 3.8 actually shows the following:

Corollary 3.9 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) If $X \in \Sigma^b_1(\alpha)$ and $X \looparrowright 2^t$, there exists a $PV_2(\alpha)$-surjection $f\colon t2^t \twoheadrightarrow X$ with a $PV_2(\alpha)$-coretraction.*

The corollary below states the important principle that approximate counting with small error reduces to exact counting whenever the latter is possible.

Corollary 3.10 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) Let $X \in \Sigma^b_1(\alpha)$, and $s \leq \varepsilon^{-1} \in \mathrm{Log}$. We have $X \precsim_\varepsilon s$ iff there exists a sequence of length at most s which includes all elements of X.*

Proof: If w is such a sequence, then $w\colon s \twoheadrightarrow X$, hence $X \precsim s$ by Corollary 3.5. (We can use $sWPHP$ by Theorem 2.1.) On the other hand, $X \precsim_\varepsilon s$ implies the existence of w by the proof of Lemma 3.7 and Theorem 3.8. □

The following corollary serves several purposes. First, it shows the basic counting principle that upper bounds on cardinality are preserved by surjections. Second, it shows that the present approximate counting generalizes the method of [21] in the following sense: if $Y \subseteq 2^n$ and $Y \preceq_\varepsilon s$, then $Y \precsim s + 2\varepsilon 2^n$. Finally, we will often use the special case when f is the identity function to reduce the error of approximation in favor of worse bounds: $X \precsim_\varepsilon s$ implies $X \precsim_\delta \lfloor s(1+\varepsilon)\rfloor$ for every $\delta^{-1} \in \mathrm{Log}$.

Corollary 3.11 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $X, Y \in \Sigma^b_1(\alpha)$, $f \in PV_2(\alpha)$, $d, \varepsilon^{-1} \in \mathrm{Log}$, $d, r > 0$. If $X \precsim_\varepsilon s$, and $f\colon r \times X^d \twoheadrightarrow r \times Y^d$, then $Y \precsim \lfloor s(1+\varepsilon)\rfloor$.*

Proof: We have $\lfloor s(1+\varepsilon)\rfloor^c \twoheadrightarrow X^c$ for some c by Theorem 3.8, hence $r^c\lfloor s(1+\varepsilon)\rfloor^{cd} \twoheadrightarrow r^c X^{cd} \twoheadrightarrow r^c Y^{cd}$, thus $Y \precsim \lfloor s(1+\varepsilon)\rfloor$ by Corollary 3.5. □

The next two results state fundamental counting principles for computing upper bounds on the size of Cartesian products and unions.

Corollary 3.12 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) If $X, Y \in \Sigma^b_1(\alpha)$, $X \precsim_\varepsilon s$, and $Y \precsim_\varepsilon t$, then $X \times Y \precsim \lfloor st(1+\varepsilon)^2\rfloor$.*

Proof: Use Corollary 3.5, and Theorems 3.8 and 2.1. □

Theorem 3.13 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) If $X, Y \in \Sigma^b_1(\alpha)$, $X \precsim_\varepsilon s$, and $Y \precsim_\varepsilon t$, then $X \cup Y \precsim \lfloor (s+t)(1+2\varepsilon) \rfloor$.*

Proof: We can use *sWPHP* by Theorem 2.1. Take $PV_2(\alpha)$-functions $f\colon S^c \twoheadrightarrow X^c$, and $g\colon T^d \twoheadrightarrow Y^d$ by Theorem 3.8, where $S = \lfloor s(1+\varepsilon)\rfloor$, $T = \lfloor t(1+\varepsilon)\rfloor$. Put $k = 3(c|s|+d|t|)\lceil \varepsilon^{-1}\rceil$. Using Lemma 3.3, we can construct smooth $PV_2(\alpha)$-surjections

$$(X \cup Y)^k \twoheadleftarrow \dot{\bigcup_{i\le k}} \binom{k}{i} X^i Y^{k-i} \twoheadleftarrow \sum_{i \le k} \binom{k}{i} S^{c\lceil i/c\rceil} T^{d\lceil (k-i)/d\rceil} \le S^c T^d \sum_{i\le k} \binom{k}{i} S^i T^{k-i} =$$
$$= S^c T^d (S+T)^k \le 2^{c|S|+d|T|}(S+T)^k \le \big((1+\varepsilon/3)(S+T)\big)^k$$

as $(1+\varepsilon/3)^{\lceil 3\varepsilon^{-1}\rceil} \ge 2$. If

$$(1+\varepsilon/3)(1+\varepsilon)(s+t) \le (s+t)(1+2\varepsilon) - 1 \le \lfloor (s+t)(1+2\varepsilon)\rfloor,$$

we are done by Corollary 3.5. Otherwise $s, t \in \mathrm{Log}$, in which case $\lfloor (s+t)(1+2\varepsilon)\rfloor \twoheadrightarrow X \cup Y$ by Lemma 3.7. □

Theorem 3.13 is one of the most important elementary counting principles. Its dual, which says that the size of a disjoint union $X \mathbin{\dot\cup} Y$ is (approximately) bounded below by the sum of the sizes of X and Y (it has to be formulated contrapositively, see Theorem 3.17), is just as fundamental, but it is considerably harder to prove in our setting. To see why, consider the case where $X \mathbin{\dot\cup} Y = [0, a)$: the obvious fact that $X \cup Y \precsim_\varepsilon a$ does not give us any useful information, hence we must be ready to produce out of thin air a function witnessing that the size of X or Y is (approximately) at most $a/2$. Theorem 2.3 comes to our rescue, as production of magic surjections is exactly what it is good for.

But first we state another consequence of [21]. It allows us to reduce the complexity of $\precsim_\varepsilon$ from $\Sigma^b_2(\alpha)$ to $\Pi^b_1(\alpha)$ in many situations, which is indispensable in proofs by induction (notice that our favourite theory has induction only for $\mathcal{B}(\Sigma^b_1(\alpha))$-formulas). We recall that this does not imply any fancy derandomization of *AM*, as $\Pi^b_1(\alpha)$ here has the meaning of *coNP/poly*, not *coNP* (see Section 2).

Recall Definition 3.1.

Lemma 3.14 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $c \in \mathrm{Log}$, $X \subseteq a \times 2^n$, $X \in \Sigma^b_1(\alpha)$, and put $X_x = \{y < 2^n;\ \langle x, y\rangle \in X\}$ for each $x < a$. There exists a $\Pi^b_1(\alpha)$-predicate C such that*

$$C(x,t) \Rightarrow X_x \hookrightarrow 2^t,$$
$$\neg C(x,t) \Rightarrow \Pr_{A_0,\dots,A_{t-1}\in 2^{t\times n}}\big(\vec{A} \text{ isolates } X_x\big) \preceq_0 1/4$$

for every $x < a$, $t < c$.

Proof: The promise problem $L = \langle L^+, L^-\rangle$, where

$$\langle x,t\rangle \in L^+ \Leftrightarrow \Pr_{\vec{A}}\big(\vec{A} \text{ isolates } X_x\big) \preceq_0 1/8,$$
$$\langle x,t\rangle \in L^- \Leftrightarrow \Pr_{\vec{A}}\big(\vec{A} \text{ isolates } X_x\big) \succeq_0 1/4,$$

is a definable $prAM(\alpha)$-problem, hence the existence of C follows from Theorem 2.4. □

Lemma 3.15 *(in $T^1_2(\alpha)+sWPHP(PV_2(\alpha))$) Let $X\in\Sigma^b_1(\alpha)$, $X\ne\varnothing$. There exists a $t\in\mathrm{Log}$ such that $X \looparrowright 2^{t+2}$, and for every $\varepsilon^{-1}\in\mathrm{Log}$ there exists a positive $r\in\mathrm{Log}$, and a smooth $PV_2(\alpha)$-surjection $f\colon r\times(X\mathbin{\dot\cup}\varepsilon 2^t)\twoheadrightarrow r2^t$ with a $PV_2(\alpha)$-coretraction.*

Proof: Assume $X\subseteq 2^n$. Let C be as in Lemma 3.14, find the minimal $k\le n$ such that $C(k)=1$ by $PV_2(\alpha)$-induction, and put $t=k-2$. We have $X\looparrowright 2^k$. Take $g\colon k2^k\twoheadrightarrow X$ and its coretraction $h\colon X\to k2^k$ from Corollary 3.9, and define

$$A=\{u<k2^k;\, h(g(u))=u\}=\mathrm{rng}(h).$$

Let $\eta=\varepsilon/4n$, and $A\approx_\eta a$ by Theorem 2.3. We have $g\colon A\simeq X$ with inverse $h\colon X\simeq A$, and there exists a $PV_2(\alpha)$-surjection $r(a+\eta k2^k)\twoheadrightarrow rA$ with a $PV_2(\alpha)$-coretraction for some $r>0$, $r\in\mathrm{Log}$, hence

$$r(a+\eta k2^k)\twoheadrightarrow r\times X.$$

By minimality of k, and Theorem 3.4 we have

$$a+\eta k2^k\ge 2^t.$$

There exists a $PV_2(\alpha)$-surjection $r(A\mathbin{\dot\cup}\eta k2^k)\twoheadrightarrow ra$ with a $PV_2(\alpha)$-coretraction by Theorem 2.3. We compose it with h to obtain $r(X\mathbin{\dot\cup}\eta k2^k)\twoheadrightarrow ra$, hence

$$r(X\mathbin{\dot\cup}\varepsilon 2^t)\supseteq r(X\mathbin{\dot\cup}2\eta k2^k)\twoheadrightarrow r(a+\eta k2^k)\ge r2^t. \qquad \square$$

Theorem 3.16 *(in $T^1_2(\alpha)+rWPHP(PV_2(\alpha))$) If $X,Y\in\Sigma^b_1(\alpha)$, and $X\times Y\precsim_\varepsilon st$, then $X\precsim\lfloor s(1+\varepsilon)\rfloor$, or $Y\precsim\lfloor t(1+\varepsilon)\rfloor$.*

Proof: We can use $sWPHP$ by Theorem 2.1. Assume that $X\ne\varnothing\ne Y$, and take a $PV_2(\alpha)$-function $f\colon\lfloor st(1+\varepsilon)\rfloor^d\twoheadrightarrow(X\times Y)^d$ by Theorem 3.8. Let $0<c\in\mathrm{Log}$, and take $k,\ell,r\in\mathrm{Log}$ such that $X^{cd}\looparrowright 2^{k+2}$, $r(X^{cd}\mathbin{\dot\cup}\eta 2^k)\twoheadrightarrow r2^k$, $Y^{cd}\looparrowright 2^{\ell+2}$, $r(Y^{cd}\mathbin{\dot\cup}\eta 2^\ell)\twoheadrightarrow r2^\ell$ by Lemma 3.15, where $\eta=(8(k+\ell+4))^{-1}$. By Corollary 3.9, there are smooth $PV_2(\alpha)$-surjections $(k+2)2^{k+2}\twoheadrightarrow X^{cd}$, $(\ell+2)2^{\ell+2}\twoheadrightarrow Y^{cd}$. As

$$\eta(k+2)2^{k+\ell+2}+\eta(\ell+2)2^{k+\ell+2}+\eta^2 2^{k+\ell}\le\tfrac58 2^{k+\ell},$$

we can construct smooth $PV_2(\alpha)$-surjections

$$\begin{aligned}r^2(X^{cd}\times Y^{cd}\mathbin{\dot\cup}\tfrac58 2^{k+\ell})&\twoheadrightarrow r^2(X^{cd}\times Y^{cd}\mathbin{\dot\cup}\eta 2^kY^{cd}\mathbin{\dot\cup}\eta 2^\ell X^{cd}\mathbin{\dot\cup}\eta^2 2^{k+\ell})\\&\simeq r(X^{cd}\mathbin{\dot\cup}\eta 2^k)\times r(Y^{cd}\mathbin{\dot\cup}\eta 2^\ell)\twoheadrightarrow r^2 2^{k+\ell}.\end{aligned}$$

