PROOF COMPLEXITY IN ALGEBRAIC SYSTEMS AND BOUNDED DEPTH FREGE SYSTEMS WITH MODULAR COUNTING

S. Buss, R. Impagliazzo, J. Krajíček, P. Pudlák,
A. A. Razborov and J. Sgall

Abstract. We prove a lower bound of the form $N^{\Omega(1)}$ on the degree of polynomials in a Nullstellensatz refutation of the $Count_q$ polynomials over $\mathbf{Z}_m$, where q is a prime not dividing m. In addition, we give an explicit construction of a degree $N^{\Omega(1)}$ design for the $Count_q$ principle over $\mathbf{Z}_m$. As a corollary, using Beame *et al.* (1994) we obtain a lower bound of the form $2^{N^{\Omega(1)}}$ for the number of formulas in a constant-depth Frege proof of the modular counting principle $Count_q^N$ from instances of the counting principle $Count_m^M$.

We discuss the polynomial calculus proof system and give a method of converting tree-like polynomial calculus derivations into low degree Nullstellensatz derivations.

Further we show that a lower bound for proofs in a bounded depth Frege system in the language with the modular counting connective MOD_p follows from a lower bound on the degree of Nullstellensatz proofs with a constant number of levels of extension axioms, where the extension axioms comprise a formalization of the approximation method of Razborov (1987), Smolensky (1987) (in fact, these two proof systems are basically equivalent).

Introduction

A propositional proof system is intuitively a system for establishing the validity of propositional tautologies in some fixed complete language. The formal definition of *propositional proof system* is that it is a polynomial time function f which maps strings over an alphabet Σ *onto* the set of propositional tautologies (Cook & Reckhow 1979). Any string π for which $f(\pi) = \tau$ is called an f-proof of τ. Although this definition seems a little counterintuitive at first, it covers ordinary propositional proof systems T; namely, let f_T map π to τ if π is a valid proof of τ in the proof system T, otherwise map π to some fixed tautology.

The efficiency of a proof system f can be measured by the length-of-proofs function

$$s_f(\tau) := \min\{|\pi| : f(\pi) = \tau\},$$

where $|\pi|$ is the length of the string π. It is a fundamental open problem of mathematical logic and computational complexity theory to show that $s_f(\tau)$ cannot be bounded by a polynomial in the length of τ, for any f. By Cook & Reckhow (1979) this is equivalent to showing that

$NP \neq coNP$. Despite extensive research (see the expository articles Buss (1995b), Krajíček (1996), Pudlak (1995a), Krajíček (1997) or the monograph Krajíček 1995), no non-trivial lower bounds for s_f are known, even for the usual text-book proof systems based on a finite number of axiom schemes and inference rules (a *Frege system* in the established terminology).

The proof systems discussed in this paper can be arranged in two hierarchies; one consisting of traditional propositional proof systems (i.e., fragments of Frege systems) and the other of algebraic proof systems (i.e., fragments of the equational theory of rings). We present here an overview of these systems; formal definitions will be given later.

Fragments of Frege systems considered in this paper are:

Bounded depth Frege. This is a Frege system where the lines are limited to be constant-depth formulas over the basis of unrestricted fan-in AND, OR and NOT operations. This system corresponds to the circuit complexity class AC^0 (in the sense that lines in polynomial size bounded depth Frege proofs are representable by AC^0-formulae), which is known not to be able to express properties involving counting or modular counting.

Bounded depth Frege with counting axioms modulo m. This is bounded depth Frege with the addition of substitution instances of the family of tautologies $Count_m^N$ where $N \not\equiv 0 \pmod{m}$, which expresses the impossibility of partitioning a set of N objects into groups of size m if N is not divisible by m. This adds some ability to use reasoning involving counting modulo m, which is not possible for bounded depth Frege systems.

Bounded depth Frege with MOD_m gates. Here we add into the basis for bounded depth Frege a new connective MOD_m computing whether the sum of its inputs is divisible by m. This (apparently) gives more power for modular reasoning than just adding counting axioms modulo m since concepts defined in terms of modular counting can then be used to define yet more complicated concepts.

Unrestricted Frege Systems. Lines in an unrestricted Frege proof can be arbitrary Boolean formulas over any fixed complete basis, e.g., $\wedge, \neg$. Any Frege system can polynomially simulate any other Frege system (Reckhow 1976). Frege systems correspond to the complexity class NC^1 in the same sense as bounded depth Frege systems corresponded above to AC^0.

As we already mentioned, no superpolynomial lower bounds on the length-of-proofs function for unrestricted Frege systems are known. Such bounds, however, are known for resolution (Tseitin 1968, Haken 1985, Urquhart 1987, Chvatal & Szemeredi 1988, Bonet *et al.* 1995, Krajíček 1994b), bounded depth Frege, with and without counting axioms modulo m (Ajtai 1988, Bellantoni *et al.* 1992, Krajíček 1994a, Krajíček *et al.* 1995, Pitassi *et al.* 1993, Ajtai 1990, Beame & Pitassi 1993, Riis 1993, Ajtai 1994, Beame *et al.* 1994, Riis 1994), and for cutting planes (Bonet *et al.* 1995, Krajíček 1994b, Pudlak 1995b). The survey Urquhart (1995) discusses many of these superpolynomial lower bounds.

The second hierarchy comprises algebraic proof systems. Let Λ be a fixed (commutative) ring, usually $\mathbf{Z}_m$ (integers modulo m) for some integer m. Instead of adding modular reasoning to propositional logic, the hierarchy here starts with almost pure equational reasoning and increases the ability to do logical arguments involving "case analysis". In each of the following systems, we try to derive consequences from a system of polynomial equations $p_i(\vec{x}) = 0$ using equational reasoning. To enforce the variables having values 0 or 1, the systems always have as axioms the equations $x_i^2 - x_i = 0$ for each variable x_i.

Nullstellensatz (Beame *et al.* 1994). A Nullstellensatz proof of a polynomial g from polynomials $f_1, \ldots, f_k$ consists of polynomials p_i, r_j such that $\sum_i p_i f_i + \sum_j r_j(x_j^2 - x_j) = g$. In other words, the proof is concrete evidence that the ideal generated by given polynomials f_i and the axioms $(x_j^2 - x_j)$ contains g, and so every 0-1 solution to the system $f_i(\vec{x}) = 0$ is also a 0-1 solution to $g(\vec{x}) = 0$. Here we insist that each p_i, r_j is represented explicitly by its vector of coefficients, so that proof size is determined by the largest degree.

Polynomial calculus (sometimes also called "*Gröbner proofs*" (Clegg *et al.* 1996)). This is a proof system that allows equational reasoning by adding two previously deduced polynomials or multiplying a previously deduced polynomial by another polynomial. As above, polynomials must be explicitly given as vectors of coefficients, which again means that size is determined by the degree. This system obviously includes the Nullstellensatz system.

Nullstellensatz with extension axioms (introduced in this paper). Here we supplement the Nullstellensatz system with additional equations in new variables that implicitly define a low degree approximation to the product of a set of polynomials. A proof is then a Nullstellensatz refutation from the original equations and additional extension equations.

Equational logic. The same reasoning is allowed as in the polynomial calculus; the difference is that the polynomial equations deduced are allowed to be represented as algebraic formulas over the variables. To show that two representations are equal must be done explicitly in the proof, using the ring laws. Equational logic over any (fixed) finite field is equivalent to unrestricted Frege systems.

In this paper we do the following:

(1) Prove an exponential lower bound for constant-depth Frege proofs in the de Morgan language of the counting modulo q principle $Count_q^N$ from instances of the counting modulo m principle (non-existence of a polynomial size proof was shown in Ajtai (1994), Beame *et al.* (1994), Riis (1994)). Our proof is based on giving a lower bound on the degree of certain Nullstellensatz refutations.

(2) Investigate the relation between polynomial calculus proofs and Nullstellensatz proofs.

(3) Prove a tight connection between the lengths of constant-depth Frege proofs with MOD_p gates and the lengths of Nullstellensatz refutations with extension axioms.

These problems are motivated by three sources: (a) the line of research in complexity of proof systems conducted so far, (b) the relationship between provability in bounded arithmetic and short constant-depth Frege proofs, and (c) facts from Boolean complexity. We refer the reader to Krajíček (1995) for detailed explanation of the first two topics and here we remark only on the third one.

A lower bound for a proof system working with formulas from a class $\mathcal{C}$ seems to be possible only after a Boolean complexity lower bound is established for the class $\mathcal{C}$. It should be stressed, however, that there is no provable relationship between the two problems, so each particular case requires a new idea. For example, all lower bounds for bounded depth Frege systems and their extensions utilize techniques developed for proving lower bounds for the class AC^0 (in particular, extensions of the switching lemmas of Hastad (1989)). Since a lower bound for the class $AC^0[p]$ of constant-depth circuits with the MOD_p gate is known by Razborov (1987), Smolensky (1987), it is natural to try to modify this technique to obtain a lower bound for bounded depth Frege systems in the language with MOD_p. So far, no one has succeeded in this; nonetheless, the circuit complexity lower bounds provide hope for success in finding lower bounds for the analogous proof systems.

