Notre Dame Journal of Formal Logic
Volume 38, Number 1, Winter 1997

Algebraic Methods and Bounded Formulas

DOMENICO ZAMBELLA

Abstract We present some algebraic tools useful to the study of the expressive power of bounded formulas in second-order arithmetic (alternatively, second-order formulas in finite models). The techniques presented here come from Boolean circuit complexity and are adapted to the context of arithmetic. The purpose of this article is to expose them to a public with interests ranging from arithmetic to finite model theory. Our exposition is self-contained.

1 Introduction We present some algebraic tools useful to the study of the expressive power of bounded formulas in second-order arithmetic (alternatively, second-order formulas in finite models). The techniques presented here come from Boolean circuit complexity and are adapted to the context of arithmetic. The purpose of this 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, to some extent, parallel to that of Smolensky [7]—our formulation is more explicit. In Section 4 we also include some related techniques developed in [8]. In [7] an alternative proof of a theorem of Yao (exponential lower bound for circuits of bounded depth computing parity) [11] is given. The techniques 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, however, a topological flavor, whereas 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, Smolenky's method (based on some ideas of Razborov [5]) yields a stronger result than what is obtainable by the topological method: namely, the nondefinability of parity is extended to a language expanded with a generalized quantifier expressing counting modulo a prime number > 2. Moreover, nondefinability extends to any function that *approximates* parity on essentially more than half of the sets.

Received December 5, 1995; revised March 21, 1997

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, $x, y, z, \ldots$ and second-order, $X, Y, Z, \ldots$ 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: $X < y$ (in words, *X is less than 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 bounded. The class of bounded formulas without second-order quantifiers is denoted by Σ_0^p. In the following we shall concentrate on this class. This class is the ground level of a hierarchy of formulas, Σ_i^p, Π_i^p that is obtained by counting the alternation of second-order quantifiers. For $i > 0$ these classes coincide 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. This model will be kept fixed throughout this note. The language L is expanded to include constants for every element of the standard model. We call these new constants *parameters*. The classes defined above are naturally extended to this expanded language (but they will keep the same name). Practically, we shall restrict the attention to formulas with a free variable X that ranges over the subsets of some fixed but arbitrary finite set S. The formulas may have arbitrary parameters. The size of these constants and the length of the formula are the relevant inputs of the theorems below.

Let $q > 1$. We write $=_q$ for the relation "congruent modulo q". The next theorem says that the formula $\|X\| =_q x$ is not Σ_0^p definable. (With $\|X\|$ we denote the cardinality of X.) Moreover, given any set S which is large enough, no Σ_0^p formula coincides with $\|X\| =_q x$ on *essentially* more than half of the subsets X of S. In Section 6 this theorem is generalized to a language containing a generalized quantifier for counting modulo a prime number p that does not divide q. We say that a formula $\varphi(x, X)$ *counts X modulo q* if and only if for all x, $\varphi(x, X)$ is true if and only if $\|X\| =_q x$.

Theorem 2.1 *Let $\varphi(x, X)$ be a Σ_0^p formula and let S be a (finite) set. The formula $\varphi(x, X)$ counts X modulo q for at most $2^{s-1}\left[1 + (a^m/\sqrt{s})\right]$ subsets X of S, where*

1. *s is the cardinality of S,*
2. *m is a constant that depends only on the syntax of $\varphi(x, X)$ (m is proportional to the length of $\varphi(x, X)$), and*
3. *a is any number which is larger than m and such that $2^{q^{\frac{1}{m}}}$ bounds S and all the parameters occurring $\varphi(x, X)$.*

To have a more concrete example in mind, fix $S = [0, s)$ and suppose that s is large enough to have that $2 \cdot \lceil \log s \rceil^{2m} < \sqrt{s}$ and that no parameter in $\varphi(x, X)$ does exceed s. Take $\lceil \log s \rceil^m$ for a and apply the theorem. We obtain that $\varphi(x, X)$ fails to count X modulo q for at least one subset X of S. The proof is given in Section 5. 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. Formulas will be interpreted as functions that take as input tuples of numbers and sets and output 1_R, if these make the formula true, 0_R otherwise (the subscript R 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. 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.