On the other hand, $f^{(c)}\colon\lfloor st(1+\varepsilon)\rfloor^{cd}\twoheadrightarrow X^{cd}\times Y^{cd}$, hence

$$r^2\big(\lfloor st(1+\varepsilon)\rfloor^{cd}+\tfrac58 2^{k+\ell}\big)\twoheadrightarrow r^2 2^{k+\ell},$$

which implies

$$(st(1+\varepsilon))^{cd}\ge 2^{k+\ell}(1-\tfrac68)=2^{k-1}2^{\ell-1}$$

by $sWPHP(PV_2(\alpha))$. Therefore $(s(1+\varepsilon/2))^{cd} \geq 2^{k-1}$ or $(t(1+\varepsilon/2))^{cd} \geq 2^{\ell-1}$, hence

$$\begin{aligned}
&8(s(1+\varepsilon/2))^{cd}\big(cd|s(1+\varepsilon/2)|+3\big) \geq 8(s(1+\varepsilon/2))^{cd}|4(s(1+\varepsilon/2))^{cd}| \geq 2^{k+2}(k+2)\\
&\quad\text{or}\quad 8(t(1+\varepsilon/2))^{cd}\big(cd|t(1+\varepsilon/2)|+3\big) \geq 8(t(1+\varepsilon/2))^{cd}|4(t(1+\varepsilon/2))^{cd}| \geq 2^{\ell+2}(\ell+2),
\end{aligned}$$

which implies

$$8(s(1+\varepsilon/2))^{cd}\big(cd|s(1+\varepsilon/2)|+3\big) \twoheadrightarrow X^{cd} \quad\text{or}\quad 8(t(1+\varepsilon/2))^{cd}\big(cd|t(1+\varepsilon/2)|+3\big) \twoheadrightarrow Y^{cd}.$$

We may fix $c \in \mathrm{Log}$ so that

$$8\big(cd|\max\{s,t\}(1+\varepsilon/2)|+3\big) \leq (1+\varepsilon/4)^{cd},$$

hence there exists a $PV_2(\alpha)$-function

$$\begin{aligned}
&\Big(s\left(1+\tfrac{7}{8}\varepsilon\right)\Big)^{cd} \geq \big(s(1+\varepsilon/2)(1+\varepsilon/4)\big)^{cd} \twoheadrightarrow X^{cd}\\
&\qquad\qquad\text{or}\quad \Big(t\left(1+\tfrac{7}{8}\varepsilon\right)\Big)^{cd} \geq \big(t(1+\varepsilon/2)(1+\varepsilon/4)\big)^{cd} \twoheadrightarrow Y^{cd}.
\end{aligned}$$

Then

$$X \precsim \lfloor s(1+\varepsilon)\rfloor \quad\text{or}\quad Y \precsim \lfloor t(1+\varepsilon)\rfloor$$

by Corollary 3.5 and Lemma 3.7. □

Theorem 3.17 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) If $X, Y \in \Sigma^b_1(\alpha)$, and $X \mathbin{\dot\cup} Y \precsim_\varepsilon s+t+1$, then $X \precsim \lfloor s(1+2\varepsilon)\rfloor$, or $Y \precsim \lfloor t(1+2\varepsilon)\rfloor$.*

Proof: We can use $sWPHP(PV_2(\alpha))$ and $BB\Sigma^b_3(\alpha)$ by Theorem 2.1. W.l.o.g. assume $s \leq t$. Put $S = s(1+\varepsilon)$, $T = (t+1)(1+\varepsilon)$. We fix a surjection $\lfloor S+T\rfloor^c \twoheadrightarrow (X \mathbin{\dot\cup} Y)^c$ by Theorem 3.8. Let $d \in \mathrm{Log}$ be such that $8cd|6(S+T)| \leq (1+\varepsilon/4)^{cd}$, and put $\eta = 1/4$. For each $i \leq cd$, we fix $k_i \leq cd|S+T|$ such that $X^i \times Y^{cd-i} \looparrowright 2^{k_i+2}$, and $PV_2(\alpha)$-functions $f_i \colon r(X^i \times Y^{cd-i} \mathbin{\dot\cup} \eta 2^{k_i}) \twoheadrightarrow r2^{k_i}$, by Lemma 3.15 and $BB\Sigma^b_3(\alpha)$. Then we can construct surjections

$$r\bigg((S+T)^{cd} + \eta\sum_i \binom{cd}{i}2^{k_i}\bigg) \twoheadrightarrow r\dot{\bigcup_i}\binom{cd}{i}(X^i \times Y^{cd-i} \mathbin{\dot\cup} \eta 2^{k_i}) \twoheadrightarrow r\sum_i\binom{cd}{i}2^{k_i}$$

using Lemma 3.3. By $sWPHP(PV_2(\alpha))$,

$$(1-\eta/4)r\sum_i\binom{cd}{i}2^{k_i} \leq r\bigg((S+T)^{cd} + \eta\sum_i\binom{cd}{i}2^{k_i}\bigg),$$

hence

$$\sum_i\binom{cd}{i}2^{k_i} \leq (S+T)^{cd}(1+2\eta) = \sum_i\binom{cd}{i}S^iT^{cd-i}(1+2\eta)$$

by $sWPHP(PV_2(\alpha))$, which implies

$$2^{k_i} \leq S^iT^{cd-i}(1+2\eta) \leq \tfrac{3}{2}S^iT^{cd-i}$$

for some $i \le cd$. It follows that

$$\tfrac{3}{4}S^iT^{cd-i}(1+\varepsilon/4)^{cd} \ge 6S^iT^{cd-i}|6S^iT^{cd-i}| \ge (k_i+2)2^{k_i+2} \twoheadrightarrow X^i \times Y^{cd-i}$$

using Corollary 3.9, hence $X^i \times Y^{cd-i} \precsim \frac{3}{4}S^iT^{cd-i}(1+\varepsilon/4)^{cd}$ by Corollary 3.5, which implies

$$X^i \precsim (S(1+\varepsilon/4))^i \quad \text{or} \quad Y^{cd-i} \precsim (T(1+\varepsilon/4))^{cd-i}$$

by Theorem 3.16. We obtain a $PV_2(\alpha)$-function

$$\big(s(1+\varepsilon)(1+\varepsilon/3)\big)^{ie} \twoheadrightarrow X^{ie} \quad \text{or} \quad \big((t+1)(1+\varepsilon)(1+\varepsilon/3)\big)^{(cd-i)e} \twoheadrightarrow Y^{(cd-i)e}$$

for some $e \in \mathrm{Log}$ by Theorem 3.8, hence

$$X \precsim \lfloor s(1+2\varepsilon)\rfloor \quad \text{or} \quad Y \precsim \lfloor (t+1)(1+\tfrac{9}{5}\varepsilon)\rfloor$$

by Corollary 3.5 and Lemma 3.7. If

$$(t+1)(1+\tfrac{9}{5}\varepsilon) \le t(1+2\varepsilon),$$

we are done. Otherwise $s \le t < 5\varepsilon^{-1} + 9 \in \mathrm{Log}$, hence the result follows by exact counting, using Corollary 3.10. □

We formulated the key theorems 3.12, 3.13, 3.16, and 3.17 for binary sums and products. It is straightforward to generalize them to sums and products of logarithmically many sets, using simple induction.

Corollary 3.18 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) Let $n, \varepsilon^{-1}, \delta^{-1} \in \mathrm{Log}$, and let $\{X_i;\, i < n\}$ be a $\Sigma^b_1(\alpha)$ parametric family of subsets of some 2^m.*

(i) If $X_i \precsim_\varepsilon s_i$ for every $i < n$, then $\bigcup_{i<n} X_i \precsim \big\lfloor (1+2\varepsilon)\sum_{i<n} s_i \big\rfloor$.

(ii) If $X_i \precsim_\varepsilon s_i$ for every $i < n$, then $\prod_{i<n} X_i \precsim \big\lfloor (1+\varepsilon)^{n+1} \prod_{i<n} s_i \big\rfloor$.

(iii) If $\dot{\bigcup}_{i<n} X_i \succsim_\varepsilon \sum_{i<n} s_i - 1$, there exists $i < n$ such that $X_i \succsim \lfloor s_i(1+2\varepsilon)\rfloor - 1$.

(iv) If $n > 0$, and $\prod_{i<n} X_i \succsim_\varepsilon \prod_{i<n} s_i$, there exists $i < n$ such that $X_i \succsim \lfloor s_i(1+\varepsilon)\rfloor$.

Proof: (i): Let $X = \bigcup_i X_i$, and $s = \sum_i s_i$. We have $X_i \precsim_\eta \lfloor s_i(1+\varepsilon)\rfloor$ by Corollary 3.11, hence

$$X \precsim_\eta \big\lfloor s(1+\varepsilon)(1+2\eta)^{n-1} \big\rfloor$$

by induction on n from Theorem 3.13 (we can make the induction hypothesis $\Pi^b_1(\alpha)$ using Lemma 3.14). We choose $\eta = \varepsilon/(12n)$ so that $(1+2\eta)^n \le (1+\varepsilon/3)$. We have

$$\big(s(1+\varepsilon)(1+\varepsilon/3)\big)^c \twoheadrightarrow X^c$$

by Theorem 3.8, hence $X \precsim \lfloor s(1+2\varepsilon)\rfloor$ by Corollary 3.5 and Lemma 3.7.

The other items are proved in a similar way, using Corollary 3.12, and Theorems 3.16 and 3.17. □

We also prove versions of Theorems 3.13 and 3.17 which apply to a "large" number of summands with a uniform description. They can be thought of as averaging arguments: if there are more than st objects in a rectangle of length s, some column must hold at least the average, which is more than $st/s = t$.

