Algebraic methods and bounded formulas

Domenico Zambella[1]

Department of Mathematics and Computer Science
University of Amsterdam
Plantage Muidergracht 24, 1018 TV Amsterdam, NL
Email: domenico@fwi.uva.nl

1 Introduction

We present some algebraic tools and apply them to the study of the expressive power of bounded formulas in second-order arithmetic. The accent is on negative results. The techniques studied here come from Boolean circuit complexity and are adapted to the context of arithmetic. One of the purposes of these article is to expose them to a public with interests ranging from arithmetic to finite model theory. Our exposition is self-contained.

The machinery developed in Section 3 runs, up to some extend, parallel to that of Smolensky [7] (but our approach reveal the intriguing role played in the construction by the Möbius inversion formula). In [7] an alternative proof of a theorem of Yao's [11] (exponential lower bound for circuits of bounded depth computing parity) is given. The thechniques used by Yao had been introduced by Ajtai [1] and, independently, by Furst, Saxe and Sipser [3]. Subsequently, these have been improved and refined by Håstad [4], Razborov [6] and others. All these proofs have, somehow, a topological flavor while the ideas introduced in [7] are of algebraic nature. Both methods can be adapted to our context and give interesting information on the combinatorial properties of Σ_0^p formulas. However, the method of Smolenky (that is based on some ideas of Razborov [5]) yields a stronger result than what is obtainable by the topological method. Namely, the non-definability result is extended to a language expanded with a generalized quantifier expressing counting modulo an odd prime. Moreover, non-definability extends to any function that 'approximates' parity on essentially more than half of the sets.

[1] Research supported by the Netherlands Foundation for Scientific Research (NWO) grant PGS 22-262

2 Preliminaries

The language L is that of second-order arithmetic. It consists of two constants: 0, 1, two binary functions: $+$, $\cdot$, and two binary relations: $<$, $\in$. Variables are of two sorts: first-order, $x, y, z, \cdots$ and second-order, $X, Y, Z, \cdots$ that are meant to range over numbers and, respectively, finite sets of numbers.

The semantics of this language is the usual one but for the following interpretations: $X<y$ holds when all elements of X are less than y. Note that terms are just polynomials in first-order variables.

Bounded quantifiers are those quantifiers that appear in the context: $(Qx{\in}X)\varphi$, $(Qx{<}t)\varphi$ or $(QX{<}t)\varphi$, where Q is either $\forall$ or $\exists$ and t is a term in which x does not occur. A formula is bounded if all of its quantifiers are. The class of bounded formulas without second-order quantifiers is denoted by Σ^p_0. In the following we shall concentrate on this class. This class is the ground level of a hierarchy of formulas, Σ^p_i, Π^p_i that is obtained by counting the alternation of second-order quantifiers.For $i{>}0$ these classes coincides with those of the polynomial time hierarchy.

The standard model of this language is the set of natural numbers together with the set of its finite subsets. To simplify our exposition we will work in an elementary extension of the standard model. This model will be kept fixed throughout this notes; we will call it simply: *the* model. The language L is expanded to include a name for every element of the model. We call these new constants parameters. The classes defined above are naturally extended to include formulas with parameters in the model (but they will keep the same name).

Consider the equivalence relation $a \sim b$ that holds iff for some standard n, either $a < b^n$ or $b < a^n$. Let f be a numerical function, write $\lesssim f(a)$ for "$\leq f(b)$ for some $b \sim a$". Only the letter a will be used with this meaning, so, there will be no risk of ambiguities.

Let $q > 1$, we write $\equiv_q$ for the relation "congruent modulo q". We will show that the formula $\|X\| \equiv_q x$ is not Σ^p_0 definable. Moreover, given any set S which is large enough, no Σ^p_0 formula coincides with $\|X\| \equiv_q x$ on 'essentially' more than half of the subsets X of S. We say that a formula $\varphi(x, X)$ **counts** X **modulo** q iff for all x, $\varphi(x, X)$ is true iff $\|X\| \equiv_q x$.

Theorem 1 Let a be arbitrary and fix a set $S \lesssim 2^a$ of cardinality s. Let $\varphi(x, X)$ be a Σ^p_0 formula with parameters $\lesssim 2^a$. Then the formula $\varphi(x, X)$ counts X modulo q for at most $2^{s-1}[1{+}(a/\sqrt{s})]$ subsets X of S. $\square$

The proof is given in Section 5. In Section 6 this theorem is generalized to a language containing a generalized quantifier for counting modulo a prime number p. The next sections are dedicated to some preparatory work which is of independent interest.