In the following, X is restricted to range over the subsets of S. The set of unary functions from the power set of S to R, constitutes a ring and, forgetting multiplication, an R-module. This module has dimension 2^s. The functions $\{E = X\}_E$—that is, the functions that map X to 1 if $X = E$ and to 0 otherwise—for E ranging over subsets of S, form the canonical base of this module. In fact, every function $\delta(X)$ can be written as $\sum_E \delta(E)(E = X)$. (When the range of a subscript is omitted this is implicitly understood to be S.) Lemma 3.6 below shows a useful way of constructing new bases. Let $\varphi(X)$ be an arbitrary function: we use the following abbreviation.

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

This is called the discrete Fourier transform of $\varphi(X)$. The following property of the Möbius function will be used repeatedly in the algebraic manipulations of functions.

Fact 3.1 $\sum_{E\subseteq X}\mu(E) \equiv (X = \varnothing)$.

Proof: If $X \neq \varnothing$, 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 $X = \varnothing$, then $\sum_{Y\subseteq X}\mu(Y)$ equals $\mu(\varnothing)$ and so, it equals 1 as required. □

Fact 3.2 $^X(A \subseteq X) \equiv \mu(A)\cdot(X = A)$.

Proof: Observe that $^X(A \subseteq X)$ is equivalent to

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

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

Now apply Fact 3.1. □

In particular, we have that $^X(x \in X) \equiv -(X = \{x\})$.

Lemma 3.3 $^{XX}\varphi(X) \equiv \varphi(X)$.

Proof: Observe that

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

The lemma follows from Fact 3.2. □

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 the converse also holds. 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 $> d$. The terminology is justified by the following observation. Suppose $\varphi(X)$ has degree d. Assign to every $x \in S$ a variable X_x and write the polynomial $2^{-s}\sum_{E}{}^{E}\varphi(E)\prod_{x\in E}\mathsf{X}_x$ in R. Then this polynomial has degree d. The value of function $\varphi(X)$ at X coincides with the value the polynomial assumes when we assign to X_x value 1 or 0 according to whether x is in X or not.

Fact 3.4 ${}^{X}[\varphi(X)\cdot\psi(X)] \equiv \sum_{A\cup B=X}\mu(A\cap B)\cdot{}^{A}\varphi(A)\cdot{}^{B}\psi(B)$.

Proof: Expressing φ and ψ through their transforms, we have

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

The fact follows from Lemma 3.3. □

The fact above generalizes easily to the operator $\prod_{x<t}$. We state it in the following fact.

Fact 3.5 ${}^{X}\prod_{x<t}\varphi(x,X) \equiv \mu(X)\sum_{\bigcup_{x<t}E_x=X}\prod_{x<t}\mu(E_x)\cdot{}^{E_x}\varphi(x,E_x)$.

From these facts we can give an upper bound to the degree of a product from the degree of the factors. For example, 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 3.6 *Let $\delta(X)$ be any function. The function $\delta(E\cap X)$ has degree $\le \|E\|$. 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 $\leq \|E\|$, compute ${}^X\delta(E \cap X)$ and check that it is 0 for $X \not\subseteq E$

$$\begin{aligned}
{}^X\delta(E \cap X) &\equiv \sum_{A \subseteq X} \mu(A)\delta(E \cap A) \equiv \sum_{B \subseteq X \setminus E} \sum_{A \subseteq X \cap E} \mu(A \cup B) \cdot \delta(A) \\
&\equiv \sum_{A \subseteq X} \mu(A)\delta(E \cap A) \equiv \sum_{A \subseteq X \cap E} \mu(A) \cdot \delta(A) \sum_{B \subseteq X \setminus E} \mu(B) \\
&\equiv \sum_{A \subseteq X \cap E} \mu(A) \cdot \delta(A) \cdot (X \subseteq E) \\
&\equiv (X \subseteq E) \sum_{A \subseteq X} \mu(A) \cdot \delta(A) \\
&\equiv (X \subseteq E)^X\delta(X).
\end{aligned}$$