After the preliminary version of this paper circulated the following developments occured. Exponential lower bounds for constant-depth Frege proofs of the *onto* version of the pigeonhole principle PHP_n^m from instances of the counting modulo m principle $Count_m$ are shown in Beame & Riis (1996). They are obtained via a reduction to Nullstellensatz proofs for this principle similar to Theorem 2.4. For proving the Nullstellensatz bound, a substantially developed variant of the degree reduction technique from Section 3 is used. The first (unconditional) lower bounds for the polynomial calculus are established in Razborov (1996). Namely, it is shown there that every polynomial calculus proof of PHP_n^m must have degree $\Omega(n)$ over any ground field.

The paper is organized as follows. In Section 1 we recall the definitions of the system $F(MOD_a)$ and of the formulas $Count_a^N$. In Section 2 we define two algebraic proof systems, the Nullstellensatz system and the polynomial calculus system. The new lower bound $N^{\Omega(1)}$ for the Nullstellensatz proof system related to proofs of $Count_q^N$ is given in Section 3. It improves upon Beame *et al.* (1994) where a lower bound of the form $\omega(1)$ was proved. In Section 4 we discuss the use of designs for proving lower bounds for Nullstellensatz proofs, and give a second, design-based proof of the $N^{\Omega(1)}$ lower bound; in this section we also prove $\Omega(N)$ lower bound for the case of characteristic zero. In Section 5 we make several observations on Nullstellensatz proofs and polynomial calculus proofs: we prove an analogue of cut-elimination, we give a constructive proof of completeness and we prove that tree-like polynomial calculus proofs can be converted into Nullstellensatz proofs without a large increase in degree. A proof-theoretic modification of the approximation method of Razborov (1987), Smolensky (1987) is defined in Section 6. We show there that the logarithm of the length-of-proofs function s_T, where T is the system of bounded depth Frege with MOD_p gates is equivalent, up to a polynomial, to the minimal possible degree of polynomials required for a refutation in the Nullstellensatz system with constantly many levels of extension axioms.

1. The counting principles

Let F be a fixed Frege proof system in the language $\{TRUE, \neg, \vee\}$ ($\vee$ is binary, and other connectives $\wedge, \rightarrow$ and $\equiv$ are used as abbreviations for their defining formulas) with finitely many axiom schemes and modus ponens as the only rule of inference. We follow Krajíček (1995) in the definitions of $F(MOD_a)$ and $Count_a^N$.

Definition 1.1. *Let $a \geq 2$ be a fixed integer, and let $0 \leq i \leq a-1$. Then $MOD_{a,i}$ are propositional connectives of unbounded arity with the intuitive meaning*

$$MOD_{a,i}(p_1, \ldots, p_k) \text{ is true} \quad \text{iff} \quad |\{j : p_j \text{ is true}\}| \equiv i \pmod a.$$

We extend the Frege system F to the language including the $MOD_{a,i}$ connectives by adding the following axiom schemes (called the MOD_a axioms):

1. *$MOD_{a,0}(\emptyset)$, where $\emptyset$ is the empty set of formulas ($k = 0$),*

2. *$\neg MOD_{a,i}(\emptyset)$, for $i = 1, \ldots, a-1$,*

3. *$MOD_{a,i}(\phi_1, \ldots, \phi_k, \phi_{k+1}) \equiv [(MOD_{a,i}(\phi_1, \ldots, \phi_k) \wedge \neg\phi_{k+1}) \vee (MOD_{a,i-1}(\phi_1, \ldots, \phi_k) \wedge \phi_{k+1})]$ for $i = 0, \ldots, a-1$, where $k \geq 0$ and $i-1$ is taken modulo a.*

The system $F(MOD_a)$ is the system F whose language is extended by the connectives $MOD_{a,i}$, $i = 0, \ldots, a-1$, and which is augmented by the above axioms.

The *depth* of a formula ϕ is defined as follows:

1. If $\phi = TRUE$ or $\phi = p_i$ then $depth(\phi) := 0$.

2. If $\phi = (\neg\psi)$ then $depth(\phi) := depth(\psi) + 1$.

3. If $\phi = MOD_{a,i}(\psi_1, \ldots, \psi_k)$ then $depth(\phi) := \max_{1 \leq j \leq k} depth(\psi_j) + 1$.

4. If $\phi = \eta(\psi_1, \ldots, \psi_m)$, where $m \geq 2$, η is a formula containing no connective other than the disjunction, and $\psi_1, \ldots, \psi_m$ already do not begin with a disjunction then
$$depth(\phi) := \max_{1 \leq j \leq m} depth(\psi_j) + 1.$$

Notice that this is just the ordinary definition of the depth if we rewrite ϕ with unbounded arity disjunctions. The *depth* of a proof is the maximal depth of formulas appearing in the proof. The *size* of a proof is the number of inferences. This differs from the usual convention of defining the "size" of a Frege proof to equal the number of symbols in the proof as we did with the length-of-proof function in the introduction. However, the number of symbols in the proof certainly bounds from above the size as we have defined it, and the lower bounds we obtain on the size are therefore also lower bounds on the number of symbols. For Frege system and its bounded depth subsystems it holds that the minimal size of a proof of a formula is proportional to the minimal number of different formulas that must occur in any proof of the formula (see Buss (1995b) or Krajíček (1995), Lemma 4.4.6).

Next we define tautologies $Count_a^N$ over the basis $\{\neg, \vee\}$ that express a modulo a counting principle:

Definition 1.2. *Let $N \not\equiv 0 \pmod a$, $N \geq a$. For an a-element subset e of $[N] = \{1, \ldots, N\}$, let p_e be a propositional variable. Let $e \perp f$ denote the condition $e \neq f \wedge e \cap f \neq \emptyset$. The formulas $Count_a^N$ are formed from atoms p_e as follows:*

$$Count_a^N := \bigvee_{e \perp f} (p_e \wedge p_f) \vee \bigvee_{v \in [N]} \bigwedge_{e \ni v} \neg p_e.$$

(The notation "$e \ni v$" means the conjunction is taken over all e such that $v \in e$).

The intuitive meaning of $Count_a^N$ is that $[N]$ can not be partitioned into a-element subsets: $Count_a^N$ is false if and only if $\{e : p_e = 1\}$ is a partition of all N elements.

An *instance* of $Count_a^N$ is any formula obtained from $Count_a^N$ by substituting arbitrary formulas for the p_e's. Let $q_1, \ldots, q_k$ be all prime factors of a. By Beame *et al.* (1994), Theorem 1.2, the formula $Count_a^N$ follows from instances of formulas $Count_{q_i}^M$, and every $Count_{q_i}^N$ follows from instances of some $Count_a^M$, where *follows* means having polynomial size constant-depth F-proofs. Thus we may confine our attention to the provability of formulas $Count_q^N$ with q a prime.

The main result of Ajtai (1994), Beame *et al.* (1994), Riis (1994) is that constant-depth Frege proofs of $Count_q^N$ from instances of $Count_m^M$, where q is a prime not dividing m, require superpolynomial size. These bounds were barely superpolynomial; Theorem 3.4 below improves them to exponential.

2. Algebraic proof systems

Algebraic proof systems are proof systems which manipulate polynomials over a ring. By translating propositional formulas $\phi(p_1, \ldots, p_n)$ in the language $\{TRUE, \neg, \vee\}$ into algebraic expressions $\phi^*(x_1, \ldots, x_n)$ over $\mathbf{Z}$, an algebraic proof system can serve as a propositional proof system. The definition of ϕ^* is quite straightforward: we translate the truth-value $TRUE$ to 0, variables p_i to x_i, $\neg\phi$ to $1 - \phi^*$, and $\phi \vee \psi$ to $\phi^* \cdot \psi^*$ (for technical reasons we switch in this paper the customary roles of 0 and 1, both for inputs and outputs). Moreover, every formula ϕ in the language of $F(MOD_p)$, where p is a prime, translates into an algebraic term over $\mathbf{F}_p$ if we additionally translate $MOD_{p,i}(\phi_1, \ldots, \phi_k)$ to $(\phi_1^* + \cdots + \phi_k^* - (k-i))^{p-1}$ (cf. Smolensky 1987).

2.1. The Nullstellensatz system. A formula ϕ in the language of $F(MOD_p)$ is a *tautological consequence* of a set of formulas $\psi_1, \ldots, \psi_k$ iff in characteristic p the polynomial ϕ^* is identically zero on all 0-1 solutions to $\psi_1^* = \cdots = \psi_k^* = 0$. By Hilbert's Nullstellensatz this is equivalent to the property that ϕ^* is in the ideal

$$\langle \psi_1^*, \ldots, \psi_k^*, x_1^2 - x_1, \ldots, x_n^2 - x_n \rangle \subseteq \mathbf{F}_p[x_1, \ldots, x_n]$$

generated by polynomials $\psi_1^*, \ldots, \psi_k^*, x_1^2 - x_1, \ldots, x_n^2 - x_n$. This, in turn, is equivalent to the existence of polynomials $P_1, \ldots, P_k, R_1, \ldots, R_n \in \mathbf{F}_p[\bar{x}]$ such that the Nullstellensatz identity

$$\sum_i P_i \psi_i^* + \sum_j R_j (x_j^2 - x_j) = \phi^*$$

holds. We note that in the general version of Nullstellensatz it is necessary to work in an algebraically closed field and take a radical of the ideal of polynomials. In our special case it is not needed due to the presence of the polynomials $x_j^2 - x_j$. We give a direct proof of the relevant case of Hilbert's Nullstellensatz as Theorem 5.2 below.