Theorem 3.19 *(in $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$) If $X, Y, Z \in \Sigma^b_1(\alpha)$, $Z \subseteq X \times Y$, $X \precsim_\varepsilon s$, and $\{y \in Y;\ \langle x, y\rangle \in Z\} \precsim_\varepsilon t$ for every $x \in X$, then $Z \precsim \lfloor st(1 + 4\varepsilon)\rfloor$.*

Proof: Assume $X \subseteq 2^n$, and fix $\eta^{-1} \in \mathrm{Log}$ such that $(1 + \eta)^{6n+2} \le 1 + \varepsilon/4$. We denote

$$Z_{a..b} = \{\langle x, y\rangle \in Z;\ a \le x < b\}$$

for every $a < b \le 2^n$, and $Z_a = Z_{a..a+1}$. By Lemma 3.14, Corollary 3.5 and Theorem 3.8, there exist $\Pi^b_1(\alpha)$-predicates $C(u, v, a), D(u, v, a)$ such that

$$\begin{aligned} C(u, v, a) \to \qquad\quad Z_{u..v} \precsim_\eta a &\to C(u, v, \lfloor a(1 + \eta)\rfloor), \\ D(u, v, a) \to X \cap [u, v) \precsim_\eta a &\to D(u, v, \lfloor a(1 + \eta)\rfloor). \end{aligned}$$

We prove

$$(*) \qquad \forall u < v \le 2^n\, \forall a \le 2^n \left(v - u = 2^k \wedge D(u, v, a) \to C\big(u, v, \lfloor at(1 + \varepsilon)(1 + \eta)^{6k+1}\rfloor\big)\right)$$

by induction on $k \le n$. The case $k = 0$ is clear, let thus $k > 0$. Assume $D(u, v, a)$. Put $w = (u + v)/2$, and find $b, c \le 2^n$ such that $D(u, w, b)$, $\neg D(u, w, b - 1)$, $D(w, v, c)$, $\neg D(w, v, c - 1)$ using induction (where "$D(\dots, -1)$" counts as false). By $(*)$ for $k - 1$, we have

$$\begin{aligned} Z_{u..w} &\precsim_\eta \lfloor bt(1 + \varepsilon)(1 + \eta)^{6k-5}\rfloor, \\ Z_{w..v} &\precsim_\eta \lfloor ct(1 + \varepsilon)(1 + \eta)^{6k-5}\rfloor, \end{aligned}$$

hence

$$Z_{u..v} \precsim_\eta \lfloor (b + c)t(1 + \varepsilon)(1 + \eta)^{6k-3}\rfloor$$

by Theorem 3.13. On the other hand, we have $X \cap [u, w) \not\precsim_\eta \lceil b(1 + \eta)^{-1}\rceil - 1$, $X \cap [w, v) \not\precsim_\eta \lceil c(1 + \eta)^{-1}\rceil - 1$, thus

$$X \cap [u, v) \not\precsim_\eta \left\lceil \frac{b}{(1 + \eta)^3}\right\rceil + \left\lceil \frac{c}{(1 + \eta)^3}\right\rceil - 1$$

by Theorem 3.17. As $X \cap [u, v) \precsim_\eta a$, we obtain $a \ge (b + c)(1 + \eta)^{-3}$, i.e., $b + c \le \lfloor a(1 + \eta)^3\rfloor$. Hence

$$Z_{u..v} \precsim_\eta \lfloor at(1 + \varepsilon)(1 + \eta)^{6k}\rfloor,$$

which implies $C\big(u, v, \lfloor a(1 + \varepsilon)(1 + \eta)^{6k+1}\rfloor\big)$.

Take $k = n$. We have $D(0, 2^n, \lfloor s(1 + \varepsilon)(1 + \eta)\rfloor)$, hence $C\big(0, 2^n, \lfloor st(1 + \varepsilon)^2(1 + \eta)^{6n+2}\rfloor\big)$ by $(*)$, which gives

$$Z = Z_{0..2^n} \precsim_\eta \lfloor st(1 + \varepsilon)^2(1 + \eta)^{6n+2}\rfloor \le \lfloor st(1 + 4\varepsilon)\rfloor.$$

As it stands, the proof needs $\Pi^b_2(\alpha)$-*LIND*. The theorem is unfortunately not $\forall\Sigma^b_2(\alpha)$, we thus cannot directly use Theorem 2.1. Nevertheless, we can decrease the complexity of the induction as follows. Let $a(u,v)$ be a $PV_2(\alpha)$-function which computes $a \le 2^n$ such that $D(u,v,a) \land \neg D(u,v,a-1)$ by binary search. We define a $PV_2(\alpha)$-function $f(k)$ (where $k \le n$ is given in unary) by

$$\begin{aligned} f(0) &= 0, \\ f(k+1) &= \begin{cases} f(k) + 2^{n-k-1}, & C\bigl(u, u+2^{n-k-1}, \lfloor a(u,u+2^{n-k-1})t(1+\varepsilon)(1+\eta)^{6(n-k)-5}\rfloor\bigr), \\ f(k) & \text{otherwise.} \end{cases} \end{aligned}$$

If we assume for contradiction $Z \not\precsim_\eta \lfloor st(1+\varepsilon)^2(1+\eta)^{6n+2}\rfloor$, we can prove

$$\neg C\bigl(f(k), f(k)+2^{n-k}, \lfloor a(f(k),f(k)+2^{n-k})t(1+\varepsilon)(1+\eta)^{6(n-k)+1}\rfloor\bigr)$$

by $PV_2(\alpha)$-*LIND* on $k \le n$, using the same argument as above. Taking $k = n$, we have $a(f(k), f(k)+1) \le 1$, and we obtain a contradiction with the assumptions. □

Theorem 3.20 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) If $X, Y, Z \in \Sigma^b_1(\alpha)$, $Z \subseteq X \times Y$, and $Z \precsim_\varepsilon st$, then $X \precsim s-1$, or there exists $x \in X$ such that $\{y \in Y;\ \langle x,y\rangle \in Z\} \precsim \lfloor t(1+2\varepsilon)\rfloor$.*

Proof: Let $n \in \mathrm{Log}$ be such that $X \subseteq 2^n$. Fix $\eta \in \mathrm{Log}$ such that $(1+\eta)^{6n} \le (1+\varepsilon/2)$, and assume $X \not\precsim_\eta s-1$. By induction on $k \le n$, we will show

$$(*)\quad \exists u < v \le 2^n\, \exists a \le 2^n \Bigl(v-u \le 2^{n-k} \land a \ne 0 \land (X \cap [u,v)) \not\precsim_\eta a-1 \\ \land Z_{u..v} \precsim_\eta \lfloor at(1+\varepsilon)(1+\eta)^{6k}\rfloor\Bigr),$$

where $Z_{u..v}$ is as in the proof of Theorem 3.19. If $k=0$, we may take $u=0$, $v=2^n$, $a=s$. Assume that $(*)$ holds for $k<n$. Put $w = \lceil (u+v)/2\rceil$, and find b,c such that $X \cap [u,w) \not\precsim_\eta b-1$, $X \cap [u,w) \precsim_\eta \lfloor b(1+\eta)\rfloor$, $X \cap [w,v) \not\precsim_\eta c-1$, $X \cap [w,v) \precsim_\eta \lfloor c(1+\eta)\rfloor$. Assume $b \ne 0 \ne c$, the other cases are easy. We have $X \cap [u,v) \precsim_\eta \lfloor (b+c)(1+\eta)^3\rfloor$ by Theorem 3.13, hence $a \le \lfloor (b+c)(1+\eta)^3\rfloor$, which implies $Z_{u..v} \precsim_\eta \lfloor bt(1+\varepsilon)(1+\eta)^{6k+4}\rfloor + \lfloor ct(1+\varepsilon)(1+\eta)^{6k+4}\rfloor + 1$. By Theorem 3.17, we obtain $Z_{u..w} \precsim_\eta \lfloor bt(1+\eta)^{6(k+1)}\rfloor$ or $Z_{w..v} \precsim_\eta \lfloor ct(1+\eta)^{6(k+1)}\rfloor$, which gives $(*)$ for $k+1$.

Take u,v,a which witness $(*)$ for $k=n$. Then $v-u \le 1$, and $X \cap [u,v) \ne \varnothing$, hence $v = u+1$, $u \in X$, $a=1$, and $Z_u \precsim_\eta \lfloor t(1+\varepsilon)(1+\eta)^{6n}\rfloor$, which implies $Z_u \precsim \lfloor t(1+2\varepsilon)\rfloor$.

As in the proof of Theorem 3.19, we can replace $\precsim_\eta$ by a $\Pi^b_1(\alpha)$-formula in $(*)$, thus the argument formalizes in $S^2_2(\alpha) + sWPHP(PV_2(\alpha))$. The result is $\forall\Sigma^b_2(\alpha)$, hence it is provable in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$ by Theorem 2.1. (We can also eliminate the instance of $\Sigma^b_2(\alpha)$-*LIND* explicitly as in Theorem 3.19.) □

In the special case $Z = X \times Y$, Theorem 3.20 implies a variant of Theorem 3.16 with slightly different parameters, which may be favourable in some applications (e.g., see the proof of Theorem 4.3): if $X \times Y \precsim_\varepsilon st$, then $X \precsim s-1$ or $Y \precsim \lfloor t(1+2\varepsilon)\rfloor$.

The next theorem shows that we can construct almost counting functions for any set X. Moreover, the conditions imposed on f and g make them very well-behaved: the "local defects” by which the functions differ from true counting functions (i.e., monotone bijections) are small, and evenly distributed across the domain. A possible use of the theorem is that we can apply various results of [21] to relatively dense subsets of a sparse set X, as we can lift the whole situation to an interval. (Lifting a $\Sigma^b_1(\alpha)$-set by a $PV_2(\alpha)$-function increases its complexity to $\Delta^b_2(\alpha)$, which is fine as $T^1_2(\alpha) + sWPHP(PV_2(\alpha))$ can count $\Delta^b_2(\alpha)$-sets in the framework of [21].)

The main idea of the construction was suggested by Neil Thapen.

Theorem 3.21 *(in $T^1_2(\alpha) + rWPHP(PV_2(\alpha))$) Let $X \in \Sigma^b_1(\alpha)$, and $\varepsilon^{-1} \in \mathrm{Log}$. There exist numbers t, s such that $s \le t \le \lfloor s(1+\varepsilon)\rfloor$, and non-decreasing $PV_2(\alpha)$-functions*

$$\begin{tikzcd} t \arrow[r, shift left, "f"] & X \arrow[l, shift left, "f'"] \arrow[r, shift left, "g"] & s \arrow[l, shift left, "g'"] \end{tikzcd}$$

such that $f \circ f' = \mathrm{id}_X$, $g \circ g' = \mathrm{id}_s$ (hence f, g are onto, and f', g' are injective), f, g are ≤ 2-to-1, and

$$\left\lfloor \frac{s}{t} u \right\rfloor \le g(f(u)) \le \left\lceil \frac{s}{t} u \right\rceil$$

for every $u < t$.