3 Formulas, functions and the Möbius inversion.

Fix a commutative ring with unity R. We assume that R is definable in the model (i.e., the domain and the operations of R are definable). Formulas will be interpreted as functions that take as input tuples of elements of the model and output 1_R if these make the formula true, 0_R otherwise (the subscript will be omitted in the sequel). We define sum and multiplication of function –as usual– by point-wise addition and product. Multiplication with elements of R is defined in a similar way, so, functions mapping sets into R constitute a ring and, forgetting multiplication of function, an R-module. We shall consider also the operators of sum $\sum_{x<t}$, $\sum_{Y<t}$ and the operator of product $\prod_{x<t}$; these act on functions in the obvious way. We denote by $\mu(X)$ the **Möbius function**: $\prod_{x\in X}(-1)$, that is, $\mu(X)$ is -1, if X has an odd number of elements, 1 otherwise.

We shall use some notions of elementary algebra. This notions are naturally formalizable in the model. In particular we shall use the fact that the module of unary functions with domain a finite (in the model) set of sets α has dimension $\|\alpha\|$. The functions $\{E{=}X\}_E$ are the canonical base of the module of unary functions: every function $\delta(X)$ trivially coincides with $\sum_E \delta(E)(E{=}X)$. In this section we shall construct other bases. Let $\varphi(X)$ be an arbitrary function: we use the following abbreviation

$$^{X}\varphi(X) \quad :\equiv \quad \sum_{E\subseteq X} \mu(E)\cdot\varphi(E).$$

The following property of the Möbius function will be used repeatedly in the algebraic manipulations of functions.

Fact 2 $\displaystyle\sum_{E\subseteq X} \mu(E) \quad\equiv\quad (X{=}\emptyset).$

Proof. If $X{\neq}\emptyset$, choose an $x{\in}X$,

$$\sum_{E\subseteq X} \mu(E) \;=\; \sum_{E\subseteq X\setminus\{x\}} \mu(E) + \mu(E\cup\{x\}) \;=\; \sum_{E\subseteq X\setminus\{x\}} \mu(E) - \mu(E) \;=\; 0.$$

On the other side, if $X=\emptyset$, then $\sum_{Y\subseteq X}\mu(Y)$ equals $\mu(\emptyset)$ and so, it equals 1 as required. $\square$

Fact 3 ${}^{X}(A\subseteq X) \quad\equiv\quad \mu(A)\cdot(X=A)$.

Proof. Observe that ${}^{X}(A\subseteq X)$ is equivalent to

$$\sum_{E\subseteq X}\mu(E)\cdot(A\subseteq E) \quad\equiv\quad (A\subseteq X)\sum_{E\subseteq X\setminus A}\mu(A\cup E) \quad\equiv\quad (A\subseteq X)\cdot\mu(A)\sum_{E\subseteq X\setminus A}\mu(E)$$

Now, apply Fact 2. $\square$

Lemma 4 ${}^{XX}\varphi(X) \quad\equiv\quad \varphi(X)$.

Proof. Observe that

$$\begin{aligned} {}^{XX}\varphi(X) \quad&\equiv\quad \sum_{E\subseteq X}\mu(E)\sum_{A\subseteq E}\mu(A)\cdot\varphi(A) \quad\equiv\quad \sum_{E\subseteq X}\mu(E)\sum_{A}\mu(A)\cdot\varphi(A)\cdot(A\subseteq E) \\ &\equiv\quad \sum_{A}\mu(A)\cdot\varphi(A)\sum_{E\subseteq X}\mu(E)\cdot(A\subseteq E) \quad\equiv\quad \sum_{A}\mu(A)\cdot\varphi(A)\cdot{}^{X}(A\subseteq X). \end{aligned}$$

The lemma follows from Fact 3. $\square$

If $\varphi(E)=0$ for all $E\subseteq X$, then ${}^{E}\varphi(E)=\sum_{Z\subseteq E}\mu(Z)\cdot 0=0$ for all $E\subseteq X$. From the lemma above it follows that also the converse holds. So, the functions $\mu(E)\cdot(E\subseteq X)$ are linearly independent and form a base; ^{X} is an invertible linear transformation. The function ${}^{E}\varphi(E)$ gives the components of $\varphi(X)$ with respect to the base $\mu(E)\cdot(E\subseteq X)$.

***

We define the **degree** of a function $\varphi(X)$ to be the least d such that ${}^{X}\varphi(X)=0$ for all X of cardinality $>d$.