We define the Nullstellensatz proof system for a general commutative ring mostly to cover the case of rings $\mathbf{Z}_m$ for composite m. This is still a sound proof system. However, in the presence of divisors of 0 it is complete only as a refutation system and not as a proof system.

Definition 2.1. *Let Λ be a commutative ring. A* Nullstellensatz proof *of a polynomial g from polynomials $f_1, \dots, f_k \in \Lambda[x_1, \dots, x_n]$ is a $(k+n)$-tuple of polynomials*

$$P_1, \dots, P_k, R_1, \dots, R_n$$

such that the identity

$$\sum_i P_i f_i + \sum_j R_j (x_j^2 - x_j) = g$$

holds in $\Lambda[x_1, \dots, x_n]$. The degree *of the proof is*

$$\max\left\{ \max_i(deg(P_i) + deg(f_i)),\ \max_j(deg(R_j) + 2) \right\}.$$

A Nullstellensatz refutation *of $f_1, \dots, f_k$ is a Nullstellensatz proof of some non-zero constant in the ring Λ from $f_1, \dots, f_k$.*

Note that the degree of a Nullstellensatz proof provides an upper bound on the maximal degree of all participating polynomials, and that this bound is tight when Λ is an integral domain. Accordingly, it is convenient to measure the size of a Nullstellensatz proof by its degree rather than by the total number of symbols. In particular, when our underlying ring Λ is a field, we can (and will) assume w.l.o.g. that all Nullstellensatz refutations are normalized so that the right-hand side is equal to 1.

The definition of degree is based on the degrees of polynomials P_i and R_j. However, for Nullstellensatz refutations one can ignore the degrees of the latter polynomials. More specifically, we have the following easy

Lemma 2.2. *Suppose that there is a Nullstellensatz proof*

$$P_1, \dots, P_k, R_1, \dots, R_n$$

of g from $f_1, \dots, f_k$, and let $d = \max_i(deg(P_i) + deg(f_i))$. Then there is another proof

$$P_1, \dots, P_k, R'_1, \dots, R'_n$$

of g from $f_1, \dots, f_k$ of degree at most $\max\{d, deg(g)\}$. In particular, if g is a constant (i.e., we have a Nullstellensatz refutation), then the degree of the new proof is at most d, and if $k = 0$ (i.e., g belongs to the ideal generated by $(x_j^2 - x_j)$), then this degree is at most $deg(g)$.

Proof. We start with the last claim. Suppose that g belongs to the ideal generated by polynomials $(x_j^2 - x_j)$ and that $g \neq 0$. Then g can not be multilinear and thus must contain a monomial of the form $x_j^2 x_{j_1} \ldots x_{j_r}$. Replacing this monomial by $x_j x_{j_1} \ldots x_{j_r}$, we obtain another polynomial g' and, obviously, $g - g'$ has a Nullstellensatz derivation from $x_j^2 - x_j$ of degree at most $deg(g)$. Since the sum of degrees of all monomials appearing in g' is strictly less than for g, we can argue by induction on this sum.

The general claim follows from this partial case applied to the polynomial $g - \sum_i P_i f_i$. □

Instead of translating the $Count_q^N$ tautologies into an algebraic term, it is more convenient to follow Beame *et al.* (1994) and represent (the negation of) $Count_q^N$ by the following set of polynomials.

Definition 2.3 (Beame *et al.* 1994). *Assume that $N \geq q$. An (N, q)-*polynomial system *is the following system of polynomial equations in variables x_e where e ranges over q-element subsets of $[N]$:*

$$(v) \qquad \textstyle\sum_{e \ni v} x_e = 1,$$

one for each $v \in [N]$, and

$$(e, f) \qquad x_e \cdot x_f = 0,$$

one for each $e \perp f$.

Let Q_v denote the polynomial $(\sum_{e \ni v} x_e) - 1$ and let $Q_{e,f}$ denote the polynomial $x_e \cdot x_f$.

Any solution of the (N, q)-system in any field must be necessarily a 0-1 solution and the set $\{e : x_e = 1\}$ would form a total partition of $[N]$ into q-element sets. Hence the (N, q)-system has no solution in any field if $N \not\equiv 0 \pmod{q}$. The link between Nullstellensatz refutations of $Q_v, Q_{e,f}$ and constant-depth F-proofs of $Count_q^N$ from instances of $Count_m^M$ (even for composite m) is the following theorem of Beame *et al.* (1994). We indicate how to modify the proof of the theorem as it is not stated there in exactly this form.

Theorem 2.4 (Beame *et al.* 1994). *Let $q \geq 2$ be a prime and $m \geq 2$ be an integer not divisible by q. Denote by $d(N)$ the minimum value d such that there is a Nullstellensatz refutation of polynomials $Q_v, Q_{e,f}$ over $\mathbf{Z}_m$ of degree d. Let $\ell \geq 1$ be a constant. Then there is $\epsilon > 0$ such that for almost all $N \not\equiv 0 \pmod{q}$ the following holds:*

> *In every depth ℓ Frege proof of $Count_q^N$ from instances of the formulas $\{Count_m^M : M \not\equiv 0 \pmod{m}\}$ at least $N^{\epsilon \cdot d(N^\epsilon)}$ different formulas must occur.*

Proof. Assume on the contrary that such proofs of $Count_q^N$ of size at most $N^{s(N)}$ exist for infinitely many $N \not\equiv 0 \pmod{q}$. By Theorem 3.5 and Lemma 4.5 of Beame *et al.* (1994) then $d(N) = O(s(N^{O(1)}))$. Hence for suitable $\epsilon > 0$, $\epsilon \cdot d(N^\epsilon) \leq s(N)$.

Theorem 3.5 in Beame *et al.* (1994) has an additional assumption that $s(N) = N^{o(1)}$. The only place in its proof where this is used is to fulfill the hypothesis of Lemma 3.7 that

$s(N) \leq N^{\epsilon_\ell}$, some $\epsilon_\ell > 0$ depending on ℓ only. Hence by choosing ϵ above small enough we get $\epsilon \cdot N^\epsilon \leq N^\epsilon \leq N^{\epsilon_d}$ as well. □

By the remarks after Definition 1.1, Theorem 2.4 immediately implies that any depth ℓ Frege proof of $Count_q^N$ from instances of the formulas $\{Count_m^M \ : \ M \not\equiv 0 \pmod{m}\}$ has also size at least $N^{\epsilon \cdot d(N^\epsilon)}$ for some $\epsilon > 0$.

2.2. The polynomial calculus. A basic proof system for ideal membership is *equational logic* based on axioms of commutative rings with identity and of a given characteristic. The lines in an equational proof are equations between terms in the language of rings with constants for all elements of the ground ring. The axioms are instances of the ring identities and identity laws; the inference rules of equational logic are

$$\frac{t_1 = s_1 \ \ldots \ t_k = s_k}{f(\bar{t}) = f(\bar{s})},$$

one for every k-ary function symbol f. In the ring language these are just the two rules

$$\frac{t_1 = s_1 \quad t_2 = s_2}{t_1 + t_2 = s_1 + s_2} \quad \text{and} \quad \frac{t_1 = s_1 \quad t_2 = s_2}{t_1 \cdot t_2 = s_1 \cdot s_2}.$$

Note that it is important to distinguish between an axiom and an axiom schema. E.g. consider a group presented by a finite set of equations E. Then we consider not only the equations from E, but also all instances of the axiom schemas of group theory, i.e., instances of the associativity law and laws for inverse elements.

The *size* of an equational logic proof is defined to equal the number of inferences in the proof. Equational logic over any fixed field $\mathbf{F}_p$ is polynomially equivalent with full Frege systems, meaning that a proof in one system can be translated into a proof in the other system with only a polynomial increase in size. It is proved in Reckhow (1976) that all Frege proof systems in all complete languages are polynomially equivalent (equational logic is just one of them).[1] In fact, also bounded depth subsystems of $F(MOD_p)$ and bounded depth subsystems of equational logic over $\mathbf{F}_p$ (measuring the depth of terms) correspond to each other.

Equational logic is thus too strong and we consider also a weaker proof system.

A *monomial* is a product of variables, and a *polynomial* in this context is a linear combination of monomials (over the underlying ring). We introduce this definition to distinguish polynomials as terms of a particular form. Of course, a term is equivalent to a polynomial; the only difference is that a term is not explicitly written out as a linear combination of monomials, and that the latter expression might be substantially longer than the original term.

[1] Reckhow actually proved the polynomial equivalence of all Frege systems when proof size is measured in terms of the number of symbols in the proof; however, his results hold also for proof size measured in terms of the number of inferences.