Proof: Fix $n \in \mathrm{Log}$ such that $X \subseteq 2^n$, and $\eta^{-1} \in \mathrm{Log}$ such that $(1+\eta)^{8n} \le 1 + \varepsilon$. Let C be a $\Pi^b_1(\alpha)$-predicate such that

$$C(u, v, w) \to X \cap [u, v) \precsim_\eta w \to C(u, v, \lfloor w(1+\eta)\rfloor)$$

for all $u, v, w \le 2^n$. Using binary search, we can define a $PV_2(\alpha)$-function S such that

$$C(u, v, S(u, v)) \wedge \neg C(u, v, S(u, v) - 1)$$

for all $u, v \le 2^n$. Put $a = S(0, 2^n)$. If $a < \eta^{-1}$, then $a \in \mathrm{Log}$, hence the required functions exist by Corollary 3.10. We thus assume $a \ge \eta^{-1}$. Consider the following algorithm (where u, u_k, v_k, w_k are rationals):

```
input: either $u \in [0, 1)$, or $x \in X$
let $x_0 := 0$, $y_0 := 2^n$, $u_0 := 0$, $v_0 := 1$, $r_0 := a$
for $k = 0, \dots, n-1$ do:
  $z_k := (x_k + y_k)/2$, $p_k := S(x_k, z_k)$, $q_k := S(z_k, y_k)$
  $w_k := (q_k u_k + p_k v_k)/(p_k + q_k)$
  if $u < w_k$ or $x < z_k$ then $\langle x_{k+1}, y_{k+1}, u_{k+1}, v_{k+1}, r_{k+1}\rangle := \langle x_k, z_k, u_k, w_k, p_k\rangle$
        else $\langle x_{k+1}, y_{k+1}, u_{k+1}, v_{k+1}, r_{k+1}\rangle := \langle z_k, y_k, w_k, v_k, q_k\rangle$
```

If it is necessary to indicate the input, we will write $p_k(u)$ for the value of p_k assigned by the algorithm on input u, and so on.

Claim 1 *Let $u \le u' < 1$, $x \le x' \in X$, and $\ell \le k \le n$.*

(i) $y_k = x_k + 2^{n-k}$.

(ii) $u_k(u) \le u < v_k(u)$, $x_k(x) \le x < y_k(x)$.

(iii) $r_k = S(x_k, y_k) \neq 0$*, and* $p_k + q_k \neq 0$ *(hence the division step makes sense).*

(iv) $x_\ell \leq x_k$*,* $y_\ell \geq y_k$*,* $u_\ell \leq u_k$*,* $v_\ell \geq v_k$*.*

(v) *Either* $\langle x_k(u), y_k(u), u_k(u), v_k(u)\rangle = \langle x_k(u'), y_k(u'), u_k(u'), v_k(u')\rangle$*, or* $y_k(u) \leq x_k(u')$*,* $v_k(u) \leq u_k(u')$*.*

(vi) *Either* $\langle x_k(x), y_k(x), u_k(x), v_k(x)\rangle = \langle x_k(x'), y_k(x'), u_k(x'), v_k(x')\rangle$*, or* $y_k(x) \leq x_k(x')$*,* $v_k(x) \leq u_k(x')$*.*

(vii) *If* $x_k(u) \leq x < y_k(u)$*, or* $u_k(x) \leq u < v_k(x)$*, then* $\langle x_k(u), y_k(u), u_k(u), v_k(u)\rangle = \langle x_k(x), y_k(x), u_k(x), v_k(x)\rangle$*.*

(viii) $(1+\eta)^{-3} r_k \leq p_k + q_k \leq (1+\eta)^3 r_k$*.*

(ix) $(1+\eta)^{-3k} r_k \leq a(v_k - u_k) \leq (1+\eta)^{3k} r_k$*.*

Proof: (i)–(vii): Straightforward induction on k.

(viii): As $p_k = S(x_k, z_k)$, we have $\neg C(x_k, z_k, p_k - 1)$, thus $X \cap [x_k, z_k) \not\precsim_\eta \lceil p_k(1+\eta)^{-1}\rceil - 1$. Similarly $X \cap [z_k, y_k) \not\precsim_\eta \lceil q_k(1+\eta)^{-1}\rceil - 1$, hence

$$\begin{aligned} X \cap [x_k, y_k) \not\precsim_\eta \bigl\lceil \lceil p_k(1+\eta)^{-1}\rceil (1+\eta)^{-2}\bigr\rceil + \bigl\lceil \lceil q_k(1+\eta)^{-1}\rceil (1+\eta)^{-2}\bigr\rceil - 1 \\ \geq \lceil (p_k + q_k)(1+\eta)^{-3}\rceil - 1 \end{aligned}$$

by Theorem 3.17. On the other hand, $r_k = S(x_k, y_k)$, hence $C(x_k, y_k, r_k)$, and $X \cap [x_k, y_k) \precsim_\eta r_k$. This implies $r_k \geq \lceil (p_k + q_k)(1+\eta)^{-3}\rceil$, hence $r_k(1+\eta)^3 \geq p_k + q_k$. In a similar way we have $X \cap [x_k, z_k) \precsim_\eta p_k$, $X \cap [z_k, y_k) \precsim_\eta q_k$, and $X \cap [x_k, y_k) \not\precsim_\eta \lceil r_k(1+\eta)^{-1}\rceil - 1$, hence $\lceil r_k(1+\eta)^{-1}\rceil \leq \lfloor (p_k + q_k)(1+\eta)^2\rfloor$ by Theorem 3.13, thus $(1+\eta)^{-3} r_k \leq (p_k + q_k)$.

(ix): By induction on k, using (viii), and the identities

$$w_k - u_k = \frac{p_k}{p_k + q_k}(v_k - u_k), \qquad v_k - w_k = \frac{q_k}{p_k + q_k}(v_k - u_k). \qquad \square \text{ (Claim 1)}$$

Let $t = \lceil a(1+\eta)^{3n}\rceil$, $s = \lfloor a(1+\eta)^{-3n}\rfloor$, $f(u) = x_n(u/t)$, $f'(x) = \lceil t v_n(x)\rceil - 1$, $g(x) = \lceil s v_n(x)\rceil - 1$, $g'(v) = x_n(v/s)$ for any integers $u < t$, $v < s$, $x \in X$. We have $t \leq a(1+\eta)^{4n} \leq \lfloor a(1+\eta)^{-3n}\rfloor (1+\eta)^{8n} \leq s(1+\varepsilon)$. Notice that $r_n = 1$ (hence $x_n \in X$) by (i) and (iii), thus

$$\frac{1}{2s} \leq \frac{1}{t} \leq v_n - u_n \leq \frac{1}{s} \leq \frac{2}{t}$$

by (ix) (in particular, $f'(x), g(x) \geq 0$). Clearly $f\colon t \to X$, $f'\colon X \to t$, $g\colon X \to s$, $g'\colon s \to X$, and all the functions are monotone by (v), (vi).

If $u = f'(x)$, we have $u_n(x) \leq v_n(x) - 1/t \leq u/t < v_n(x)$, hence $f(u) = x_n(u/t) = x_n(x) = x$ by (vii), thus $f \circ f' = \mathrm{id}$. If $f(u) = x$, then $t u_n(x) = t u_n(u/t) \leq u < t v_n(u/t) = t v_n(x)$ by (vii). As there are at most two integers in $[t u_n(x), t v_n(x))$, we have $|f^{-1}(x)| \leq 2$.

If $x = g'(v)$, then $s v_n(x) - 1 \leq s u_n(x) = s u_n(v/s) \leq v < s v_n(v/s) = s v_n(x)$ by (vii), hence $v = g(x)$, thus $g \circ g' = \mathrm{id}$. Assume that $x, x', x'' \in X$, $x < x' < x''$. We have

$y_n(x) = x + 1 \leq x_n(x') = x'$ by (i) and (ii), hence $sv_n(x) \leq su_n(x') \leq sv_n(x') - 1/2$ by (vi). Similarly $sv_n(x') \leq sv_n(x'') - 1/2$, hence $g(x'') = \lceil sv_n(x'') \rceil - 1 \geq \lceil sv_n(x) \rceil > g(x)$. It follows that $|g^{-1}(v)| \leq 2$ for any $v < s$.

Let $u < t$, and put $x = f(u)$, $v = g(x)$. We have $v_n(x) - 1/s \leq u_n(x) = u_n(u/t) \leq u/t < v_n(u/t) = v_n(x)$ by (vii) and (ii), and $v_n(x) - 1/s \leq v/s < v_n(x)$ by definition, hence $-1/s < u/t - v/s < 1/s$. □

4 Applications

We begin with a classical theorem which cannot be avoided by any self-respecting theory of counting.

Theorem 4.1 (Ramsey theorem) *(in $T^1_2(G) + rWPHP(PV_2(G))$) An undirected graph G on N vertices contains a clique or independent set of size at least $|N|/2$.*

Proof: We formalize the following well-known proof. We pick a node a_0, and let c_0 be the majority colour among edges incident with a_0. We continue with nodes connected to a_0 by a c_0-coloured edge, and repeat the process. In this way, we construct a sequence $a_0, \dots, a_{k-1}$ of nodes and a sequence $c_0, \dots, c_{k-1}$ of colours such that the edge from a_i to a_j is c_i-coloured for $i < j$, and there are at least (roughly) $N/2^k$ nodes connected to every a_i by a c_i-coloured edge. We can carry on as long as $k < \log_2 N$, obtaining a prehomogeneous set of size $\log_2 N$, from which we select a homogeneous set of size $\log_2 N/2$ by taking the majority colour among $\vec{c}$. We proceed with the formal details.

We can use $sWPHP$ by Theorem 2.1. For every $a \neq b$, we define $C(a, b) < 2$ so that $C(a, b) = 1$ iff there is an edge between a and b in G. Let $\varepsilon^{-1} \in \mathrm{Log}$ be such that $(1-2\varepsilon)^{|N|} \geq 1/2$. By induction on $k \leq |N| - 2$, we prove that there exists a sequence $\langle c_i;\, i < k \rangle$ of $c_i < 2$, and a sequence $\langle a_i;\, i < k \rangle$ of pairwise distinct $a_i < N$ such that $C(a_i, a_j) = c_i$ whenever $i < j < k$, and

$$S(\vec{a};\vec{c}) := \{x;\, \forall i < k\, C(a_i, x) = c_i\} \not\precsim_\varepsilon \left\lfloor \frac{N}{2^k}(1-2\varepsilon)^{k+1} \right\rfloor - 1.$$

(We can make the induction hypothesis $\Sigma^b_1(G)$ by Lemma 3.14, as in the proof of Theorem 3.19.) The base step $k = 0$ amounts to $N \not\precsim_\varepsilon \lfloor N(1-2\varepsilon) \rfloor - 1$, which follows from Theorem 3.8 and $rWPHP$. Assume the statement holds for k, we will show it for $k+1$. We have $S(\vec{a};\vec{c}) \neq \varnothing$, we may thus pick any $a_k \in S(\vec{a};\vec{c})$. The set $S(\vec{a};\vec{c})$ can be divided into nodes x such that $C(a_k, x) = 0$, nodes such that $C(a_k, x) = 1$, and node a_k itself, hence

$$S(\vec{a};\vec{c}) = \{a_k\} \cup S(\vec{a}, a_k; \vec{c}, 0) \cup S(\vec{a}, a_k; \vec{c}, 1).$$

We have $1 \precsim_\varepsilon 1$, and

$$\Big\lfloor (1+2\varepsilon)\Big(1 + 2\big(\lfloor N2^{-(k+1)}(1-2\varepsilon)^{k+2} \rfloor - 1\big)\Big) \Big\rfloor \leq$$
$$\leq \Big\lfloor 2(1+2\varepsilon)N2^{-(k+1)}(1-2\varepsilon)^{k+2} - (1+2\varepsilon) \Big\rfloor \leq \lfloor N2^{-k}(1-2\varepsilon)^{k+1} - 1 \rfloor,$$

hence

$$S(\vec{a}, a_k; \vec{c}, c_k) \not\gtrsim_\varepsilon \left\lfloor \frac{N}{2^{k+1}}(1-2\varepsilon)^{k+2} \right\rfloor - 1$$

for some $c_k < 2$ by Theorem 3.13.