Fact 5 ${}^{X}[\varphi(X)\cdot\psi(X)] \quad\equiv\quad \sum_{A\cup B=X}\mu(A\cap B)\cdot{}^{A}\varphi(A)\cdot{}^{B}\psi(B)$.

Proof. Expressing φ and ψ trough their transforms, we have

$$\varphi(X)\cdot\psi(X) \quad\equiv\quad \sum_{A,B\subseteq X}\mu(A)\cdot{}^{A}\varphi(A)\cdot\mu(B)\cdot{}^{B}\psi(B)$$

$$\equiv \sum_{Y\subseteq X} \sum_{A\cup B=Y} \mu(A)\cdot{}^{A}\varphi(A)\cdot\mu(B)\cdot{}^{B}\psi(B)$$

$$\equiv \sum_{Y\subseteq X} \mu(Y) \sum_{A\cup B=Y} \mu(A\cap B)\cdot{}^{A}\varphi(A)\cdot{}^{B}\psi(B).$$

The fact follows from Lemma 4. □

The fact above generalizes easily to the operator $\prod$. We state it in the following fact. (We write E_x for the x-th set coded by E in some natural way, e.g., $y{\in}E_x$ iff $\langle x,y\rangle{\in}E$.)

Fact 6 $${}^{X}\prod_{x<t} \varphi(x,X) \quad\equiv\quad \sum_{\bigcup\limits_{x<t} E_x = X} \quad \prod_{x<t} {}^{E_x}\varphi(x,E_x).$$ □

From these facts we can give an upper bound to the degree of a products from the degree of the factors. E.g., the degree of $\prod_{x<t}\varphi(x)$ is at most the sum of the degrees of $\varphi(x)$ for $x<t$. Needless to say, the degree of $\sum_{x<t}\varphi(x)$ is just the maximum of the degrees of $\varphi(x)$ for $x<t$.

***

The following lemma shows a simple way of obtaining new bases of the module of unary functions.

Lemma 7 Let $\delta(X)$ be any function. The function $\delta(E\cap X)$ has degree $\|E\|$. Moreover, if for all X, ${}^{X}\delta(X)$ has an inverse in R then $\{\delta(E\cap X)\}_E$ is a base of the module of the unary functions.

Proof. To check that $\delta(E\cap X)$ has degree $\|E\|$, compute ${}^{X}\delta(E\cap X)$ and check that it is 0 for $X \not\subseteq E$

$$\sum_{A\subseteq X} \mu(A)\delta(E\cap A) \quad\equiv\quad \sum_{B\subseteq X\setminus E} \; \sum_{A\subseteq X\cap E} \mu(A\cup B)\cdot\delta(A)$$

$$\equiv \sum_{A\subseteq X\cap E} \mu(A)\cdot\delta(A) \sum_{B\subseteq X\setminus E} \mu(B)$$

$$\equiv \sum_{A\subseteq X\cap E} \mu(A)\cdot\delta(A)\cdot(X{\subseteq}E)$$

$$\equiv \quad (X{\subseteq}E)\sum_{A\subseteq X}\mu(A){\cdot}\delta(A) \quad \equiv \quad (X{\subseteq}E)^{X}\delta(X).$$

To check that $\{\delta(E\cap X)\}_E$ generates all the unary functions, observe that from the equivalence above (up to renaming of variables) we have

$$[{}^{A}\delta(A)]^{-1}\sum_{E\subseteq A}\mu(E)\delta(E\cap X) \quad \equiv \quad (X{\subseteq}A).$$

Since $\{(A{\subseteq}X)\}_A$ is a base, the claim follows. We also have that if we restrict the range of X to the subsets of some set S, then $\{\delta(E{\cap}X)\}_{E\subseteq S}$ generates all functions. So, to check the linear independence of $\{\delta(E{\cap}X)\}_E$, restrict E and X to range over the subsets of S and apply an obvious cardinality argument. By the arbitrariness of S the claim follows. □

Let g be any element of R such that $(1{-}g)^{-1}$ exists in R, to prove Theorem 1 we shall need the fact that the functions $g^{\|E\cap X\|}$, form a base. We check that the hypothesis of the lemma are satisfied. If $X=\emptyset$ then ${}^{X}(g^{\|X\|})$ is 1, otherwise, pick an arbitrary $a{\in}X$,

$$\sum_{E\subseteq X}\mu(E){\cdot}g^{\|E\|} \;=\; \sum_{E\subseteq X\setminus\{a\}}\mu(E){\cdot}[g^{\|E\|}-g^{\|E\|+1}] \;=\; (1{-}g)\sum_{E\subseteq X\setminus\{a\}}\mu(E){\cdot}g^{\|E\|}$$

So, ${}^{X}(g^{\|X\|})$ is $(1{-}g)^{\|X\|}$ and our claim follows from the lemma.