The polynomial calculus is an algebraic proof system in which each line of the proof is a polynomial. This system has also been called the *sequential ideal generation system* (Impagliazzo 1995) and the *Gröbner proof system* (Clegg *et al.* 1996).

Definition 2.5. *A* polynomial calculus proof *is a directed acyclic graph; each line (node) in the proof is a polynomial over a fixed ring (e.g., $\mathbf{F}_p$). The rules of inference are:*

$$\frac{f \quad g}{f+g} \quad \text{and} \quad \frac{f}{f \cdot g},$$

where f and g are arbitrary polynomials. These two rules are called addition *and* multiplication, *respectively. For polynomials $f_1, \ldots, f_k, g$, a* polynomial calculus proof *of g from $f_1, \ldots, f_k$ is a proof in which initial polynomials are among $f_1, \ldots, f_k$ and the final polynomial is g. The* degree *of the proof is the maximum of the degrees of the polynomials appearing in the proof.*

An interesting observation about the polynomial calculus was made in Impagliazzo (1995), Clegg *et al.* (1996). It is shown there that for any fixed d there is a polynomial time algorithm deciding whether g has a degree d proof from $f_1, \ldots, f_k$, and if such a proof exists the algorithm constructs it.

Despite various syntactic conditions in the definition of the polynomial calculus, the property of having a low degree proof can be characterized in a syntax-free manner, at least for the case when the underlying ring is $\mathbf{F}_p$. The characterization uses the notion of *semantic derivations* introduced in Krajíček (1994b).

For a polynomial $f \in \mathbf{F}_p[x_1, \ldots, x_n]$ denote by $V(f)$ the variety

$$\{\bar{a} \in \{0,1\}^n : f(\bar{a}) = 0\}.$$

Theorem 2.6. *Let $f_1, \ldots, f_k, g \in \mathbf{F}_p[\bar{x}]$, and let d_1, d_2 be the minimal numbers that are greater or equal than $\max\{\max_i deg(f_i),\ deg(g)\}$ and have the following two properties.*

1. *There is a degree d_1 polynomial calculus proof of g from $f_i, (x_j^2 - x_j)$.*
2. *There is a sequence $V_1, \ldots, V_\ell$ of subsets of $\{0,1\}^n$ such that*
   (a) *Every V_t is either $\{0,1\}^n$, or one of $V(f_i)$, or $V_t \supseteq V_r \cap V_s$ for some $r, s < t$.*
   (b) *$V_\ell = V(g)$.*
   (c) *Every V_t has the form $V(h)$ for some polynomial h of degree at most d_2.*

Then

$$d_2 \leq d_1 \leq (2p-1)d_2.$$

Proof. Given a polynomial calculus proof, replace every polynomial h in the proof with the set $V(h)$. This shows $d_2 \leq d_1$.

In the opposite direction assume that $V_1, \dots, V_\ell$ satisfy the assumptions above; in particular, let $V_t = V(h_t)$ for some degree d_2 polynomial h_t. W.l.o.g. we can assume that $h_t = 0$ if $V_t = \{0,1\}^n$, $h_t = f_i$ if $V_t = V(f_i)$, and $h_\ell = g$. We show by induction on t that h_t can be derived from $f_i, (x_j^2 - x_j)$ by a polynomial calculus proof of degree at most $(2p-1)d_2$.

This is clear when V_t is one of the initial sets, so suppose that $V_t \supseteq V_r \cap V_s$ for some $r, s < t$. First we derive from h_r, h_s the polynomials $g_r = h_r^{p-1}$ and $g_s = h_s^{p-1}$: they represent the same sets V_r, V_s but have the additional property that they are 0-1 valued. Then we derive

$$h'_t = h_t(1 - (1-g_r)(1-g_s)) = h_t(g_r + g_s - g_r g_s).$$

Since $V_t \supseteq V_r \cap V_s$, $h'_t(\bar{a}) = h_t(\bar{a})$ for every $\bar{a} \in \{0,1\}^n$. Hence $h'_t - h_t$ belongs to the ideal generated by $(x_j^2 - x_j)$ and by Lemma 2.2 can be derived from them via a Nullstellensatz proof (which can be treated as a special case of a polynomial calculus proof) of degree at most $deg(h'_t) \leq (2p-1)d_2$. □

3. Improved lower bound for Nullstellensatz

In this section we prove a lower bound of the form $N^{\Omega(1)}$ on the degree of Nullstellensatz refutations of the (N,q)-system over $\mathbf{Z}_m$, $N \not\equiv 0 \pmod{q}$, q a prime not dividing m. This improves upon Beame *et al.* (1994) where it was shown that the degree of such refutations cannot be bounded by a constant independent of N. As a consequence, we improve the barely superpolynomial lower bound for bounded depth Frege with counting axioms from Ajtai (1994), Beame *et al.* (1994), Riis (1994) to exponential (Theorem 3.4 below).

We note that for this application we only need to prove the bound on the degree for fields $\mathbf{F}_p$. This is seen as follows: as we noted at the end of Section 1, instances of the $Count_m$ principle follow from instances of the $Count_p$ principles, p all prime divisors of m. By Beame *et al.* (1994), Theorem 3.5, to show a lower bound for constant-depth Frege proofs of $Count_q$ from these counting principles $Count_p$ it is enough to refute the existence of small (p,q,ℓ,M)-generic systems (parameters ℓ, M as in that theorem) for all primes p dividing m. The non-existence of such generic systems follows by Beame *et al.* (1994), Lemma 4.5 from lower bounds for the Nullstellensatz refutations of the (N,q)-system over all $\mathbf{F}_p$. But since our technique (as well as the one from Beame *et al.* 1994) for proving lower bounds on the degree of Nullstellensatz refutations applies with the same success to arbitrary rings $\mathbf{Z}_m$, we consider at once this more general case.

First we should introduce some more notation. Let I_N denote the ideal generated in $\mathbf{Z}_m[\bar{x}_e]$ by all $Q_{e,f}$ with $e \perp f$, and all $(x_e^2 - x_e)$. A monomial $x_{e_1} \cdots x_{e_t}$ in variables of the (N,q)-system will be abbreviated by x_E, where $E = \{e_1, \dots, e_t\}$ (strictly speaking, E can be a multiset, but in the presence of axioms $(x_e^2 - x_e)$ this is inessential).

Now we recall (and slightly modify) a lemma saying that for Nullstellensatz refutations of the (N,q)-system the relevant monomials occur only as coefficients of axioms Q_v and, moreover, correspond to partial q-partitions not containing v (we already saw in Lemma 2.2 that the degrees of R_j are inessential for general Nullstellensatz proofs). We give a direct combinatorial proof. However, we note the lemma corresponds to the algebraic fact that the factor ring $\mathbf{Z}_m[\bar{x}_e]/I_N$ is a free $\mathbf{Z}_m$-module, and the monomials x_E, where E runs over all partial q-partitions of $[N]$ (i.e., all $e \in E$ are pairwise different and disjoint), form a basis of this module.

Lemma 3.1 (Beame *et al.* 1994). *Assume that*

$$\sum_{v \in [N]} P_v \cdot Q_v = c \ (\mathrm{mod}\ I_N),$$

where $P_v \in \mathbf{Z}_m[\bar{x}_e]$, $d = \max_v deg(P_v) + 1$ and $c \in \mathbf{Z}_m$ is a constant. Then:

1. *there exist polynomials P'_v, with $deg(P'_v) \leq d-1$ such that if a monomial x_E occurs in P'_v with a non-zero coefficient then E is a partial q-partition of $[N] - \{v\}$ and*

$$\sum_{v \in [N]} P'_v \cdot Q_v = c \ (\mathrm{mod}\ I_N),$$

2. *in addition there exist polynomials $P'_{e,f}$ with $deg(P'_{e,f}) \leq d-2$ such that in $\mathbf{Z}_m[\bar{x}_e]$*

$$\sum_{v \in [N]} P'_v \cdot Q_v + \sum_{e \perp f} P'_{e,f} \cdot Q_{e,f} = c.$$

Proof. Let P'_v be the result of first removing multiple edges from every E, x_E a monomial of P_v, and then removing from P_v all monomials x_E where E is not a partition or $v \in \bigcup E$. Then

$$\sum_{v \in [N]} P'_v \cdot Q_v = \sum_{v \in [N]} P_v \cdot Q_v \ = c \ (\mathrm{mod}\ I_N),$$

since for $v \in e$

$$x_e Q_v = (x_e^2 - x_e) + \sum_{\substack{f \ni v \\ f \neq e}} x_e x_f .$$

To prove the second part, note that $\sum P'_v \cdot Q_v - c$ is multilinear. Express it as a linear combination of polynomials $Q_{e,f}$ and $x_e^2 - x_e$. Since $Q_{e,f}$ are multilinear monomials, we can make their linear combination multilinear at the expense of changing the coefficients of $x_e^2 - x_e$. Now the coefficients of $x_e^2 - x_e$ have to be 0 as their ideal contains no non-zero multilinear polynomial. Finally, obtain polynomials $P'_{e,f}$ from the coefficients of $Q_{e,f}$ by removing all monomials of degree at least $(d-1)$; they have to cancel since $Q_{e,f}$ are homogeneous of degree 2. □