Let $\vec{a}$, $\vec{c}$ be the sequences given by the statement above for $k = |N|-2$. We have $S(\vec{a}; \vec{c}) \not\gtrsim_\varepsilon 0$, hence there exists $a_k \in S(\vec{a}; \vec{c})$. There exists $c < 2$ such that $|\{i < k;\, c_i = c\}| \geq \lceil k/2 \rceil$, then $\{a_i;\, c_i = c\} \cup \{a_k\}$ is a homogeneous set of size $\lceil |N|/2 \rceil$. □

The Ramsey theorem was, of course, proved in bounded arithmetic by Pudlák [30]. The point of Theorem 4.1 is that (apart from a few ε sprinkled here and there) the argument follows almost literally the usual combinatorial proof of the theorem, without the need to resort to *ad hoc* functions for simulation of counting by *WPHP*.

Our first real result will be the tournament principle (originally discovered by Erdős [13]), whose provability in bounded arithmetic was posed as a problem by Krajíček [9, 23]. Recall that a *tournament* is a directed graph G in which there exists exactly one directed edge between any pair of distinct vertices ("players"); if there is an edge going from a to b, we write $a \to b$, and say that a beats b. A *dominating set* is a set D of vertices such that every player outside of D is beaten by some player in D.

Theorem 4.2 (Tournament principle) *(in $T^1_2(G) + rWPHP(PV_2(G))$) A tournament G with N players has a dominating set of size at most $|N|$.*

Proof: Informally, the argument is as follows. There are $N(N-1)/2$ edges in the tournament, hence we may choose a player a_0 who beats at least $(N-1)/2$ other players. We repeat the process with the subtournament consisting of the unbeaten players, halving the size at each step. After at most $|N|$ steps, we reach the empty set, hence we obtaining a dominating set of size $|N|$. We now give the formal proof.

We can work in $S^2_2(G) + sWPHP(PV_2(G))$ by Theorem 2.1. Choose $\varepsilon^{-1} \in \mathrm{Log}$ such that $(1+\varepsilon)^{8(|N|+1)} < 2$. If $\langle a_i;\, i < k\rangle$ is a sequence of vertices, we denote

$$G(\vec{a}) = \{x < N;\, \forall i < k\, x \to a_i\}.$$

By $\Sigma^b_2(G)$-*LIND* on $k \leq |N|+1$, we will prove that there exists a sequence $\langle a_i;\, i < k\rangle$ such that

$$(*) \qquad G(\vec{a}) \precsim_\varepsilon \left\lfloor \frac{N}{2^k}(1+\varepsilon)^{8k} \right\rfloor .$$

The case $k = |N|+1$ then gives $G(\vec{a}) = \varnothing$, i.e., $\vec{a}$ is a dominating set of size $|N|+1$. (How to get rid of the $+1$ is left as an exercise. Hint: in real world, the bound $|N|$ is not tight.)

The base case $k = 0$ is obvious. Assume that $(*)$ holds for k, we will show it for $k+1$. Find s such that $G(\vec{a}) \precsim_\varepsilon \lfloor s(1+\varepsilon)\rfloor$, $G(\vec{a}) \not\gtrsim_\varepsilon s-1$. Notice that $s \leq N2^{-k}(1+\varepsilon)^{8k}$. We have

$$\{\langle x, y\rangle \in G(\vec{a});\, x \neq y\} \subseteq G(\vec{a})^2 \precsim_\varepsilon \lfloor s^2(1+\varepsilon)^4 \rfloor \leq 2\left\lfloor \frac{s^2}{2}(1+\varepsilon)^4 \right\rfloor + 1$$

by Corollary 3.12, hence

$$\{\langle x,y\rangle \in G(\vec{a})^2;\ y \to x\} \precsim_\varepsilon \left\lfloor \frac{s^2}{2}(1+\varepsilon)^6 \right\rfloor \quad \text{or} \quad \{\langle x,y\rangle \in G(\vec{a})^2;\ x \to y\} \precsim_\varepsilon \left\lfloor \frac{s^2}{2}(1+\varepsilon)^6 \right\rfloor$$

by Theorem 3.17, and properties of the tournament. In the former case, there exists an $x \in G(\vec{a})$ such that

$$G(\vec{a},x) = \{y \in G(\vec{a});\ y \to x\} \precsim_\varepsilon \left\lfloor \frac{s}{2}(1+\varepsilon)^8 \right\rfloor \le \left\lfloor \frac{N}{2^{k+1}}(1+\varepsilon)^{8(k+1)} \right\rfloor$$

by Theorem 3.20. The latter case is symmetric. □

As proved by E. and G. Szekeres [35], every tournament has a dominating set of size $|N| - ||N|| + O(1)$. We could formalize this stronger result with no additional difficulty; we skip the proof as it involves lengthy quotes from [35] with no particular benefit for our purpose (which is to illustrate the machinery developed in Section 3).

For the sake of completeness, we mention that Erdős [13] proved a lower bound of $|N| - 2||N|| + O(1)$ on the minimal size of a dominating set in random tournaments, and Razborov [31] provided tournaments computable by $AC^0[2]$-circuits with the same property. We do not know how to prove these lower bounds in bounded arithmetic. (Ojakian [28] formalizes Erdős's proof in a different setting, where $N \in \mathrm{Log}$.) An explicit construction of tournaments without small dominating sets was given in [16]: if $p \equiv -1 \pmod 4$ is a prime, the tournament with p players defined by

$$a \to b \quad \text{iff} \quad \left(\frac{a-b}{p}\right) = 1$$

has no dominating set of size $\frac{1}{2}|p| - ||p||$. However, their proof depends on Weil's Riemann hypothesis for curves over finite fields, which we cannot expect to prove in bounded arithmetic by any stretch of imagination.

It turns out that generalizations of the tournament principle are more useful in applications than the principle itself. We provide such a generalization next. The statement seems to be new even outside the context of bounded arithmetic; it was inspired by a variant of the tournament principle introduced in [14] (our Corollary 4.4), and a combinatorial principle implicit in [26] (Corollary 4.5).

In order to explain it, let us consider first Corollary 4.4, which is a symmetric generalization of the tournament principle to arbitrary binary relations that may not be tournaments. We can reformulate it as follows: given a colouring of ordered pairs of points of a by two colours, there is a colour $i < 2$, and a set D of size $\log a$ with the following property: for any point x, there is an i-coloured pair whose ith coordinate is x, and the other coordinate belongs to D. Now we can generalize the statement to higher dimensions as follows (this is the special case of Theorem 4.3 with $a_i = a$, $p_i = 1/d$, $m_i = 1$): given a colouring of d-tuples of points of a by d colours, there exists a colour $i < d$, and a set D of $(d-1)$-tuples of size $(d-1)\log a$ with the following property: for any point x, there exists an i-coloured d-tuple whose ith coordinate is x, and the tuple consisting of the remaining coordinates belongs to D.

In order to accommodate Corollary 4.5, we introduce as an extra complication the possibility that the colouring is not total. We only require that it is "dense", in the sense that every hypercube with sufficiently large sides (sets of size m) contains a tuple whose colour is defined. The conclusion is modified so that the d-tuple only needs to be i-coloured if its colour is defined, and there will be an exceptional small (of size less than m) set M whose points $x \in M$ are exempt from the existence condition. To guard against trivializing the conclusion, we also require that any tuple from D can be extended to a d-tuple with defined colour ($D \subseteq S_i$ in the notation below). Finally, we allow each coordinate to use a different value of a and m for extra generality, as it does not change the proof, and indeed it simplifies the notation used in the proof in that it allows us to conveniently specify which coordinate in the product a^d are we referring to.

Theorem 4.3 *(in $T^1_2(C)$ + $rWPHP(PV_2(C))$) Let $0 < d \in \mathrm{Log}$. Let $\langle a_i; i < d\rangle$ and $\langle m_i; i < d\rangle$ be sequences of positive integers such that $m_i \in \mathrm{Log}$, $\langle p_i; i < d\rangle$ a sequence of rationals $p_i \in \mathbb{Q}_{\mathrm{Log}}$ such that $0 < p_i < 1$, and C a partial function from $\prod_{i<d} a_i$ to d.*

Assume that $\sum_{i<d} p_i \le 1$, and $\mathrm{dom}(C) \cap \prod_{i<d} M_i \neq \varnothing$ for every sequence $\langle M_i; i < d\rangle$ of subsets $M_i \subseteq a_i$ such that $|M_i| = m_i$. Put

$$S_i = \Big\{\langle x_j; j \neq i\rangle \in \prod_{j\neq i} a_j;\ \exists x_i \in a_i\, \vec{x} \in \mathrm{dom}(C)\Big\}.$$

Then there exists an $i < d$, a set $D \subseteq S_i$ of size at most

$$2 + \lfloor \log_{(1-p_i)^{-1}}(a_i/m_i)\rfloor \le 1 + (p_i^{-1} - 1)|\lfloor a_i/m_i\rfloor|,$$

and a set $M \subseteq a_i$ of size $|M| < m_i$ with the following property: for every $x_i \in a_i \smallsetminus M$ there exists $\langle x_j; j \neq i\rangle \in D$ such that $C(\vec{x}) = i$ or $\vec{x} \notin \mathrm{dom}(C)$.

Proof: The statement is $\forall\Sigma^b_2(C)$, we can thus work in $S^2_2(C)$ + $sWPHP(PV_2(C))$. We write $C(\vec{x})\uparrow$ for $\vec{x} \notin \mathrm{dom}(C)$. If $\vec{x} \in \prod_{j\neq i} a_j$, and $x \in a_i$, we will write $C(\vec{x}, x)$ instead of $C(x_0, \dots, x_{i-1}, x, x_{i+1}, \dots, x_{d-1})$ if i is clear from the context.

For each $i < d$, put $c_i = 2 + \lfloor \log_{(1-p_i)^{-1}}(a_i/m_i)\rfloor$. As $c_i \in \mathrm{Log}$, and $a_i(1-p_i)^{c_i-1} < m_i$, we can construct $\delta_i \in \mathbb{Q}_{\mathrm{Log}}$ such that $0 < \delta_i < p_i$, and $a_i(1-\delta_i)^{c_i} < m_i$. Then there exists an $0 < \varepsilon \in \mathbb{Q}_{\mathrm{Log}}$ such that $\big(1 - p_i(1+\varepsilon)^{-19}\big)(1+\varepsilon)^3 \le 1 - \delta_i$ for every $i < d$.