4 Approximations

In general, even very simple formulas may have high degree. For instance, by Fact 2, the formula $A{\subseteq}X$ has degree $\|A\|$. Nevertheless, we shall see that every formula φ in Σ_0^p can be approximated by a function ψ of low degree, namely, of degree that is polynomial in the logarithm of the parameters in occurring in φ. By 'approximating' we mean that for all but an infinitesimal fraction of the sets $X{\in}\alpha$ the functions $\varphi(X)$ and $\psi(X)$ are equivalent. The rest of this section is devoted to the proof of this theorem. We will give two proofs: the first works only when R has characteristic between 1 and a, the second is completely general. The first proof is due to Smolensky [7] and uses combinatorial techniques of Razborov [5]. The second is of Tauri [8], it uses ideas of Vazirani and Vardi [10]. The first proof we give is simpler and it is sufficient to prove Theorem 1 and Corollary 12. The second proof is included for completeness and because of the general combinatorial ideas used there that make the method interesting in itself.

Theorem 8 Let a be arbitrary and fix a set α be an arbitrary set of sets $\lesssim 2^a$. Let $\varphi(X)$ be a Σ_0^p formula with parameters $\lesssim 2^a$. Then, there is a function $\psi(X)$ of degree $\lesssim a$, that is equal to $\varphi(X)$ for all but at most $2^{-a}\|\alpha\|$ sets $X\in\alpha$.

Proof. We call the function ψ an **approximation** of φ and the fraction of sets in α such that $\varphi(X)$ differs from $\psi(X)$, the **error probability**. We will proceed by induction on the syntax of formulas. To keep induction going some natural uniformity in the parameters is needed. We claim (and leave to the reader to check) that the approximation given in the construction below is uniform on the first order parameters of the formula.

To prove the basic step of the induction let us assume that second-order equality does not occur in φ (if it does we eliminate it using extensionality). By Fact 2 and Fact 3, the atomic formulas $t{=}s$, $t{<}s$ and $t{\in}X$ have degree ≤ 1.

Eliminate in φ all connectives different of $\neg$, $\wedge$, and $\exists$. Also, for definiteness, replace the quantifiers of the form $(\exists x{\in}T)$ with $(\exists x{<}t)(x{\in}T) \to$, where t is an appropriate closed term. So, we may assume that all quantifiers occurring in the formulas are of the form $(\exists x{<}t)$.

The induction step for negation is trivial. In fact, negation coincides with '$1-$', so, it is a linear operator and it does not increase the degree. Conjunction (i.e., multiplication of functions) is easy. Suppose that φ is of the form $\varphi_1 \wedge \varphi_2$ and that φ_1, φ_2 have approximations ψ_1, ψ_2 of degree $\lesssim a$. We claim that the product $\psi_1{\cdot}\psi_2$ is the required approximation of φ. From Fact 5 above it follows that $\psi_1{\cdot}\psi_2$ has degree $\lesssim 2a^n \sim a$. The error probability is at most the sum of that of the two conjuncts separately, i.e., $< 22^{-a}$. The claim follows by applying the induction hypothesis with $a{+}1$ for a.

The difficult part of the proof consists in proving the induction step for the existential quantifier. A 'brute force' strategy: replace quantification with product (as we did with conjunction), is doomed to fail. In fact, large products make us loose any control on the degree of the functions. So, we will express (though, in an approximate form) existential quantification using sums and 'small' products, i.e., products of the form $\prod_{x<a^n}\varphi(x,X)$. Then, if for all $x{<}a^n$ the degree of $\varphi(x,X)$ is less than a^n than the degree of $\prod_{x<a^n}\varphi(x,X)$ is less than $a^{2n} \sim a$.

For the expository reasons explained above, we first prove the theorem in a special case, that is, when R has nonzero characteristic $n \lesssim a$. Let $t \lesssim 2^a$ be a closed term. Assume that, for all $x < t$, the function $\psi(x,X)$ of degree $\lesssim a$ is an approximation of $\varphi(x,X)$. We can assume that these approximations have error probability $< 2^{-2a}$

(apply the inductive hypothesis with $2a$ for a). So, all but at most $2^{-a}\|\alpha\|$ sets $X\in\alpha$ are such that $\psi(x,X)$ equals $\varphi(x,X)$ for all $x<t$.