Theorem 3.2. *Let $q \geq 2$ be a prime not dividing a (fixed) integer m. Any Nullstellensatz refutation of the (N,q)-system over $\mathbf{Z}_m$ for $N \not\equiv 0 \ (\mathrm{mod}\ q)$ must have degree at least $\Omega\left(N^{1/\log_2(m+q)}\right)$.*

The proof of Theorem 3.2 is based on the following lemma:

Lemma 3.3. *Suppose we have a Nullstellensatz refutation of degree $2d$ for the $((m+q)N, q)$-system over $\mathbf{Z}_m$, where q, m are as above and $N \not\equiv 0 \pmod{q}$. Then there exists a Nullstellensatz refutation of degree d for the (N, q)-system over $\mathbf{Z}_m$.*

Proof of Theorem 3.2 from Lemma 3.3. Lemma 3.3 immediately implies that (N, q)-system requires refutations of degree $2^i = \Omega\left(N^{1/\log_2(m+q)}\right)$ for $N = (m+q)^i$, $i \geq 1$ (the base case is just the fact that there is no degree one refutation of (N, q)-system over $\mathbf{Z}_m$ for $N > q$, $N \not\equiv 0 \pmod{q}$). To bridge the gaps between the values of N which are not of this form we make a restriction corresponding to a partial q-partition (i.e., we substitute 0's and 1's for all variables involving some vertices); thus we obtain a refutation of (M, q)-system from a refutation of (N, q)-system for $M < N$, $M \equiv N \pmod{q}$, with no increase of the degree. □

Proof of Lemma 3.3. By Lemma 3.1 it is equivalent to reduce any Nullstellensatz refutation of $((m+q)N, q)$-system of degree $2d$ modulo $I_{(m+q)N}$ to a Nullstellensatz refutation of (N, q)-system of degree d modulo I_N.

Let S be a domain of $(m+q)N$ vertices. The variables x_e of the Nullstellensatz refutation are indexed by unordered q-tuples $e \subseteq S$. Suppose that

$$\sum_v P_v \cdot Q_v = c \pmod{I_{(m+q)N}},$$

where $deg(P_v) \leq 2d - 1$, $c \in \mathbf{Z}_m$ and $c \neq 0$. By Lemma 3.1, we may assume that all monomials in P_v are indexed by partial q-partitions E of $S - \{v\}$.

Fix a permutation *Next* on S all of whose cycles have length $m+q$ (equivalently, partition the vertices into N cyclically ordered blocks of size $m+q$ each). For a given vertex v, let $v^0 = v$, $v^{i+1} = \mathit{Next}(v^i)$, …, so that $v^{m+q} = v^0$. Let $v^{-i} = v^{m+q-i}$ so that $(v^{-i})^i = v^0$. Let $\mathit{Orbit}(v) = \{v^1, v^2, \ldots, v^{m+q}\}$. For $V \subseteq S$, we denote $V^i = \{v^i : v \in V\}$ and $\mathit{Orbit}(V) = \bigcup_{v \in V} \mathit{Orbit}(v)$. Intuitively, the orbit of a set of vertices contains all the elements of the blocks intersecting the set.

Given the refutation on S of degree $2d$ we will construct a Nullstellensatz refutation on a set T with N elements of degree d. The elements $w \in T$ are indexed by the cycles of *Next*. Respectively, the Nullstellensatz variables $x_{e'}$ over T are indexed by (unordered) sets e' of q such cycles.

Now V will range over all selections of one element from each cycle, i.e., sets satisfying $|V| = N$, $\mathit{Orbit}(V) = S$. For each V we define a function ρ_V mapping each q-tuple $e \subseteq S$ to a q-tuple $e' \subseteq T$, 0, or 1. This naturally extends to a unique homomorphism ρ_V mapping the polynomials over (the variables defined by q-tuples in) S to polynomials over T, and we will make sure that ρ_V maps the ideal $I_{(m+q)N}$ into the ideal I_N corresponding to the domain T.

The mapping ρ_V is defined as follows. If $e \subseteq V^j$ for some $j \in \{1, \ldots, m\}$, then $\mathit{Orbit}(e)$ consists of q cycles and we put $\rho_V(e) = \mathit{Orbit}(e)$; call such edges *cross-edges*. If

$e = \{v^{m+1}, \ldots, v^{m+q}\}$ for some $v \in V$, then put $\rho_V(e) = 1$; call such edges *inner-edges*. Otherwise put $\rho_V(e) = 0$.

Inspection shows that for every V and v,

$$\rho_V(Q_v) = \begin{cases} Q_{Orbit(v)} & \text{if } v \in V^1 \cup \ldots \cup V^m \\ 0 & \text{if } v \in V^{m+1} \cup \ldots \cup V^{m+q}, \end{cases}$$

and for every V and $e \perp f$, either $\rho_V(Q_{e,f}) = 0$ or $\rho_V(Q_{e,f}) = Q_{\rho_V(e),\rho_V(f)}$. The latter fact in particular implies $\rho_V(I_{(m+q)N}) \subseteq I_N$. Hence for every particular V we have a Nullstellensatz refutation

$$\rho_V\left(\sum_v P_v \cdot Q_v\right) = c \pmod{I_N},$$

and summing over all V, we have

$$\sum_V \rho_V\left(\sum_v P_v \cdot Q_v\right) = \sum_w P'_w \cdot Q_w = c(m+q)^N \pmod{I_N}$$

with w running over all vertices in T. Notice that this is a Nullstellensatz refutation as $(m+q)^N$ is invertible in $\mathbf{Z}_m$ and, therefore, the right-hand side does not equal 0. We need to prove that this equation modulo I_N can be rearranged so that $\deg(P'_w) \leq d-1$. Hence it is sufficient to show that for every v and every monomial x_E in P_v either

(A) for every V, $\rho_V(x_E \cdot Q_v)$ can be written modulo I_N as $P \cdot Q_{Orbit(v)}$, for some P with $deg(P) \leq d-1$, or

(B) $\sum_V \rho_V(x_E \cdot Q_v) = 0$ (in which case x_E can be simply removed from P_v without changing the value of the sum).

Suppose that $\rho_W(x_E \cdot Q_v) \notin I_N$ for some W. Then in particular (cf. the proof of Lemma 3.1) x_E may not contain two cross-edges (w.r.t. W) with intersecting but different orbits, or an edge whose orbit contains v.

If E contains cross-edges with only $d-1$ different orbits, then $\rho_V(x_E)$ has only $d-1$ distinct variables, and (A) is satisfied.

In the remaining case E contains d cross-edges with disjoint orbits, and each other edge in E intersects an orbit of at most one of them (for inner edges this is trivial). In this case there exists a cross-edge $e \in E$ whose orbit is disjoint from the orbits of all other edges in E and also does not contain v. Fix such a cross-edge e in E. We prove that (B) is satisfied, i.e., $\sum_V \rho_V(x_E \cdot Q_v) = 0$.

Every V can be written as a disjoint union of $X = Orbit(e) \cap V$ and $Y = V \setminus Orbit(e)$. If $X^j = e$ for some $j \in \{1, \ldots, m\}$, then $\rho_{X\cup Y}(e) = Orbit(e)$, otherwise $\rho_{X\cup Y}(e) = 0$. Thus for a fixed Y, $\rho_{X\cup Y}(x_E)$ has the same value for $X = e^{-1}, \ldots, e^{-m}$, and is 0 otherwise. The same is true for $\rho_{X\cup Y}(x_E \cdot Q_v)$, since $v \notin Orbit(e)$, and hence $\rho_{X\cup Y}(Q_v)$ does not depend on

X. Thus every non-zero term $\rho_V(x_E \cdot Q_v)$ will actually get counted m times, which leads to the calculation

$$\sum_V \rho_V(x_E \cdot Q_v) = \sum_Y \sum_X \rho_{X\cup Y}(x_E \cdot Q_v) = \sum_Y \sum_{j=1}^m \rho_{(e^{-j})\cup Y}(x_E \cdot Q_v) \equiv 0 \pmod{m}.$$

This finishes the proof that the degree of the new Nullstellensatz proof modulo I_N is at most d. □

Theorems 2.4, 3.2, and the remarks on the reducibility of $Count_q^N$ to $Count_a^N$ made at the end of Section 1, have the following corollary.

Theorem 3.4. *Let $m, n \geq 2$ be fixed integers such that there is a prime factor q of n which is not a prime factor of m. Then the size of any constant-depth F-proof of $Count_n^N$ from instances of the axiom schema $Count_m^M$ is at least $2^{N^{\Omega(1)}}$.*

4. Design-based lower bounds

We now introduce the combinatorial notion of designs (cf. Beame *et al.* 1995, Clegg *et al.* 1996) which can be used for proving lower bounds on the degree of Nullstellensatz proofs. We'll give two applications of designs: first, we give an alternative proof of Theorem 3.2, and second, we give a linear lower bound for the degree of Nullstellensatz refutations of the (N, q)-systems in characteristic zero.