By $\Sigma^b_2(C)$-$LMAX$, we can find the maximal k such that there exist sequences $\langle k_i; i < d\rangle$, $\langle \vec{x}^{i,j}; i < d, j < k_i\rangle$ satisfying $k = \sum_i k_i$, $k_i \le c_i$, $\vec{x}^{i,j} \in S_i$, and

$$M_i := \{x < a_i;\ \forall j < k_i\, \exists \ell \neq i\, C(\vec{x}^{i,j}, x) = \ell\} \gtrsim_\varepsilon \lfloor a_i(1-\delta_i)^{k_i}\rfloor$$

for every $i < d$. If $|M_i| < m_i$ for some $i < d$, the conclusion of the theorem holds with $M = M_i$, $D = \{\vec{x}^{i,j}; j < k_i\}$. We thus assume $|M_i| \ge m_i$ (which implies $k_i < c_i$ by the choice of δ_i) for every $i < d$, and we intend to reach a contradiction. We put $X = \prod_j M_j \cap \mathrm{dom}(C)$, $X_i = \{\vec{x} \in X;\ C(\vec{x}) = i\}$, $N_i = S_i \cap \prod_{j\neq i} M_j$, and $O_i = (N_i \times M_i) \smallsetminus X$.

The intuition is as follows. For any i and $\vec{x} \in S_i$, we have

$$|\{x \in M_i;\ \exists \ell \neq i\, C(\vec{x}, x) = \ell\}| \ge a_i(1-\delta_i)^{k_i+1} \ge |M_i|(1-\delta_i)$$

by maximality of k_i, hence $\Pr_{x\in M_i}(C(\vec{x}, x)\uparrow \vee C(\vec{x}, x) = i) \le \delta_i$. Consequently,

$$1 = \sum_i \Pr_{\vec{x}\in X}(C(\vec{x}) = i) \le \sum_i \Pr_{\vec{x}\in N_i\times M_i}(C(\vec{x})\uparrow \vee C(\vec{x}) = i) \le \sum_i \delta_i < 1$$

using nonemptiness of X, which is a contradiction. Now we formalize this argument using approximate counting.

By assumption, $X \neq \varnothing$, hence we can find a $t > 0$ such that $X \precsim_\varepsilon \lfloor t(1+\varepsilon)\rfloor$, $X \not\succsim_\varepsilon t-1$ by $\Sigma^b_2(C)$-*LIND*. Fix $i < d$, and let w_i be such that $M_i \precsim_\varepsilon \lfloor w_i(1+\varepsilon)\rfloor$, and $M_i \not\succsim_\varepsilon w_i - 1$. Consequently, $w_i \leq \lfloor a_i(1-\delta_i)^{k_i}\rfloor$.

Take any $\vec{x} \in S_i$. By Theorem 3.17, we have

$$(*) \qquad \{x \in M_i;\, C(\vec{x},x)\!\uparrow \vee\, C(\vec{x},x) = i\} \precsim_\varepsilon \big\lfloor \lfloor w_i p_i (1+\varepsilon)^{-18}\rfloor (1+\varepsilon)^2 \big\rfloor \leq \lfloor w_i p_i (1+\varepsilon)^{-16}\rfloor$$

or

$$\begin{aligned}\{x \in M_i;\, \exists \ell \neq i\, C(\vec{x},x) = \ell\} \precsim_\varepsilon{}& \big\lfloor \big(\lfloor w_i(1+\varepsilon)\rfloor - \lfloor w_i p_i(1+\varepsilon)^{-18}\rfloor - 1\big)\,(1+\varepsilon)^2\big\rfloor \\ &\leq \lfloor w_i\big(1 - p_i(1+\varepsilon)^{-19}\big)(1+\varepsilon)^3\rfloor \leq \lfloor a_i(1-\delta_i)^{k_i+1}\rfloor.\end{aligned}$$

The latter however contradicts the maximality of k_i, hence $(*)$ holds for every $\vec{x} \in S_i$. Find v_i such that $N_i \precsim_\varepsilon \lfloor v_i(1+\varepsilon)\rfloor$, $N_i \not\succsim_\varepsilon v_i - 1$. We have

$$(**) \qquad P_i := \{\vec{x} \in N_i \times M_i;\, C(\vec{x})\!\uparrow \vee\, C(\vec{x}) = i\} \precsim_\varepsilon \lfloor v_i w_i p_i (1+\varepsilon)^{-11}\rfloor$$

by Theorem 3.19, and $(*)$. Let $O_i \precsim_\varepsilon \lfloor u_i(1+\varepsilon)\rfloor$, $O_i \not\succsim_\varepsilon u_i - 1$. We claim

$$(***) \qquad v_i w_i \leq \lfloor (t+u_i)(1+\varepsilon)^6\rfloor.$$

Note that $N_i \times M_i \subseteq X \cup O_i \precsim_\varepsilon \lfloor (t+u_i)(1+\varepsilon)^3\rfloor$ by Theorem 3.13. Assume first $v_i \geq 2/\varepsilon$. We have $N_i \not\succsim_\varepsilon v_i - 1 \geq \big\lfloor \lfloor v_i(1+\varepsilon)^{-2}\rfloor(1+2\varepsilon)\big\rfloor$, hence

$$\lfloor (t+u_i)(1+\varepsilon)^3\rfloor > w_i \lfloor v_i(1+\varepsilon)^{-2}\rfloor \geq v_i w_i (1+\varepsilon)^{-3}$$

by Theorem 3.20. The case $w_i \geq 2/\varepsilon$ is symmetric. If $v_i, w_i \leq 2/\varepsilon$, then in particular $v_i, w_i \in \mathrm{Log}$, and we can derive $N_i \times M_i \not\succsim_\varepsilon \lceil v_i w_i (1+\varepsilon)^{-1}\rceil - 1$ easily by exact counting, which implies $(***)$ as above.

The definition of S_i implies $X \subseteq N_i \times M_i$, hence $P_i = O_i \,\dot\cup\, X_i$. We thus obtain

$$O_i \,\dot\cup\, X_i \precsim_\varepsilon \lfloor v_i w_i p_i (1+\varepsilon)^{-11}\rfloor \leq \lfloor (t+u_i) p_i (1+\varepsilon)^{-5}\rfloor \leq \lfloor t p_i (1+\varepsilon)^{-5}\rfloor + \lceil u_i p_i (1+\varepsilon)^{-5}\rceil$$

from $(**)$ and $(***)$, which implies

$$O_i \precsim_\varepsilon \big\lfloor \big(\lceil u_i p_i (1+\varepsilon)^{-5}\rceil - 1\big)\,(1+\varepsilon)^2\big\rfloor \leq \lceil u_i(1+\varepsilon)^{-3}\rceil - 1 \quad \text{or} \quad X_i \precsim_\varepsilon \lfloor t p_i (1+\varepsilon)^{-3}\rfloor$$

by Theorem 3.17. The former contradicts the choice of u_i, hence the latter holds for every $i < d$. As $t > 0$, we obtain

$$X = \bigcup_i X_i \precsim_\varepsilon \Big\lfloor (1+\varepsilon)^2 \sum_i t p_i (1+\varepsilon)^{-3}\Big\rfloor \leq \lfloor t(1+\varepsilon)^{-1}\rfloor \leq t-1$$

from Corollary 3.18, which contradicts the definition of t. □

Corollary 4.4 *(in $T^1_2(R) + rWPHP(PV_2(R))$) Let R be a binary relation on a. There exists a set $D \subseteq a$ of size at most $|a| + 1$ such that*

$$\forall x < a\, \exists y \in D\, R(x, y) \lor \forall y < a\, \exists x \in D\, \neg R(x, y).$$

Proof: Use Theorem 4.3 with $d = 2$, $a_i = a$, $p_i = 1/2$, $m_i = 1$, and C the (total) characteristic function of R. □

Corollary 4.5 *(in $T^1_2(R) + rWPHP(PV_2(R))$) Let $c \in \mathrm{Log}$, and let $a^{[i]}$ denote the set of i-element subsets of a. Assume that $R \subseteq a^{[c]} \times a$ is a relation satisfying*

$$\forall X \in a^{[c+1]}\, \exists x \in X\, R((X \smallsetminus \{x\}), x).$$

Then there exists a set $D \subseteq a^{[c]}$ of size $|D| \le c|a|$, and a set $M \subseteq a$ of size at most c, such that

$$\forall x \in a \smallsetminus \big(M \cup \bigcup D\big)\, \exists X \in D\, R(X, x).$$

Proof: Apply Theorem 4.3 with $d = m_i = c + 1$, $a_i = a$, $p_i = 1/d$, and

$$C(x_0, \ldots, x_c) = \begin{cases} \min\{i \le c;\ R(\{x_j;\ j \ne i\}, x_i)\} & \text{if } x_i \text{ are pairwise distinct,} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Observe that $\vec{x} \in S_i$ iff the elements of $\vec{x}$ are pairwise distinct. □

The collapse of the bounded arithmetic hierarchy implies the collapse of the polynomial-time hierarchy. The original result is by Krajíček et al. [26], who prove that $T^i_2 = S_2$ implies $\Sigma^P_{i+1} \subseteq \Delta^P_{i+1}/poly$ (hence also $PH = \Sigma^P_{i+2} = \Pi^P_{i+2}$). Buss [3] formalized a weaker conclusion inside the bounded arithmetic: if $T^i_2 = S^{i+1}_2$, then T^i_2 proves $\Sigma^b_{i+1} \subseteq \Pi^b_{i+1}/poly$, and $\Sigma^b_\infty = \mathcal{B}(\Sigma^b_{i+2})$ (cf. also Zambella [37]). We will show that the stronger collapse from [26] can be formalized in bounded arithmetic as well. Surprisingly, this also allows us to strengthen the collapse to $PH = \mathcal{B}(\Sigma^P_{i+1})$, using a result from [12].

Theorem 4.6 *If $T^i_2 = S^{i+1}_2$, then T^i_2 proves $\Sigma^b_{i+1} \subseteq \Delta^b_{i+1}/poly$.*

Proof: It suffices to formalize in T^i_2 the proofs of [26, Thm. B, L. 2.2], also repeated (with a slightly different notation) in [23, Thm. 10.2.4, L. 10.2.2]. We assume the reader has one of these two proofs at hand, but we sketch an outline of the proof here for convenience. We consider a Σ^b_{i+1}-predicate $\exists w \le v\, B(v, w)$, where $B \in \Pi^b_i$, we need to show that there is an $FP^{\Sigma^b_i}$-function $g(u, v)$ and a polynomially bounded advice function $h(n)$ such that

$$\exists w \le v\, B(v, w) \to B(v, g(h(|v|), v)).$$

We consider the Δ^b_{i+1}-relation

$$R(\langle v_1, \ldots, v_r\rangle, \langle w_1, \ldots, w_s\rangle) \Leftrightarrow s \le r \land \forall \ell \le s\, (w_\ell \le v_\ell \land B(v_\ell, w_\ell)).$$

By an application of the KPT witnessing theorem to an instance of Σ^b_{i+1}-$LMAX$, provable in T^i_2 by assumption, we obtain a combinatorial principle called Ω_i which states that we can compute a length-maximal b such that $R(a,b)$ from a by a certain counterexample computation in constantly many rounds. Using Ω_i, we define a certain algorithm for computing a pair $\langle \ell, w\rangle$ from $a = \langle v_1, \dots, v_k\rangle$. Let $V_1 = \{|v| = n; \exists w \leq v\, B(v,w)\}$. If Q is a $(k-1)$-element subset of V_1, and $v \in V_1 \smallsetminus Q$, we say that Q *helps* v [23] or $\langle Q, v\rangle$ *is good* [26], if there is an ordering $\{v_1, \dots, v_{\ell-1}, v_{\ell+1}, \dots, v_k\}$ of Q such that the algorithm assigns $\langle \ell, w\rangle$ to $\langle v_1, \dots, v_{\ell-1}, v, v_{\ell+1}, \dots, v_k\rangle$, where w is a witness for v (i.e., $w \leq v \wedge B(v,w)$). (Here, k is a constant parameter we obtain along with the principle Ω_i.) Using a counting argument, we constructs sets $V_1 \supseteq V_2 \supseteq \cdots \supseteq V_t$ and $Q_j \subseteq V_j$ for some $t = O(n)$ so that Q_j have $k-1$ elements, Q_j helps all elements of $V_j \smallsetminus V_{j+1}$, and $|V_t| \leq k$. Then we can compute a witness for v by a $FP^{\Sigma^b_i}$-function g given the sets $Q_1, \dots, Q_t, V_t$ as well as witnesses for all their elements, which will be encoded in the advice $h(n)$.