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

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

Since $\{(E \subseteq X)\}_E$ is a base, the claim follows. To check the linear independence apply a cardinality argument. □

Let g be any element of R such that $(1-g)^{-1}$ exists in R. To prove Theorem 2.1 we shall need the fact that the functions $g^{\|E \cap X\|}$ form a base. We check that the hypothesis of the lemma is satisfied. It is easy to check that when $X = \varnothing$ then ${}^X\left(g^{\|X\|}\right)$ is 1. If instead $X \neq \varnothing$, pick an arbitrary $c \in X$,

$$\begin{aligned}
\sum_{E \subseteq X} \mu(E) \cdot g^{\|E\|} &= \sum_{E \subseteq X \setminus \{c\}} \mu(E) \cdot [g^{\|E\|} - g^{\|E\|+1}] \\
&= (1-g) \sum_{E \subseteq X \setminus \{c\}} \mu(E) \cdot g^{\|E\|}
\end{aligned}$$

Iterating this argument for all elements of X we can conclude that ${}^X\left(g^{\|X\|}\right)$ 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 3.2, 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 occurring in φ. By "approximating" we mean that for all but a small fraction of the sets X 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 is a field of characteristic > 0, the second is 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 2.1 and Corollary 6.2. The second proof is included for completeness and because of the general combinatorial ideas used there that make the method interesting in itself.

Theorem 4.1 *Let $\varphi(X)$ be a Σ_0^p formula and let α be a subset of the power set of S. There is a function $\psi(X)$ of degree $\leq a^n$ that is equal to $\varphi(X)$ for all but at most $2^{a(n-a)}\|\alpha\|$ sets $X \in \alpha$, where*

1. *n is proportional to the length of $\varphi(X)$, and*
2. *$a > n$ and $2^{a^{\frac{1}{n}}}$ is larger than S and of all the parameters occurring in $\varphi(X)$.*

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. In the statement of the theorem there is no claim on uniformity: namely, nothing is claimed on the syntax of the function ψ. Indeed the way one constructs ψ from atomic functions by means of the operator of products and sums is intimately connected with the syntax of φ. Some nonuniformity occurring in the construction results in the presence of some extra parameters. We shall not make this explicit.

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). Eliminate in φ all connectives but $\neg$, $\wedge$, and $\exists$. Also, for definiteness, replace the quantifiers of the form $(\exists x \in T)$ with $(\exists x < t)(x \in T) \rightarrow$, where t is an appropriate parameter. We can assume that $t < 2^a$, so we may assume that all quantifiers occurring in the formulas are of the form $(\exists x < t)$.

By Fact 3.1 and Fact 3.2, the atomic formulas $t = s$, $t < s$, and $t \in X$ have degree ≤ 1. 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 $< a^n$. We claim that the product $\psi_1 \cdot \psi_2$ is the required approximation of φ. From Fact 3.4 above it follows that $\psi_1 \cdot \psi_2$ has degree $< 2a^n$. The error probability is at most the sum of that of the two conjuncts separately, that is, $< 2 \cdot 2^{a(n-a)}$. The claim (with $n+1$ for n) follows from the induction hypothesis.

The relevant 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—has no chance. In fact, large products make us lose any control on the degree of the functions. We will express (though, in an approximate form) existential quantification using sums and *small* products: that is, products of the form $\prod_{x<a^k} \psi(x, X)$. Then, if for all $x < a^k$ the degree of $\psi(x, X)$ is less than a^n then the degree of $\prod_{x<a^k} \psi(x, X)$ is less than a^{n+k}.