Definition 4.1. *Let Λ be any commutative ring, and let $N \geq q \geq 2$ be arbitrary. Let $\mathcal{M}_{N,q}$ be the set of all partial q-partitions of $[N]$. An* (N, q)-design *of degree d over Λ is a function*

$$s : \mathcal{M}_{N,q} \to \Lambda$$

such that:

1. *$s(\emptyset) = 1$.*

2. *For any E with less than d classes and for any $v \notin \bigcup E$:*

$$s(E) = \sum_{\substack{e \ni v \\ e \cap (\cup E) = \emptyset}} s(E \cup \{e\}).$$

Lemma 4.2. *Let $N \geq q \geq 2$ be such that $N \not\equiv 0 \pmod{q}$, and let $\frac{N}{q} > d \geq 1$ be arbitrary.*

If there is an (N, q)-design of degree d over some ring Λ, then the (N, q)-system has no Nullstellensatz refutation over Λ of degree at most d.

Proof. Suppose P_v are polynomials of degree at most $(d-1)$ satisfying property 1 of Lemma 3.1. Assume

$$P_v = \sum_E a_{E,v} x_E,$$

where E ranges over partial q-partitions of $[N]$, $a_{E,v} \in \Lambda$. Summing the coefficients in the conclusion of Lemma 3.1 implies

1. $\sum_v a_{\emptyset,v} = -c$,
2. $\sum_{v \notin (\cup E)} a_{E,v} = \sum_{e\in E} \sum_{w \in e} a_{E\setminus\{e\},w}$, for $1 \leq |E| \leq d$.

Define the function

$$A_i := \sum_{|E|=i} s(E) \cdot (\sum_{v\notin(\cup E)} a_{E,v}).$$

Then

$$A_0 = s(\emptyset) \cdot (-c) = -c$$

and

$$A_{i+1} = \sum_{|E|=i+1} s(E) \cdot (\sum_{v\notin(\cup E)} a_{E,v}) = \sum_{|E|=i+1} s(E) \cdot (\sum_{e\in E}\sum_{w\in e} a_{E\setminus\{e\},w}) =$$
$$\sum_{|F|=i}\sum_{w\notin(\cup F)}\sum_{\substack{e\ni w\\ e\cap(\cup F)=\emptyset}} s(F\cup\{e\}) \cdot a_{F,w} = \sum_{|F|=i}\sum_{w\notin(\cup F)} s(F)\cdot a_{F,w} = A_i.$$

For this computation we use property 2 of Definition 4.1 for F, i.e., we need $i < d$. So, $A_d = -c \neq 0$. But that contradicts the fact that $deg(P_v) \leq d-1$ and so $a_{E,v} = 0$ for all $|E| \geq d$, which would imply that $A_d = 0$. □

We note that if Λ is a field, the converse statement to the last lemma also holds. In fact, this is a special case of a general transformation (essentially a linear algebra duality) which can be applied to any Nullstellensatz system.

We now give a re-proof of Theorem 3.2 by constructing appropriate designs. Fix m and q as in the previous section. We construct the design by a similar construction as we used in the first proof of Theorem 3.2. By a slight alteration of this proof we are able to carry the induction step from d to $2d+1$ rather than $2d$, however this only improves the multiplicative constant in the final result, not the exponent.

Lemma 4.3. *Let $N \geq 1$ and $N \not\equiv 0 \pmod q$. If there is a design of degree d on N nodes, then there is a design of degree $2d+1$ on $(m+q)N$ nodes.*

Proof. For this proof we use the notation of the proof of Lemma 3.3. Let s be a degree d design on T. We can assume w.l.o.g. that $s(E) = 0$ whenever $|E| > d$. Let V as before be a selection of nodes, one from each block. We define a degree d design s^V on S by the following construction.

First we define a partial mapping $V(E)$ from partial q-partitions on S to partial q-partitions on T. If all edges in E are either cross-edges or inner-edges, and, moreover, the set $\{\mathit{Orbit}(e) : e \text{ is a cross-edge in } E\}$ is a partial partition, we let $V(E)$ equal this partition. In all other cases $V(E)$ is undefined. In other words, $V(E)$ is defined if and only if $\rho_V(x_E) \notin I_N$, and in that case $\rho_V(x_E) = x_{V(E)}$.

Next, we let

$$s^V(E) = \begin{cases} s(V(E)) & \text{if } V(E) \text{ is defined} \\ 0 & \text{otherwise.} \end{cases}$$

Claim 1: *s^V is a degree d design on S. Furthermore, suppose that for some E and $v \notin \bigcup E$ at least one of the following conditions holds:*

1. $|V(E)| < d$ or $V(E)$ is undefined,

2. v is in a block which intersects a cross-edge of E.

Then the design condition holds for v and E.

The claim is easily proved by inspection. □

Definition 4.4. *s^+ is defined by summing all possible s^V's as follows. Let $c = (m+q)^N$ (the number of possible V's). Then*

$$s^+(E) \;=\; c^{-1} \sum_V s^V(E),$$

where c^{-1} is the multiplicative inverse of c modulo m.

Claim 2: *Suppose that there is a cross-edge e in E such that no other edge in E intersects any of the q blocks that e intersects. Then $s^+(E) = 0$.*

To prove the claim, we argue by symmetry. Namely, for each V such that $V(E)$ is defined, there are exactly $m-1$ other sets W of representatives agreeing with V on the blocks which e does not intersect, and such that $W(E)$ is also defined. By symmetry, $s^W(E) = s^V(E)$ for all these W's. Summing these m values therefore equals 0 modulo m. Since this can be done for every V, $s^+(E) = 0$. □

Claim 3: *s^+ is a degree $2d+1$ design on S.*

Proving the claim will suffice to prove Lemma 4.3. The condition for $s_\emptyset$ is satisfied by our choice of c.

Consider any partial partition E on S of degree at most $2d$ and any v not in any member of E. W.l.o.g. we can assume that $V(E)$ is defined for at least one V, and we (arbitrarily) choose one such V. In the following cases the design condition is satisfied by the previous claims:

(1) If the block of v intersects some cross-edge of E, then s^+ satisfies the design condition for v and E since it is a linear combination of designs that satisfy this design condition by Claim 1.

(2) If $|V(E)| < d$, then the same is true for all W's for which $W(E)$ is defined, and the design condition is again satisfied by Claim 1.

(3) If the block of v intersects some inner-edge (and hence contains it), then we show that there exists an edge satisfying the condition of Claim 2 or one of (1) and (2) holds. If (2) is false, E has at least d cross-edges with non-intersecting blocks. Blocks of each other edge intersect with at most one of the d cross-edges (otherwise $V(E)$ is not a partition), and since E has at most $2d$ edges, either (1) holds or we use Claim 2 to conclude that the design condition is satisfied.

It remains to consider the situation when the block containing v does not intersect any edge of E, and $V(E)$ has exactly d edges for each V (this is the case not covered by the previous proof). Let W range over all selections of representatives from the blocks not containing v, and let $W_i = W \cup v^i$. By symmetry, $s^{W_i}(E) = s^W(E)$ is independent of i. Therefore,

$$s^+(E) = \sum_W \sum_{i=1}^{m+q} s^{W_i}(E) = (m+q)\sum_W s^W(E).$$

Define $e_j = \{v^{-j}, v^{-j+1}, \ldots, v^{q-j-1}\}$ for $0 \le j < q$; note that by our assumption on the design s, these q edges are the only edges $e \ni v$ for which $s^{W_i}(E \cup \{e\})$ might be non-zero, since otherwise $W_i(E \cup \{e\}) > d$ (if defined at all). Moreover, $s^{W_i}(E \cup \{e_j\}) = s^W(E)$ if $i = -m-1-j$ (i.e., when e_j is an inner-edge of W_i) and 0 otherwise. Thus,

$$\sum_{j=0}^{q-1} s^+(E \cup \{e_j\}) = \sum_W \sum_{i=1}^{m+q} \sum_{j=0}^{q-1} s^{W_i}(E \cup \{e_j\}) = q \sum_W s^W(E)$$

which is equal to $s^+(E)$ modulo m.

That completes the proof of the claim and Lemma 4.3. □

Lemma 4.3 immediately implies the following theorem, which, in turn, implies Theorems 3.2 and 3.4 by Lemma 4.2:

Theorem 4.5. *For any $N > 0$, there is a degree $d = \Omega\left(N^{1/\log_2(m+q)}\right)$ design on $[N]$.*

Proof. By induction on N. The base case is just the fact that there is a degree zero design on $N < q$ nodes; the gaps between the induction steps are again easily filled by restrictions. □

As a second application of designs, we use Lemma 4.2 to prove a stronger lower bound in characteristic zero (this does not seem to have a proof-theoretic corollary so far). In the proof we simply make the value of the design equal for all partitions of the same size, and choose this value inductively so that the design condition is satisfied.