Now we turn to the formalization. Notice that the assumption $T^i_2 = S^{i+1}_2$ implies $T^i_2 = S_2$ by [3], hence we actually work in full bounded arithmetic; in particular, we can apply our results above to approximately count sets defined by arbitrary bounded formulas.

By inspection of the proof as given in [23] or [26], we see that T^i_2 proves the principle Ω_i (as the conclusion of the KPT witnessing theorem is provable, not just true), the analysis of the algorithm constructing $\langle \ell, w\rangle$ (straightforward, as the number of steps is a standard constant), as well as the final definition of the function g (obvious). The missing part is the construction the sets $Q_1, \dots, Q_{t-1}, V_t$ and the advice string $h(n)$. We close this gap by an application of Corollary 4.5, where $c = k-1$, and $R(Q,v)$ is the "Q helps v" relation. □

Corollary 4.7 *If $T^i_2 = S^{i+1}_2$, then T^i_2 proves $\Sigma^b_\infty = \mathcal{B}(\Sigma^b_{i+1})$, and $\Sigma^b_{i+1} \subseteq \Pi^b_{i+1}/O(1)$.*

Proof: Cook and Krajíček [12] show (in a two-sorted setting) that Theorem 4.6 implies Corollary 4.7 when $i = 0$. Their results relativize in a straightforward way. □

After showing that $PV_1 \vdash NP \subseteq P/poly$ implies $PV_1 \vdash PH = BH$ (where $BH = \mathcal{B}(NP)$ is the Boolean hierarchy), Cook and Krajíček [12] also asked whether the converse holds. We can answer their question affirmatively:

Corollary 4.8 *If T^i_2 proves $\Sigma^b_\infty = \mathcal{B}(\Sigma^b_{i+1})$, then T^i_2 proves $\Sigma^b_{i+1} \subseteq \Delta^b_{i+1}/poly$.*

Proof: The assumption implies $T^i_2 = S_2$ by Zambella [37], which gives the conclusion by Theorem 4.6. □

The base case $i = 0$ of Corollary 4.8 was meanwhile independently shown by Beyersdorff and Müller [1] using a direct proof.

Krajíček [24, 25] has studied connections between validity of variants of *PHP* in first-order structures M, and existence of certain types of abstract counting functions which map definable sets of M to elements of a ring (or semiring), and behave reasonably wrt embeddings, disjoint unions, and Cartesian products. In particular, a structure which admits a so-called nontrivial *approximate Euler characteristic* (see below) satisfies $iWPHP^{2n}_n$, and conversely, any structure which satisfies $iWPHP^{2n}_n$ and an additional principle (any two definable sets are comparable wrt definable embedding) admits a nontrivial approximate Euler characteristic.

Definition 4.9 If R is a partially ordered commutative ring, we write $a \mathrel{\dot{\le}} b$ if for every rational $q > 1$ there exist $k, l \in \mathbb{N}$ such that $l/k < q$ and $ka \le lb$. We also put $a \doteq b$ iff $a \mathrel{\dot{\le}} b \land b \mathrel{\dot{\le}} a$.

Let M be a first-order structure, and $\mathrm{Def}(M)$ the set of all subsets of M^k, $k \in \mathbb{N}$, definable with parameters from M. An *approximate Euler characteristic* is a function $\xi\colon \mathrm{Def}(M) \to R$, where R is a partially ordered commutative ring, such that

(i) $\xi(A) = |A|$ for finite A,

(ii) $\xi(A \mathbin{\dot\cup} B) \doteq \xi(A) + \xi(B)$,

(iii) $\xi(A \times B) \doteq \xi(A) \cdot \xi(B)$,

(iv) $\xi(A) \mathrel{\dot{\le}} \xi(B)$ if A is definably embeddable into B,

for all $A, B \in \mathrm{Def}(M)$. ξ is *trivial* if $R = 0$.

We also consider extra conditions

(v) $\xi(A) \mathrel{\dot{\le}} c\xi(B)$ if $\xi(f^{-1}[b]) \mathrel{\dot{\le}} c$ for all $b \in B$,

(vi) $c\xi(B) \mathrel{\dot{\le}} \xi(A)$ if $c \mathrel{\dot{\le}} \xi(f^{-1}[b])$ for all $b \in B$,

where $c \in R$, and $f\colon A \to B$ is a definable injection.

Let M be a model of bounded arithmetic formulated in a purely relational language (i.e., we replace functions with their graphs), and consider an interval $[0, a]_M$ as its substructure. Then definable sets in $[0, a]$ are definable in M by a bounded formula, hence $[0, a]$ satisfies $iWPHP^{2n}_n$ if $M \vDash iWPHP^{2n}_n(\Sigma^b_\infty)$. On the other hand, it is not known to satisfy the principle of comparing cardinalities (and it seems rather unlikely to hold in general). Nevertheless, we can show the following.

Theorem 4.10 *Let M be a model of $S_2(\alpha)$, and $a \in M$. Then $[0, a]_M$ with the induced structure admits a* (*totally ordered*) *nontrivial approximate Euler characteristic satisfying the extra conditions* (*v,vi*).

Proof: W.l.o.g. assume that a is nonstandard. Let R be the totally ordered ring whose nonnegative part is M. Notice that $x \mathrel{\dot{\le}} y$ iff $x \le (1 + c^{-1})y$ for some $c > \omega$. Fix $\varepsilon = 1/n$, where $n \in \mathrm{Log}(M) \smallsetminus \omega$ (say, $n = |a|$). If A is a definable set in $[0, a]$, then A is definable in M by a $\Sigma^b_\infty(\alpha)$-formula, hence there exists an $s \in M$ such that $M \vDash (A \precsim_\varepsilon s \land A \not\precsim_\varepsilon s - 1)$; we define $\xi(A) = s$. Then ξ is an approximate Euler characteristic by 3.10, 3.13, 3.17, 3.12, 3.16, and 3.11 (as any injection defined by a bounded formula has a retraction definable by a bounded formula). The extra conditions hold for ξ because of Theorems 3.19 and 3.20. □

The complexity class S^P_2, defined independently by Russell and Sundaram [33], and Canetti [6], consists of languages L for which there exists a poly-time predicate R such that

$$\begin{aligned} x \in L &\Rightarrow \exists y\, \forall z\, R(x, y, z), \\ x \notin L &\Rightarrow \exists z\, \forall y\, \neg R(x, y, z), \end{aligned}$$

where $|y|, |z|$ are implicitly bounded by a polynomial in $|x|$. The class S_2^P occupies an interesting position inside the second level of *PH*: obviously $S_2^P \subseteq \Sigma_2^P \cap \Pi_2^P$, we also know that $MA \subseteq S_2^P$ (hence $BPP \subseteq S_2^P$), and $P^{S_2^P} = S_2^P$ (hence $\Delta_2^P \subseteq S_2^P$) [33], and the standard proof of the Karp–Lipton theorem shows that $NP \subseteq P/poly$ implies $PH = S_2^P$. The definition of S_2^P does in no way guarantee abundance of witnesses for the existential quantifiers; surprisingly, Cai [5] has shown that nevertheless $S_2^P \subseteq ZPP^{NP}$. We will formalize this result in bounded arithmetic. (The other results mentioned above are also easy to prove in bounded arithmetic, we leave the details to the reader.)

Theorem 4.11 (*in* $T^1_2 + rWPHP(PV_2)$) *The complexity class* S_2^P *is contained in* ZPP^{NP}*.*

Proof: Let $L \in S_2^P$. Fix a constant c, and a poly-time relation R such that

$$x \in L \Rightarrow \exists y < 2^{|x|^c}\, \forall z < 2^{|x|^c}\, R(x, y, z),$$
$$x \notin L \Rightarrow \exists z < 2^{|x|^c}\, \forall y < 2^{|x|^c}\, \neg R(x, y, z).$$

By the relativization of the formalized Wilkie's witnessing theorem [18, P. 1.16] applied to Corollary 4.4, there exists a ZPP^{NP}-predicate P definable in $T^1_2 + rWPHP(PV_2)$ such that the same theory proves

$$P(x) \Rightarrow \exists D \subseteq 2^{|x|^c}\, \forall z < 2^{|x|^c}\, \exists y \in D\, R(x, y, z),$$
$$\neg P(x) \Rightarrow \exists D \subseteq 2^{|x|^c}\, \forall y < 2^{|x|^c}\, \exists z \in D\, \neg R(x, y, z).$$

Clearly, the conditions implied by $x \in L$ and $\neg P(x)$ are contradictory, and vice versa, hence $x \in L$ iff $P(x)$. □

Another application of approximate counting in computational complexity is the equivalence of public-coin and private-coin interactive protocols [15]. We illustrate it on the example of the isomorphism problem: given two structures G_0 and G_1 (as tables) of the same signature, determine whether $G_0 \simeq G_1$. (The most prominent, and indeed universal, special case is when the structures are graphs.) The problem is obviously in *NP*, and its complement admits a simple private-coin interactive proof system: the verifier picks randomly an $i < 2$, and a permutation π, and sends $\pi(G_i)$ to the prover, who has to determine i. If $G_0 \not\simeq G_1$, a (computationally unlimited) prover can succeed with probability 1, whereas if $G_0 \simeq G_1$, no prover can do any better (or worse, for that matter) than 1/2. It is much harder to construct a public-coin proof system (i.e., an *AM*-algorithm) for the same problem, and it requires approximate counting.