For the expository reasons explained above, we first prove the theorem in a special case, that is, when R is a field of nonzero characteristic p. We need also assume that $p < a$. Let $t < 2^a$. Assume that, for all $x < t$, the function $\psi(x, X)$ of degree $< a^n$ is an approximation of $\varphi(x, X)$. We can assume that these approximations have error probability $< 2^{a(n-a)}$. Therefore, all but at most $2^a \cdot 2^{a(n-a)}\|\alpha\|$ sets $X \in \alpha$ are such that $\psi(x, X)$ equals $\varphi(x, X)$ for all $x < t$.

The following is the reason for a fixed X. Suppose that $(\exists x < t)\varphi(x, X)$ and fix $\hat{x}$ such that $\varphi(\hat{x}, X)$. Choose at random (with respect to the uniform distribution) a function l from $[0, t)$ into $[0, p-1)$. We show that, with probability at least $1/p$, we have that $\sum_{x<t} l(x)\varphi(x, X)$ is 1. In fact, for every choice of $l(x)$ for $x \neq \hat{x}$ there is

one choice (out of p) of $l(\hat{x})$ that makes the sum equal to 1. Now let h be a function from $[0, t) \times [0, pa^2)$ into $[0, p-1)$ obtained by choosing independently pa^2 times a function such as l as above. The probability that $\sum_{x<t} h(x, y)\varphi(x, X) \neq 1$ for every $y < pa^2$ is $(1-1/p)^{pa^2} \leq 2^{-a^2}$.

By counting, we conclude that for all X in α but at most $\left(2^{-a^2} + 2^{a(n+1-a)}\right)\|\alpha\|$ the formula $(\exists x < t)\varphi(x, X)$ holds if and only if for some $y < pa^2$,

$$\sum_{x<t} h(x, y)\varphi(x, X) \;=\; 1.$$

It follows that for these X the function $(\exists x < t)\varphi(x, X)$ equals

$$1 \;-\; \prod_{y<pa^2} \left[1 - \sum_{x<t} h(x, y)\varphi(x, X)\right]. \qquad (*)$$

This function has degree $< a^{n+3}$ and the error probability is $< 2^{a(n+2-a)}$. Therefore, the claim of the lemma is established for $(\exists x < t)\varphi(x, X)$ with $n+2$ for n. This completes the proof of the special case of the theorem.

Now we resume the general proof. We need to prove the inductive step for the existential quantifier in the case in which R has characteristic 0. The following lemma of Valiant and Vazirani [10] gives 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, that is, $(\exists z \leq a)(\exists! x \in C_z)\ \vartheta(x, X)$. 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 sets C_z are a suitable parameter depending on φ. They need to succeed for a 'large' fraction of the $X \in \alpha$.

Lemma 4.2 *Fix a formula of the form $(\exists x < t)\vartheta(x, X)$ where $t < 2^a$. Let α be an arbitrary set of subsets of S. There are some sets $C_z < t$, for $z = 0, \ldots, a$ such that for all but at most $(1/2)\|\alpha\|$ sets $X \in \alpha$,*

$$(\exists x < t)\vartheta(x, X) \longleftrightarrow (\exists z \leq a)(\exists!\, x \in C_z)\vartheta(x, X). \qquad (*)$$

Proof: Fix an arbitrary $\hat{X} \in \alpha$ such that $(\exists x < t)\vartheta(x, \hat{X})$. We shall see that, choosing C_z at random (with respect to the distribution specified below), we have that $(\exists z \leq a)(\exists!\, x \in C_z)\vartheta(x, \hat{X})$ with probability $> 1/2$. By counting, there are sets $C_0, \ldots, C_a$ that satisfy $(*)$ for all but at most $(1/2)\|\alpha\|$ sets $X \in \alpha$. The lemma follows.