We claim that there is a function $\lambda(x,y)$ such that, for all but at most $2^{-a}\alpha$, $(\exists x{<}t)\varphi(x,X)$ holds iff there is a $y{<}na$ such that $\sum_{x<t}\lambda(x,y)\varphi(x,X)$ is one. Assume the claim, then for all these X in α, we have that $(\exists x{<}t)\varphi(x,X)$ equals

$$(*)\qquad 1-\prod_{y<na}\left[1-\sum_{x<t}\lambda(x,y)\varphi(x,X)\right].$$

In fact, if $(\exists x{<}t)\varphi(x,X)$ holds then for some $y{<}na$ the sum above is 0 then the product is 0. Vice versa, if $(\exists x{<}t)\varphi(x,X)$ does not hold then the sum above is 0 for every $y{<}na$ (note that this does not depend on the function λ), so the product is 1. The induction step follows since the required approximation can be found by substituting $\psi(x,X)$ for $\varphi(x,X)$ in (*). The error probability of these approximation is at most $2^{-a}+2^{-a}=2^{-(a-1)}$. The claim follows by applying the induction hypothesis with $a{+}1$ for a.

To prove the claim fix an arbitrary $\hat{X}$ such that for some $\hat{x}$, the formula $\varphi(\hat{x},\hat{X})$ is true. We choose at random (w.r.t. the uniform distribution), a function λ^* from $[0,t)$ into $[0,n{-}1)$. We show that, with probability at least $1/n$, we have that $\sum_{x<t}\lambda(x)\varphi(x,\hat{X})$ is 1. In fact, for every choice of $\lambda(x)$ for $x{\neq}\hat{x}$ there is one choice (out of n) of $\lambda(\hat{x})$ that makes the sum equal to 1. Now, let $\lambda(y,x)$ be obtained by choosing independently na times a function λ^* as above. The probability that for every $y{<}n(a{+}1)$ the sum is $\neq 1$ is $(1-(1/n))^{n(a+1)}\leq 2^{-(a+1)}$. The claim follows by simple counting.

This completes the proof of the special case of the theorem; now, let us resume the general proof. We need to prove the inductive step for the existential quantifier in the case R has characteristic 0. The following lemma of Valiant and Vazirani [10] yields us the technical tools we need to complete the proof of the theorem. The idea of applying it in this context is of Tauri [8]. To better understand the statement of the lemma and its role in the proof. Let us make some simple considerations.

The idea of the lemma is to hash the interval $[0,t)$ into $a{+}1$ subsets $C_0,\ldots,C_a$ such that, if $(\exists x{<}t)\vartheta(x,X)$ is true, then at least one of these subsets isolates exactly one witness, i.e., $(\exists z{\leq}a)(\exists! x{\in}C_z\ \varphi(x,X)$. So, we can rewrite $(\exists x{<}t)\vartheta(x,X)$ as

$$1-\prod_{z<a}\left[1-\sum_{x\in C_z}\vartheta(x,X)\right].$$

The set C is a suitable a parameter depending on φ. We want to find a C that works for a 'large' fraction of the $X{\in}\alpha$. Recall that C_z is the z-th element of the sequence of sets coded (in some natural way) by the set C.

Lemma 9 Fix a formula of the form $(\exists x{<}2^c)\vartheta(x,X)$. Let α an arbitrary set of sets X. There is a set $C < \langle c, 2^c\rangle$ such that for all but $< (1/2)\|\alpha\|$ sets $X{\in}\alpha$,

$$(*) \qquad (\exists x{<}t)\vartheta(x,X) \leftrightarrow (\exists z{\leq}c)(\exists! x{\in}C_z)\vartheta(x,X).$$

Proof. Fix an arbitrary $\hat{X}{\in}\alpha$ such that $(\exists x{<}2^c)\vartheta(x,\hat{X})$. We shall see that, choosing C at random (w.r.t. the distribution specified below), we have that $(\exists z{\leq}c)(\exists! x{\in}C_z)\vartheta(x,X)$ with probability $> 1/2$. So, by counting, there is a set C that satisfy (*) for all but at most $(1/2)\|\alpha\|$ sets $X{\in}\alpha$. The lemma follows.