Theorem 4.6. *Let $N \ge q \ge 2$ such that $N \not\equiv 0 \pmod q$ be arbitrary. Then every Nullstellensatz refutation of the (N, q)-system over $\mathbf{Q}$ must have degree bigger than $\lfloor \frac{N}{q} \rfloor$.*

Proof. We let $s_E = s_{|E|}$, where the values $s_0, s_1, \ldots, s_{\lfloor \frac{N}{q} \rfloor}$ are defined inductively as follows:

$$\begin{aligned} s_0 &:= 1, \\ s_{t+1} &:= s_t \cdot \binom{N - tq - 1}{q-1}^{-1}. \end{aligned}$$

This function satisfies the properties required from a design as long as $N - tq - 1 \ge q - 1$ (i.e., $\frac{N}{q} > t$), as $\binom{N-tq-1}{q-1}$ is the number of partitions extending some fixed partition E of cardinality t by one class containing a fixed element $v \in ([N] \setminus \bigcup E)$.

The lower bound then follows by Lemma 4.2. □

5. Cut-elimination and completeness for algebraic systems

For the rest of the paper we assume that our underlying ring Λ is always a finite field $\mathbf{F}_p$ for some fixed prime p.

In this section, we shall prove an analogue of the cut-elimination theorem, and then use this to give a constructive proof of the completeness theorem for the Nullstellensatz and polynomial calculus systems and to prove that tree-like polynomial calculus proofs can be transformed into Nullstellensatz proofs with only a small increase in degree. Another accessible account of the completeness of the Nullstellensatz system is given in the survey Pitassi (1996).

What we mean by an analogue of cut-elimination is the following. Suppose F is a set of polynomials over $\mathbf{F}_p$ and there is a derivation, D_1, of h from $F \cup g$ and there is another derivation, D_2, from $F \cup \{1 - g^{p-1}\}$. Then there ought to be a derivation of h from just F. Indeed there is such a derivation; the theorems below give bounds on the degree of the derivation of h from F in terms of the degrees of D_1 and D_2.

We shall make two changes to the polynomial calculus proof system in order to make Theorem 5.1 easier to state and prove. Firstly, instead of counting the number of lines (polynomials) in a polynomial calculus derivation, we count the number of addition inferences in the proof. We call the number of addition inferences in a derivation the *A-size* of the derivation. Note that the total number of inferences is at most three times the A-size plus one, and the sole reason for using A-size instead of number of steps is to simplify the counting in the proof below.

Secondly, instead of adjoining polynomials $x^2 - x$ as initial polynomials, we would like to confine our attention to multilinear polynomials from the very beginning as this is a canonical way to represent ($\mathbf{F}_p$-valued) functions on $\{0,1\}^n$. Let $\mathit{Mult}(f)$ denote the unique multilinear polynomial equal to f modulo the ideal generated by $(x_i^2 - x_i)$. For example, $\mathit{Mult}(x^2y + xy^2) = 2xy$. Notice that $\mathit{Mult}(f) = \mathit{Mult}(g)$ if and only if f and g represent the same function on $\{0,1\}^n$, and that $\mathit{Mult}(f) = 0$ iff f belongs to the ideal generated by all $(x_i^2 - x_i)$ (Smolensky 1987).

We define the *multilinear polynomial calculus* proof system similarly to the polynomial calculus with the exception that it operates only with multilinear polynomials, and the multiplication rule is modified accordingly to that:

$$\frac{f}{\mathit{Mult}(f \cdot g)},$$

where f and g are arbitrary multilinear polynomials. Notice that every (ordinary) polynomial calculus proof of a multilinear polynomial from a set of multilinear polynomials can be transformed into a multilinear polynomial calculus proof simply by applying the operator *Mult* to every line. Also, the multilinear polynomial calculus proof system is clearly a subsystem of semantic derivations from Theorem 2.6 and, therefore, it can be simulated by ordinary polynomial calculus. To simplify the notation, we omit the operator *Mult* throughout the rest of this section whenever this can not create confusion. Thus, expressions like $g \cdot h$ or $1 - g^{p-1}$ actually mean $\mathit{Mult}(g \cdot h)$ and $\mathit{Mult}(1 - g^{p-1})$, respectively.

Theorem 5.1.

(1) *Let F be a set of multilinear polynomials. Suppose there are multilinear polynomial calculus (resp., Nullstellensatz) derivations D_0 and D_1 of h from $F \cup \{g\}$ and from $F \cup \{1 - g^{p-1}\}$, respectively, where g, h are some multilinear polynomials. Let D_0 and D_1 both have degree less than or equal to d. Then there is a multilinear polynomial calculus (or Nullstellensatz, respectively) derivation D of h from F such that the degree of D is less than or equal to $d + (p-1) \cdot deg(g)$. For the multilinear polynomial calculus case, the A-size of D is at most the sum of the A-sizes of D_0 and D_1.*

(2) *Let $F = \{f_i(\vec{x}, y)\}_i$ be a set of multilinear polynomials in the variables $\vec{x}, y$. Let $F_0 = \{f_i(\vec{x}, 0)\}_i$ and $F_1 = \{f_i(\vec{x}, 1)\}_i$. Suppose D_0 and D_1 are multilinear polynomial calculus (or Nullstellensatz) refutations of F_0 and F_1, respectively, both of degree at most d. Then there is a multilinear polynomial calculus (or Nullstellensatz, respectively) refutation of F with degree at most $d + 1$. For the multilinear polynomial calculus case, the A-size of D is at most the sum of the A-sizes of D_0 and D_1 plus one.*

(3) *Let F be a set of multilinear polynomials and g, h be some multilinear polynomials such that for all $0 \leq c < p$, there are multilinear polynomial calculus (resp., Nullstellensatz) derivations D_c of h from $F \cup \{c - g\}$. Suppose also that each D_c has degree less than or equal to d. Then there is a multilinear polynomial calculus (or Nullstellensatz, respectively) derivation D of h from F such that the degree of D is less than or equal to $d + (p-1) \cdot deg(g)$. For the multilinear polynomial calculus case, the A-size of D is at most the sum of the A-sizes of the D_c's plus $(p-1)$.*

Proof. We'll just give the proofs for the multilinear polynomial calculus; they are identical (or even easier) for the Nullstellensatz system.

(1) If either g does not appear in D_0 or $1 - g^{p-1}$ does not appear in D_1, then the theorem is trivially true. So suppose both do appear in the derivation. Modify the derivation D_0 by multiplying every line in D_0 by the polynomial $1 - g^{p-1}$: the line g simplifies to just 0, so it can be removed from the derivation; the initial polynomials $f \in F$ become $(1 - g^{p-1})f$, each of which can be derived in one step. The last line, h, becomes $(1 - g^{p-1})h$. Thus we obtain a derivation D_0' of $(1 - g^{p-1})h$. Similarly, multiplying every line in D_1 by g results in a derivation D_1' of gh. Combining D_0' and D_1', by first multiplying gh by g^{p-2} and then adding $g^{p-1}h$ and $(1 - g^{p-1})h$ yields the desired derivation of h from F. Note that D_0' and D_1' have A-sizes strictly less than the A-sizes of D_0 and D_1, since at least one addition inference involving g (respectively, $1 - g^{p-1}$) was removed when forming D_0' (respectively, D_1'). Therefore, D satisfies the desired degree bound and A-size bound.

(2) This can be obtained as a corollary to (1); however, it is more interesting to give a direct proof. First form a derivation D_1' of y from $yF_1 = \{yf_i(\vec{x}, 1)\}_i$, by multiplying every polynomial in D_1 by y. Now $Mult(yf(\vec{x}, 1))$ is the same polynomial as $Mult(yf(\vec{x}, y))$ since they represent the same functions on 0-1-inputs (more generally, on all inputs where y is

either 0 or 1). Thus D_1' is a derivation of y from yF, and by adding members of F to the beginning, it becomes a derivation of y from F. By a dual argument, we also have a derivation D_0' of $(1-y)$ from F. Now combine D_0' and D_1' with a single addition inference to form the desired refutation of F.

(3) For each $c < p$, form the polynomial

$$g_c = \prod_{\substack{0 \le i < p \\ i \ne c}} (i - g).$$

Note that $\mathit{Mult}(g_c \cdot (c-g)) = 0$. Then multiply each derivation D_c by g_c to get a derivation D_c' of $g_c h$ from F. Finally, note that $\mathit{Mult}(\sum_c g_c) = -1$, so the derivations D_c' can be combined to get a derivation of h from F. □

Remark: Theorem 5.1(2) gives a constructive proof of Hilbert's Nullstellensatz when the polynomials $x^2 - x$ are available for every variable x. This is equivalent to the completeness of the Nullstellensatz system and the polynomial calculus system. We shall do a little better and prove a stronger version of completeness for these systems:

Theorem 5.2.

(1) *If F is a set of polynomials with no 0-1 solutions, then there is a Nullstellensatz refutation from F.*

(2) *If g is a tautological consequence of F, then there is a Nullstellensatz proof of g from F.*

Proof. To prove part (1), proceed by induction on the number of variables used in F. Theorem 5.1(2) is exactly what is needed for the induction step.

To prove part (2), suppose g is a polynomial which is zero for all 0-1 assignments that make every member of F zero. By already proven part (1), there is a proof of 1 from $F \cup \{1 - g^{p-1}\}$. Multiplying this proof by g, one obtains a proof of g from F. □

Recall that a proof is *tree-like* if every line in the proof is a hypothesis of at most one inference, and that the *height* of a tree-like proof is the height of its proof-tree (cf. Krajíček 1995). At the end of this section we show how to convert tree-like (multilinear) polynomial calculus proofs into Nullstellensatz proofs with only a modest increase in degree. To illustrate the general idea, we begin with the comparatively easy case of low height proofs.