Fix an arbitrary injection of the interval $[0, 2^a)$ into the set of binary strings of length a. We can view $\{0, 1\}^a$ as a vector space on the finite field $\{0, 1\}$. The canonical base for this vector space is denoted with $e_1, \ldots, e_a$. Below we shall identify numbers

$< 2^a$ and vectors; to simplify notation we assume that 0 corresponds to the zero of the vector space. The scalar product of two vectors $u, v \in \{0,1\}^a$ is defined in the natural way: it is 0 if the coordinates of u and v coincide on an even number of entries, 1 otherwise.

The sets C_z are constructed from a sequence of a mutually orthogonal vectors $v_0, \ldots, v_{a-1}$. Let $v_0 := 0$ (the zero of the vector space). If $v_0, \ldots, v_{z-1}$ are defined, let C_z be the set of vectors orthogonal to all $v_0, \ldots, v_{z-1}$. Let v_z be chosen at random in C_z (with uniform distribution). Observe that $C_0 = \{0,1\}^a$ and $C_a = \{0\}$.

The *rank* of a subset of $\{0,1\}^a$ is the minimal dimension of a subspace containing it. Let us write F for the set $\{x < t : \vartheta(x, \hat{X})\}$. We shall show that whatever the nonempty set F is, by choosing $C_0, \ldots, C_a$ at random as explained above we obtain that, with probability $> 1/2$, there is a $z \le a$ such that $C_z \cap F$ has cardinality 1, that is, $(\exists! x \in C_z)\vartheta(x, \hat{X})$ for some $z \le a$.

It suffices to prove that with probability greater than $1/2$ for some $z \le a$ the rank of $C_z \cap F$ is 1. 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 finished—or 0 belongs $C_z \cap F$, so since $C_a = \{0\}$, stage a will be successful.

Let $z \le a$ be arbitrary and let the rank of $C_z \cap F$ be $d > 1$. We claim that $C_{z+1} \cap F \ne \varnothing$ with probability greater than $1 - 2^{-d}$. To prove the claim, observe that the probability distribution is invariant under orthonormal transformations. We can assume that the vectors $v_1, \ldots, v_z$ are the base vectors $e_1, \ldots, e_z$, that $C_z \cap F$ contains the vectors $e_{z+1}, \ldots, e_{z+d}$ and these generate the subspace containing $C_z \cap F$. Now the vector v_{z+1} is chosen randomly among the vectors orthogonal to $e_1, \ldots, e_z$, that is, among the nonzero linear combinations of $e_{z+1}, \ldots, e_a$. We have

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

Therefore, $C^{z+1} \cap F$ 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 claim 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 lemma follows. □

Observe that applying the lemma a^2 times and concatenating the results, we can obtain that for some $\{C_z\}_{z<(a+1)a^2}$,

$$(\exists x < t)\varphi(x, X) \longleftrightarrow (\exists z < (a+1)a^2)(\exists! x \in C_z)\varphi(x, X)$$

holds for all $X \in \alpha$ but for at most $2^{-a^2}\|\alpha\|$. We shall use the lemma in this form.

Finally, we are ready to prove the inductive step for the existential quantifier. As in the special case above, using the *uniform* inductive hypothesis: let $t < 2^a$ and assume that, for all $x < t$, the function $\psi(x, X)$ of degree $< a^n$ is an approximation of $\varphi(x, X)$. We can assume that each approximation has error probability $< 2^{a(n-a)}$.

Therefore, all but at most $2^{a(n+1-a)}\|\alpha\|$ sets $X \in \alpha$ are such that $\psi(x, X)$ equals $\varphi(x, X)$ for all $x < t$. The function

$$1 - \prod_{z<(a+1)a^2} \left[1 - \sum_{x \in C_z} \psi(x, X)\right]$$

is the required approximation of $(\exists x < t)\varphi(x, X)$. Its degree is $< a^{n+4}$. The new error probability is $2^{-a^2} + 2^{a(n+1-a)} < 2^{a(n+2-a)}$. The claim of the lemma is established for $(\exists x < t)\varphi(x, X)$ with $n+4$ for n. This completes the proof of the induction step for the existential quantifier.