Fix an arbitrary injection of the interval $[0,2^c)$ into the set of binary strings of length c. We can view $\{0,1\}^c$ as a vector space on the finite field $\{0,1\}$. The canonical base for this vector space is denoted with $e_1,..,e_c$. Below, we shall identify numbers $<2^c$ and vectors; to simplify notation we assume that 0 corresponds to the zero of the vector space. The scalar product of two vectors u, $v \in \{0,1\}^c$ is defined in the natural way: it is 0 if the coordinates of u and v coincides on an even number entries, 1 otherwise.

The set C is constructed as follows. Let $v_0 := 0$ (the zero of the vector space) and choose consecutively a mutually orthogonal vectors $v_1,..,v_c$ by picking, v_{z+1} at random (with uniform distribution) among all vectors orthogonal to $v_0,..,v_z$. For every $z \leq c$ let C_z be the set of the vectors orthogonal to $v_0,..,v_z$.

The **rank** of a subset of $[0,2^c)$ is the dimension of the minimal subspace containing it. Let us write F for the set $\{x{<}t : \vartheta(x,\hat{X})\}$. We shall show that, whatever the non-empty set F is, by choosing C at random we obtain that, with probability $> 1/2$, there is a $z \leq c$ such that $C_z \cap F$ has cardinality 1, that is, $(\exists! x{\in}C_z)\vartheta(x,\hat{X})$ for some $z \leq c$.

It suffices to prove that with probability greater than $1/2$ the rank of $C_z \cap F$ is 1 for some $z \leq c$. In fact, suppose the rank of $C_z \cap F$ is 1. Then, either the cardinality of $C_z \cap F$ is 1 –in this case we are done– or 0 belongs $C_z \cap F$, so, since $C_c = \{0\}$, stage a will be successful.

Let $z \leq c$ be arbitrary and let the rank of $C_z{\cap}F$ be $d > 1$. We claim that $C_{z+1}{\cap}F \neq \emptyset$ with probability greater than $1{-}2^{-d}$. To prove the claim, observe that all probability distribution are invariant under orthonormal transformations. So, we can assume that the vector $v_1,..,v_z$ are just the base vectors $e_1,..,e_z$ and that $(C_z \cap F)$ is contained in the subspace generated by $e_{z+1},..,e_{z+d}$ and contains at least these vectors. Now, the vector v_{z+1} is chosen randomly among the vectors orthogonal to $e_1,..,e_z$, i.e., among the non-zero linear combinations of $e_{z+1},..,e_c$. Since

$$C^{z+1} \cap F \;=\; C^{z+1} \cap (C^z \cap F) \;\supseteq\; C^{z+1} \cap \{e_{z+1}, .., e_{z+d}\},$$

$C^{z+1} \cap F$ is empty if and only if the $(z{+}1)$-th,...,$(z{+}d)$-th coordinates of v_{z+1} are all 0. This happens with probability less than 2^{-d}. This proves the claim. From this clam it follows easily that the probability of $(\exists z{\le}c)\|C^z \cap F\| = 1$ is, at least,

$$\prod_{d=2}^{a}(1{-}2^{-d}) \;>\; \prod_{d=2}^{\infty}(1{-}2^{-d}) \;\ge\; \prod_{d=2}^{\infty} 2^{-2^{1-d}} \;=\; \frac{1}{2}$$

(for the last inequality we have used that $1{-}x \le 2^{-2x}$ for all positive $x \le 1/2$). The theorem follows. □

Assume that, for all $x < t$, the function $\psi(x, X)$ of degree $\lesssim a$ is an approximation of $\varphi(x, X)$. Apply the lemma, a times, iteratively, to those X for which (*) does not hold. Concatenating the results we obtain a set C such that

$$(**) \qquad (\exists x{<}t)\varphi(x, X) \leftrightarrow (\exists w{\le}a)(\exists z{\le}a)(\exists! x{\in}C_{w,z})\varphi(x, X)$$

holds for all sets of α but $< 2^{-a}\|\alpha\|$. So, the approximation of $(\exists x{<}t)\varphi(x, X)$, is

$$\prod_{w<a}\prod_{z<a}\left[1{-}\sum_{x\in C_{w,z}} \psi(x, X)\right]$$

The degree of this function is a^2 times the degree of ψ. This completes the proof of the induction step for the existential quantifier and hence of Theorem 8. □

5 Proof of Theorem 1

We have now assembled all what is needed in order to prove Theorem 1.