Theorem 4.12 (*in* $T^1_2 + sWPHP(PV_2)$) *The isomorphism problem is in coAM.*

Proof: For simplicity, we will ignore floor and ceiling signs. Put $\varepsilon = 1/42$. As in the proof of Lemma 3.14, there exists a definable *prAM*-problem $L = \langle L^+, L^- \rangle$ such that

$$X_b \not\precsim_\varepsilon a \Rightarrow \langle a, b \rangle \in L^+,$$
$$X_b \precsim_\varepsilon \tfrac{42}{43} a \Rightarrow \langle a, b \rangle \in L^-$$

for any parametric family of *NP*-sets X_b. As *prAM* is closed under bounded existential quantification, conjunction, and disjunction, we can define a *prAM* problem $L = \langle L^+, L^- \rangle$ such that

$$\exists a \left(\left(A_0 \not\precsim_\varepsilon a \vee A_1 \not\precsim_\varepsilon a \right) \wedge W_0 \cup W_1 \not\precsim_\varepsilon \frac{3n!}{2a} \right) \Rightarrow \langle G_0, G_1 \rangle \in L^+,$$
$$\forall a \left(\left(A_0 \precsim_\varepsilon \tfrac{42}{43} a \wedge A_1 \precsim_\varepsilon \tfrac{42}{43} a \right) \vee W_0 \cup W_1 \precsim_\varepsilon \frac{4n!}{3a} \right) \Rightarrow \langle G_0, G_1 \rangle \in L^-,$$

where G_0, G_1 are structures with domain n, and A_i and W_i are the Σ^b_1-sets

$$A_i = \mathrm{Aut}(G_i) = \{\pi \in S_n;\, \pi(G_i) = G_i\},$$
$$W_i = \{\pi(G_i);\, \pi \in S_n\},$$

where S_n is the set of all permutations of n. It suffices to show

$$G_0 \not\simeq G_1 \Rightarrow \langle G_0, G_1 \rangle \in L^+,$$
$$G_0 \simeq G_1 \Rightarrow \langle G_0, G_1 \rangle \in L^-.$$

Claim 1

(*i*) *If* $A_i \precsim_\varepsilon a$, *and* $W_i \precsim_\varepsilon b$, *then* $ab \geq \frac{5}{6} n!$.

(*ii*) *If* $A_i \not\precsim_\varepsilon a$, *and* $W_i \not\precsim_\varepsilon b$, *then* $ab \leq \frac{10}{9} n!$.

Proof: (i): if $H \in W_i$, and π_0 is any permutation such that $H = \pi_0(G_i)$, then the mapping $\pi \mapsto \pi_0 \circ \pi$ is a poly-time bijection of A_i onto $M(H) := \{\pi;\, \pi(G_i) = H\}$, with $\pi \mapsto \pi_0^{-1} \circ \pi$ being its inverse. It follows that $M(H) \precsim_\varepsilon \frac{43}{42} a$ by Corollary 3.11, thus

$$M := \{\langle \pi, \pi(G_i) \rangle;\, \pi \in S_n\} = \dot{\bigcup_{H \in W_i}} M(H) \precsim_\varepsilon \tfrac{9}{8} ab$$

by Theorem 3.19. Clearly $\pi \mapsto \langle \pi, \pi(G_i) \rangle$ is a bijection of S_n onto M. Moreover, there exists a poly-time enumeration of S_n by $n!$, hence $\frac{7}{6} ab \twoheadrightarrow n!$, which implies $n! \leq \frac{6}{5} ab$ by *sWPHP*.

(ii): similar. □ (Claim 1)

Assume $G_0 \not\simeq G_1$, and find a_0, a_1 such that $A_i \not\precsim_\varepsilon a_i$, $A_i \precsim_\varepsilon \frac{43}{42} a_i$. We have $W_i \not\precsim_\varepsilon 13n!/16a_i$ by (i). The sets W_i are disjoint, hence

$$W_0 \cup W_1 \not\precsim_\varepsilon \frac{3n!}{4} \left(\frac{1}{a_0} + \frac{1}{a_1} \right) \geq \frac{3n!}{2a_i}$$

for some i by Theorem 3.17, thus $\langle G_0, G_1 \rangle \in L^+$.

On the other hand, assume $G_0 \simeq G_1$, and let a_i be such that $A_i \precsim_\varepsilon \frac{42}{43} a_i$, $A_i \not\precsim_\varepsilon \frac{20}{21}$. Then $W_0 \cup W_1 = W_i \precsim_\varepsilon 7n!/6a_i$ by (ii), as $W_0 = W_1$. Consequently $\langle G_0, G_1 \rangle \in L^-$. □

Notice that if we change the definition of *AM* formalized in bounded arithmetic to use $\precsim$ instead of $\preceq$ (which might be a good idea anyway), the statement of Theorem 4.12 becomes $\forall \Sigma^b_2$, hence we can prove it already in $T^1_2 + rWPHP(PV_2)$.

References

[1] Olaf Beyersdorff and Sebastian Müller, *A tight Karp–Lipton collapse result in bounded arithmetic*, ACM Transactions on Computational Logic, to appear.

[2] Samuel R. Buss, *Bounded arithmetic*, Bibliopolis, Naples, 1986, revision of 1985 Princeton University Ph.D. thesis.

[3] ———, *Relating the bounded arithmetic and polynomial time hierarchies*, Annals of Pure and Applied Logic 75 (1995), no. 1–2, pp. 67–77.

[4] ———, *First-order proof theory of arithmetic*, in: Handbook of Proof Theory (S. R. Buss, ed.), Studies in Logic and the Foundations of Mathematics vol. 137, Elsevier, Amsterdam, 1998, pp. 79–147.

[5] Jin-Yi Cai, $\mathrm{S}_2^p \subseteq \mathrm{ZPP}^{\mathrm{NP}}$, Journal of Computer and System Sciences 73 (2007), no. 1, pp. 25–35.

[6] Ran Canetti, *More on BPP and the polynomial-time hierarchy*, Information Processing Letters 57 (1996), no. 5, pp. 237–241.

[7] J. Lawrence Carter and Mark N. Wegman, *Universal classes of hash functions*, Journal of Computer and System Sciences 18 (1979), no. 2, pp. 143–154.

[8] Peter Clote and Jan Krajíček (eds.), *Arithmetic, proof theory, and computational complexity*, Oxford Logic Guides vol. 23, Oxford University Press, 1993.

[9] ———, *Open problems*, in *Arithmetic, proof theory, and computational complexity* [8], pp. 1–19.

[10] Alan Cobham, *The intrinsic computational difficulty of functions*, in: Proceedings of the 2nd International Congress of Logic, Methodology and Philosophy of Science (Y. Bar-Hillel, ed.), North–Holland, 1965, pp. 24–30.

[11] Stephen A. Cook, *Feasibly constructive proofs and the propositional calculus*, in: Proceedings of the 7th Annual ACM Symposium on Theory of Computing, 1975, pp. 83–97.

[12] Stephen A. Cook and Jan Krajíček, *Consequences of the provability of* $\mathbf{NP} \subseteq \mathbf{P/poly}$, Journal of Symbolic Logic 72 (2007), no. 4, pp. 1353–1371.

[13] Paul Erdős, *On a problem in graph theory*, Mathematical Gazette 47 (1963), no. 361, pp. 220–223.

[14] Oded Goldreich and Avi Wigderson, *Improved derandomization of BPP using a hitting set generator*, in: Proceedings of RANDOM-APPROX '99 (D. S. Hochbaum, K. Jansen, J. D. P. Rolim, and A. Sinclair, eds.), Lecture Notes in Computer Science vol. 1671, Springer, 1999, pp. 131–137.

[15] Shafi Goldwasser and Michael Sipser, *Private coins versus public coins in interactive proof systems*, in: Randomness and Computation (S. Micali, ed.), Advances in Computing Research vol. 5, JAI Press, Greenwich, 1989, pp. 73–90.

[16] Ronald L. Graham and Joel H. Spencer, *A constructive solution to a tournament problem*, Canadian Mathematical Bulletin 14 (1971), no. 1, pp. 45–48.

[17] Petr Hájek and Pavel Pudlák, *Metamathematics of first-order arithmetic*, Perspectives in Mathematical Logic, Springer, 1993, second edition 1998.

[18] Emil Jeřábek, *Dual weak pigeonhole principle, Boolean complexity, and derandomization*, Annals of Pure and Applied Logic 129 (2004), pp. 1–37.

[19] ———, *The strength of sharply bounded induction*, Mathematical Logic Quarterly 52 (2006), no. 6, pp. 613–624.

[20] ———, *On independence of variants of the weak pigeonhole principle*, Journal of Logic and Computation 17 (2007), no. 3, pp. 587–604.

[21] ———, *Approximate counting in bounded arithmetic*, Journal of Symbolic Logic 72 (2007), no. 3, pp. 959–993.

[22] Jan Krajíček, *No counter-example interpretation and interactive computation*, in: Logic From Computer Science, Proceedings of a Workshop held November 13–17, 1989 in Berkeley (Y. N. Moschovakis, ed.), Mathematical Sciences Research Institute Publications vol. 21, Springer, 1992, pp. 287–293.

[23] ———, *Bounded arithmetic, propositional logic, and complexity theory*, Encyclopedia of Mathematics and Its Applications vol. 60, Cambridge University Press, 1995.

[24] ———, *Uniform families of polynomial equations over a finite field and structures admitting an Euler characteristic of definable sets*, Proceedings of the London Mathematical Society 81 (2000), no. 3, pp. 257–284.

[25] ———, *Approximate Euler characteristic, dimension, and weak pigeonhole principles*, Journal of Symbolic Logic 69 (2004), no. 1, pp. 201–214.

[26] Jan Krajíček, Pavel Pudlák, and Gaisi Takeuti, *Bounded arithmetic and the polynomial hierarchy*, Annals of Pure and Applied Logic 52 (1991), no. 1–2, pp. 143–153.

[27] Alexis Maciel, Toniann Pitassi, and Alan R. Woods, *A new proof of the weak pigeonhole principle*, Journal of Computer and System Sciences 64 (2002), no. 4, pp. 843–872.

[28] Kerry E. Ojakian, *Combinatorics in bounded arithmetic*, Ph.D. thesis, Carnegie Mellon University, Pittsburgh, 2004.

[29] Jeff B. Paris, Alex J. Wilkie, and Alan R. Woods, *Provability of the pigeonhole principle and the existence of infinitely many primes*, Journal of Symbolic Logic 53 (1988), no. 4, pp. 1235–1244.

[30] Pavel Pudlák, *Ramsey's theorem in bounded arithmetic*, in: Proceedings of Computer Science Logic '90 (E. Börger, H. K. Büning, M. M. Richter, and W. Schönfeld, eds.), Lecture Notes in Computer Science vol. 533, Springer, 1991, pp. 308–317.

[31] Александр А. Разборов, *Формулы ограниченной глубины в базисе* $\{\&, \oplus\}$ *и некоторые комбинаторные задачи*, in: Сложность вычислений и прикладная математическая логика (С. И. Адян, ed.), Вопросы кибернетики vol. 134, VINITI, Moscow, 1988, pp. 149–166 (in Russian).

[32] Søren M. Riis, *Making infinite structures finite in models of second order bounded arithmetic*, in Clote and Krajíček [8], pp. 289–319.

[33] Alexander Russell and Ravi Sundaram, *Symmetric alternation captures* **BPP**, Computational Complexity 7 (1998), no. 2, pp. 152–162.

[34] Michael Sipser, *A complexity theoretic approach to randomness*, in: Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 1983, pp. 330–335.

[35] Esther Szekeres and George Szekeres, *On a problem of Schütte and Erdös*, Mathematical Gazette 49 (1965), no. 369, pp. 290–293.

[36] Seinosuke Toda, *On the computational power of PP and* $\oplus P$, in: Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science, 1989, pp. 514–519.

[37] Domenico Zambella, *Notes on polynomially bounded arithmetic*, Journal of Symbolic Logic 61 (1996), no. 3, pp. 942–966.