To complete the theorem we are left only to verify the claim on the size of n. Indeed n increases by at most 4 at the inductive steps for $\exists$ and $\wedge$. Though to apply the proof we may need to change the syntax of φ (i.e., to eliminate $\forall$, $\rightarrow$, etc.), the growth in size is limited by a fixed factor. The proof of Theorem 4.1 is complete. □

5 Proof of Theorem 2.1 We have now established all that is needed in order to prove Theorem 2.1.

Proof: Let α be the set of those $X \subseteq S$ such that the formula $\varphi(x, X)$ counts X modulo q. Fix a ring R that has a qth 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 requires 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

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

Therefore, every function is the sum of a function of degree at most $s/2$ and a function which is $g^{\|X\|}$ multiplied by a function of degree $\leq s/2$. Since g is a q-root of unity, the function $g^{\|X\|}$ coincides in α with the linear combination $\sum_{x<q} g^x \cdot \varphi(x, X)$. By Theorem 4.1, on a set of cardinality $< \left(1 - 2^{a(n-a)}\right)\|\alpha\|$ the function $g^{\|X\|}$ coincides with a function of degree a^n where n is fixed by the theorem. There is a submodule of dimension $> \left(1 - 2^{a(n-a)}\right)\|\alpha\|$ where every function has degree $< (s/2) + a^n$. We can derive the claimed bound on the cardinality of α from a simple argument of dimensionality. The functions of degree $< a^n$ are

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

(For the last inequality use Robbins's sharp form of Stirling's formula, see [2], II.9) Therefore,

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

and

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

Recall that $S<2^a$, so in particular $s<2^a$. If $a>n$ we obtain

$$\|\alpha\| \;<\; \left(1+\frac{1}{s}\right)2^{s-1}\left(1+\frac{2a^n}{\sqrt{s}}\right) \;<\; 2^{s-1}\left(1+\frac{5a^n}{\sqrt{s}}\right).$$

Assuming $a>5$, we conclude that

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

Theorem 2.1 follows. □

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 nesting generalized quantifiers) that $(Q_\tau x<t)\varphi(x)$ holds if and only if $\tau(X)$ holds when $X=\{x<t:\varphi(x)\}$. In the example considered below $\tau(X)$ is $X=_p 0$. When $p=2$, $(Q_\tau x<t)\varphi(x)$ spells out: there is an even number of $x<t$ that satisfy $\varphi(x)$. The class $\Sigma_0^p(Q_\tau)$ is defined as Σ_0^p but it is also closed under Q_τ. Theorem 4.1 can be easily extended to the following.

Theorem 6.1 *Let $\varphi(X)$ be a $\Sigma_0^p(Q_\tau)$ formula and let α be a subset of the power set of S. There is a function $\psi(X)$ of degree $\le a^n$, that is equal to $\varphi(X)$ for all but at most $2^{a(n-a)}\|\alpha\|$ sets $X\in\alpha$, where*

1. *n is proportional to the length of $\varphi(X)$, and*
2. *$a>n$ and $2^{a^{\frac{1}{n}}}$ is larger than S and of all the parameters occurring in $\varphi(X)$; moreover, ${}^X\tau(X)\neq 0$ for at most 2^a subsets X of S.*

Proof: We only need to prove the induction step of the proof of Theorem 4.1 for the generalized quantifier Q_τ. Consider the identity

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

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

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

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

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

where $\psi(E, X)$ is the approximation of degree $< a^n$ of the formula $(\forall x \in E)\varphi(x, X)$ as given by the induction hypothesis. Note that we can only claim that for each E the function $\psi(E, X)$ approximates $(\forall x \in E)\varphi(x, X)$ on some large set. This set may depend on E, but only the sets such that ${}^E\tau(E) \neq 0$ are relevant. Since ${}^E\tau(E) \neq 0$ for less than 2^a sets, we can assume that $(*)$ coincides with $(Q_\tau x < t)\varphi(x, X)$ for all but $2^{a(n+1-a)}$. This proves the theorem. □

Theorem 2.1 generalizes to the following.