Proof of Theorem 1. Let α be the set of those $X{\subseteq}S$ such that the formula $\varphi(x, X)$ is counts X modulo q. Fix a ring R that has a q-th root of the unity g such that $1{-}g$ has an inverse in R. By the observation above, $g^{\|E\cap X\|}$ is a base, so, every function $\psi(X)$ can be written as

$$\psi(X) \equiv \sum_E g^{\|E\cap X\|}\cdot{}^*\psi(E),$$

for some function ${}^*\psi(E)$. Assume s is odd (the case when s is even is similar but require somewhat lengthier writing). Every subset of S has either cardinality $<s/2$ or it is the complement of a set of cardinality $<s/2$, so, every function $\psi(X)$ can be written as follows

$$\begin{aligned}\psi(X) &\equiv \sum_{E\,:\,\|E\|<\frac{s}{2}} \left[g^{\|E\cap X\|}\cdot{}^*\psi(E) + g^{\|E^c\cap X\|}\cdot{}^*\psi(E^c)\right]\\ &\equiv \sum_{E\,:\,\|E\|<\frac{s}{2}} g^{\|E\cap X\|}\cdot{}^*\psi(E) + g^{\|X\|}\sum_{E\,:\,\|E\|<\frac{s}{2}} g^{-\|E\cap X\|}\cdot{}^*\psi(E^c).\end{aligned}$$

So, every function is the sum of a function of degree at most $s/2$ and a function which is $g^{\|X\|}$ multiplied by a functions of degree $\leq s/2$. Since g is a q-root of unity, the function $g^{\|X\|}$ coincides in α with the linear combination $\sum_{n<q} g^n\cdot\varphi(n,X)$. Consequently, on a set of cardinality $<(1-2^{-a^n})\|\alpha\|$ the function $g^{\|X\|}$ coincides with a function of degree $\lesssim a$ (the standard constant n will be fixed below). So, there is a submodule of dimension $<(1-2^{-a^n})\|\alpha\|$ where every function has degree $\lesssim (s/2)$ $+\ a$. We can derive the claimed bound on the cardinality of α from a a simple argument of dimensionality. The functions of degree $\lesssim a$ are

$$\lesssim \sum_{i<\frac{s}{2}+a}\binom{s}{i} \leq 2^{s-1}+\sum_{i=\frac{s}{2}}^{\frac{s}{2}+a}\binom{s}{i} \leq 2^{s-1}+\binom{s}{s/2}a \leq 2^{s-1}\left(1+\frac{2a}{\sqrt{s}}\right).$$

(For the last inequality use Robbins' sharp form of Stirling's formula, see [2] II.9) So, assuming n is large enough,

$$(1-2^{-a^n})\|\alpha\| < \left(1+\frac{a^n}{\sqrt{s}}\right).$$

Recall that $S\lesssim 2^a$, so, letting n be large enough, $s<2^{a^n}$. Therefore, we may conclude that

$$\|\alpha\| < 2^{s-1}\left(1+\frac{3a^n}{\sqrt{s}}\right),$$

so Theorem 1 follows. $\square$

6 Expansions of the language

Given a formula $\tau(X)$ with only free variable X we define the (bounded first-order) generalized quantifier Q_τ stipulating (by induction on the number of nested generalized quantifiers) that $(Q_\tau x{<}t)\varphi(x)$ holds iff $\tau(\{x{<}t\,:\,\varphi(x)\})$. The class $\Sigma_0^p(Q_\tau)$ is the minimal class that contains Σ_0^p and is closed under Q_τ quantification. Theorem 8 can be easily extended to the following.

Theorem 10 Let α be an arbitrary set of sets $\lesssim 2^a$. Let $\tau(X)$ such that ${}^X\tau(X) \neq 0$ for $\lesssim 2^a$ subsets $X \lesssim 2^a$. Let $\varphi(X)$ be a $\Sigma_0^p(Q_\tau)$ formula with parameters $\lesssim 2^a$. Then, there is a function $\psi(X)$ of degree $\lesssim a$, that is equal to $\varphi(X)$ for all but at most $2^{-a}\|\alpha\|$ sets $X{\in}\alpha$.

Proof. We only need to prove the induction step of the proof of Theorem 8 for the generalized quantifier Q_τ. Consider the identity

$$\tau(X) \quad \equiv \quad \sum_{E\subseteq X} \mu(E)\cdot{}^E\tau(E).$$

Substituting the set $\{x{<}t\,:\,\varphi(x)\}$ for X, we obtain

$$(Q_\tau x{<}t)\varphi(x,X) \quad \equiv \quad \tau(\{x{<}t\,:\,\varphi(x,X)\}) \quad \equiv \quad \sum_{E<t} [(\forall x{\in}E)\varphi(x,X)]\cdot\mu(E)\cdot{}^E\tau(E).$$

Take as approximation of $(Q_\tau x{<}t)\varphi(x,X)$ the linear combination

$$(*) \qquad \sum_{E<t} \psi(E,X)\cdot\mu(E)\cdot{}^E\tau(E),$$

where, for E such that ${}^E\tau(E) \neq 0$, we let $\psi(E,X)$ be the approximation of the formula $(\forall x{\in}E)\varphi(x,X)$ as given by the induction hypothesis. For all other E, we let $\psi(E,X)$ be $(\forall x{\in}E)\varphi(x,X)$. It is clear that the transform of (*) has degree $\lesssim a$. Since ${}^X\tau(X) \neq 0$ for $\lesssim 2^a$ we can make the error probability of the approximations $\psi(E,X)$ small enough so that the error probability of (*) is $< 2^{-a}$. This proves the theorem. □

Lemma 11 Let S be a set and p a prime number. If R has characteristic p the formula $X{\subseteq}S \ \wedge\ \|X\| \equiv_p 0$ has degree p.

Proof. By Fermat's little theorem the formula $\|X\| \equiv_p 0$ is equivalent to

$$\left(\sum_{x\in S} x{\in}X\right)^p,$$

by Fact 5, $\|X\|^p$ has degree p. $\square$

Corollary 12 Let a be arbitrary and fix a set $S \lesssim 2^a$ of cardinality s. Let $p \lesssim a$ be a prime that does not divide q. Let $\varphi(x, X)$ be a $\Sigma_0^p(Q_{\|X\| \equiv_p 0})$ formula with parameters $\lesssim 2^a$. Then the formula $\varphi(x, X)$ counts X modulo q for at most $2^{s-1}[1+(a/\sqrt{s})]$ subsets X of S.

Proof. If q does not divide p there is a field F of characteristic p with a q-th root of the unity. By Theorem 10 and Lemma 11 the proof of Theorem 1 remains valid when Σ_0^p is replaced with $\Sigma_0^p(Q_{\|X\| \equiv_p 0})$ provided that we set $R := F$. $\square$

We conclude by remarking that for the proof above it is essential that p is a prime. It is open whether the corollary above holds for a composite numbers. Also it is not known whether it holds for $p \not\lesssim a$ or whether the lower bound in the error probability is optimal.

References

[1] M.Ajtai, Σ_1^1 formulas on finite structures. *Annals of Pure and Applied Logic*, vol. 24, pp. 1–48, 1983

[2] W.Feller, *An Introduction to Probability Theory and its Applications*, vol. I. John Wiley and Sons (1966)

[3] M.Furst, J.B.Saxe and M.Sipser, Parity circuits and the polynomial time hierarchy, In the proceedings of the 22nd Annual *Symposium on Foundations of Computer Science*, (1981) pp. 260–270

[4] J.Håstad, Almost optimal lower bounds for small depth circuits. In S.Micali, ed., *Randomness and Computation*, vol. 5 of *Advances in Computing Research*, JAI Press, 1989, pp. 143–170

[5] A.A.Razborov, Lower bounds for the size of circuits of bounded depth with basis $\{\wedge, \oplus\}$. *Math. notes of the Academy of Science of the USSR*, vol. 41(4) (1987) pp.333–338

[6] A.A.Razborov, Bounded arithmetic and lower bounds in Boolean complexity. *Feasible Mathematics II* (1993)

[7] R.Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In the proceedings of the 19th Annual *ACM Symposium on the Theory of Computing*, (1987) pp. 77–82

[8] J.Tauri, Probabilistic polynomials, AC_0 functions, and the polynomial-time hierarchy. *Theoretical Computer Science*, vol. 113 (1993) 167–183

[9] S.Toda and M.Ogiwara, Counting classes are at least as hard as the polynomial-time hierarchy. In the proceedings of the 8th Annual *Conference on Structure in Complexity Theory*, (1991) pp. 2–12

[10] L.G.Valiant and V.V.Vazirani, NP is as easy as detecting unique solutions. *Theoretical Computer Science*, vol. 47 (1986) 85–93

[11] A.C.Yao, Separating the polynomial-time hierarchy by oracles. In the proceedings of the 26th Annual *Symposium on Foundations of Computer Science*, (1985) pp. 1–10