Corollary 6.2 *Let p and q be two different prime numbers. Let $\varphi(x, X)$ be a $\Sigma_0^p(Q_{\|X\|=_p 0})$ formula and let S be a set. The formula $\varphi(x, X)$ counts X modulo q for at most $2^{s-1}\left[1 + (a^m/\sqrt{s})\right]$ subsets X of S, where*

1. *s is the cardinality of S,*
2. *m is proportional to length of $\varphi(x, X)$, and*
3. *a is any number which is larger than m and p and such that $2^{a^{\frac{1}{m}}}$ bounds S and all the parameters occurring $\varphi(x, X)$.*

Proof: Let F be a field of characteristic p with a qth root of the unity. The formula $\|X\| =_p 0$ is equivalent to

$$1 - \|X\|^{p-1} \equiv 1 - \left(\sum_{x \in S} x \in X\right)^{p-1},$$

by Facts 3.2 and 3.4, $\|X\|^{p-1}$ has degree $< p$. By Theorem 6.1 the proof of Theorem 2.1 remains valid (with F for R) when Σ_0^p is replaced with $\Sigma_0^p(Q_{\|X\|=_p 0})$. □

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 if we can sharp the bound in the error probability as p increases.

Acknowledgments This research was supported by the Netherlands Foundation for Scientific Research (NWO) grant PGS 22-262.

REFERENCES

[1] Ajtai, M., "Σ_1^1 formulas on finite structures," *Annals of Pure and Applied Logic*, vol. 24 (1983), pp. 1–48. MR 85b:03048 1

[2] Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. 1, 2d edition, J. Wiley and Sons, New York, 1957. Zbl 0077.12201 MR 19,466a 5

[3] Furst, M., J. B. Saxe, and M. Sipser, "Parity circuits and the polynomial-time hierarchy," *Mathematical Systems Theory*, vol. 17 (1984), pp. 13–27. Zbl 0534.94008 MR 86e:68048 1

[4] Håstad, J., "Almost optimal lower bounds for small depth circuits," pp. 143–70 in *Randomness and Computation*, vol. 5, *Advances in Computing Research*, edited by S. Micali, JAI Press, Greenwich, 1989. 1

[5] Razborov, A. A., "Lower bounds for the size of circuits of bounded depth with basis $\{\wedge, \oplus\}$," *Mathematical Notes of the Academy of Science of the USSR*, vol. 41 (1987), pp. 333–38. 1, 4

[6] Razborov, A. A., "Bounded arithmetic and lower bounds in Boolean complexity," pp. 344–86 in *Feasible Mathematics 2*, Birkhäuser, Boston, 1995. Zbl 0838.03044 MR 96d:03057 1

[7] Smolensky, R., "Algebraic methods in the theory of lower bounds for Boolean circuit complexity," pp. 77–82 in *Proceedings of the 19th Annual ACM Symposium on the Theory of Computing*, 1987. 1, 1, 1, 4

[8] Tauri, J., "Probabilistic polynomials, AC_0 functions, and the polynomial-time hierarchy," *Theoretical Computer Science*, vol. 113 (1993), pp. 167–83. 1, 4, 4

[9] Toda S., and M. Ogiwara, "Counting classes are at least as hard as the polynomial-time hierarchy," *SIAM Journal on Computing*, vol. 21 (1992), pp. 316–28. Zbl 0755.68055 MR 93h:68052

[10] Valiant, L. G., and V. V. Vazirani, "NP is as easy as detecting unique solutions," *Theoretical Computer Science*, vol. 47 (1986), pp. 85–93. Zbl 0621.68030 MR 88i:68021 4, 4

[11] Yao, A. C., "Separating the polynomial-time hierarchy by oracles," pp. 1–10 in *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, 1985. 1

Department of Mathematics and Computer Science
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
THE NETHERLANDS
email: domenico@fwi.uva.nl