Theorem 5.3. *Let $p \ge 2$ be a prime, $x_1, \dots, x_n$ variables and let $f_1, \dots, f_k, g \in \mathbf{F}_p[\bar{x}]$ be multilinear.*

(1) *If there is a degree d Nullstellensatz proof over $\mathbf{F}_p$ of g from $f_1, \dots, f_k$, then there is also a tree-like multilinear polynomial calculus proof over $\mathbf{F}_p$ of g from $f_1, \dots, f_k$ of degree at most d and of height at most $\lceil \log_2 k \rceil + 1$.*

(2) *If there is a tree-like multilinear polynomial calculus proof over* $\mathbf{F}_p$ *of* g *from* $f_1, \ldots, f_k$ *that has degree* d *and height* h*, then there is a degree* $(p-1)(h+1)d$ *Nullstellensatz proof over* $\mathbf{F}_p$ *of* g *from* $f_1, \ldots, f_k$.

Proof. (1) Assume that there is a Nullstellensatz proof

$$\sum_i P_i f_i + \sum_j R_j(x_j^2 - x_j) = g$$

of degree d. This immediately gives a degree d multilinear polynomial calculus proof which just computes the sum $\sum_i P_i f_i$ arranged in the form of an almost balanced binary tree.

(2) By induction on the height h. The base case $h = 0$ is obvious. If $h \geq 1$ and g is inferred by the addition rule, the inductive step is also obvious. Finally, if g is inferred from some g' by the multiplication rule, then, by inductive assumption, g' has a Nullstellensatz proof of degree at most $(p-1)hd$. Now we simply multiply this proof by $g(g')^{p-2}$ (which gives us a proof of $g(g')^{p-1}$), apply the last part of Lemma 2.2 to get a Nullstellensatz proof of the polynomial $g - g(g')^{p-1}$ (note that $\mathit{Mult}(g - g(g')^{p-1}) = 0$) and sum these two proofs together to get the required Nullstellensatz proof of g. □

Now we use Theorem 5.1(1) for extending this result to the case of tree-like polynomial calculus proofs that do not necessarily have low height.

Theorem 5.4. *Let* P *be a tree-like* (*multilinear or regular*) *polynomial calculus derivation of a multilinear polynomial* g *from a set* F *of multilinear polynomials. Let* P *have degree* d *and have A-size* S*. Then there is a Nullstellensatz derivation of* g *from* F *of degree* $O(d \log S)$.

A corollary to Theorem 5.4 is that non-treelike polynomial calculus refutations cannot in general be converted into tree-like polynomial calculus refutations with only a small increase in the degree. This is because of the "house-sitting" tautologies of Clegg *et al.* (1996) which have polynomial size, constant degree polynomial calculus refutations, but for which any Nullstellensatz refutation requires degree $\Omega(\sqrt{n})$.[2]

Proof of Theorem 5.4. It will suffice to prove the theorem for the more general case of the multilinear polynomial calculus. The basic idea of the proof is to apply the $\frac{1}{3}$-$\frac{2}{3}$ trick to a tree-like derivation. We'll use induction on S to prove that any degree d, A-size S multilinear polynomial calculus derivation D can be converted into a degree $Cpd \log S$ Nullstellensatz derivation where $C = 2/(\log 3 - 1)$.[3] This is obvious for $S = 1$. For $S > 1$, use the Brent-Spira trick to find a line h in the derivation D such that h is derived with S_1 addition inferences, where $S/3 \leq S_1 \leq (2S)/3$. Consider the subderivation D_1 of D ending with h; also, let D_2 be the derivation of g from $F \cup \{h\}$ which is obtained by removing the subderivation D_1 from D.

[2] The design of Clegg *et al.* (1996) can be modified to work for all values of p (actually for all rings), see Buss (1995a). The polynomials used for the house-sitting principle are all of degree 1 or 2.

[3] We write $\log S$ to mean $\max\{1, \log_2 S\}$.

Applying the induction hypothesis to D_1, we know that there is a Nullstellensatz derivation D_1' of h from F of degree at most $d' = Cpd\log(2S/3)$. Likewise, there is a Nullstellensatz derivation of g from $F \cup \{h\}$ also with degree at most d'. It is easy to modify D_1' to obtain a Nullstellensatz derivation D_1'' of g from $F \cup \{1 - h^{p-1}\}$ of degree at most $d'' = d' + (p-2)deg(h) + deg(g)$: namely, multiply the derivation D_1' by h^{p-2} and then add $(1 - h^{p-1})$ and finally multiply the derivation by g. Combining D_1'' and D_2' using Theorem 5.1(1), we get a Nullstellensatz refutation of degree at most $d'' + (p-1)deg(h)$. Since both $deg(g)$ and $deg(h)$ are at most d, the degree is bounded by

$$\begin{aligned} d'' + (p-1)deg(h) &\leq d' + 2(p-1)d \\ &\leq Cpd\log(2S/3) + 2pd \\ &= Cpd\log S + 2pd - Cpd(\log 3 - 1) \\ &= Cpd\log S. \end{aligned}$$

That completes the proof of Theorem 5.4. □

6. Nullstellensatz with extension polynomials

Our aim in this section is to reduce the problem of proving lower bounds for constant-depth $F(MOD_p)$-proofs to a question about degrees in Nustellensatz. We define *extension polynomials* and we show that the degree of Nullstellensatz proofs with these polynomials is related to the size of constant-depth $F(MOD_p)$-proofs (see Theorem 6.7). The idea behind these polynomials is to replace bounded depth formulas, which are *apriori* polynomials of large degree, by polynomials of small degree. We use the same method as in the boolean complexity (Razborov 1987, Smolensky 1987) with the difference that instead of the random bits we use for approximation generic bits called in our framework *extension variables*. The extension polynomials then simply express soundness of the approximation, and every formula becomes equivalent to a low-degree polynomial modulo the ideal generated by the extension polynomials. In this way we can reduce the question of proving lower bounds for bounded depth Frege systems with MOD_p gates to proving lower bounds for Nullstellensatz refutations. However, we have to add to the initial polynomials unknown in advance extension polynomials, and this makes the task of proving lower bounds for Nullstellensatz refutations harder. In particular, current methods apparently cannot be applied to such systems.

Since we will not consider any longer algebraic proof systems other than Nullstellensatz, we introduce a handy notation that simplifies later presentation.

DEFINITION 6.1. *Let $f_1, \ldots, f_k, g$ be polynomials over $\mathbf{F}_p$ with variables*

$$x_1, \ldots, x_n, r_1, \ldots, r_m.$$

Then

$$f_1, \ldots, f_k \vdash_d g$$

denotes the fact that g has a Nullstellensatz proof from $f_1, \ldots, f_k$ of degree at most d, where we automatically include as initial polynomials all $(x_i^2 - x_i)$ (cf. Definition 2.1) and all $(r_j^p - r_j)$.

The translation ϕ^* used in the following lemma was defined at the beginning of Section 2.

Lemma 6.2. *Let Γ be a set of polynomials, and let f, g and f_i's and g_i's be polynomials over $\mathbf{F}_p$. Then*

(1) *If f and g represent the same function on $\mathbf{F}_p$ then $\vdash_{deg(f-g)}\ (f-g)$. In particular, $\vdash_{p \cdot deg(f)}\ (f^p - f)$ for arbitrary f.*

(2) *If $\Gamma \vdash_d\ (1-f)g$ and $\Gamma \vdash_d\ (1-g)f$ then $\Gamma \vdash_d\ (f-g)$.*

(3) *Let $\phi(p_1, \ldots, p_k) = MOD_{p,i}(p_1, \ldots, p_k)$, and assume that $\Gamma \vdash_d\ (f_j - g_j)$ for $j = 1, \ldots, k$. Then*

$$\Gamma \vdash_{d'}\ \phi^*(\bar{f}) - \phi^*(\bar{g}),$$

where $d' = d + (p-2) \cdot \max\{\max_j deg(f_j),\ \max_j deg(g_j)\}$.

(4) *Let $\phi(p_1, p_2) = (p_1 \vee p_2)$, and assume that $\Gamma \vdash_{d_1} (f_1 - g_1)$, $\Gamma \vdash_{d_2} (f_2 - g_2)$, where*

$$d_1 \geq \max\{deg(f_1), deg(g_1)\}, \quad d_2 \geq \max\{deg(f_2), deg(g_2)\}.$$

Then

$$\Gamma \vdash_{d_1+d_2}\ \phi^*(\bar{f}) - \phi^*(\bar{g}).$$

Proof. (1) Similarly to Lemma 2.2.

(2) Obvious.

(3,4) We prove both statements simultaneously. Write out the algebraic term $\phi^*(x_1 + y_1, \ldots, x_k + y_k)$ in the polynomial form $\phi^*(\bar{x}) + \phi'(\bar{x}, \bar{y})$, where every monomial in ϕ' has at least one occurrence of y's. Substituting g_j for x_j and $(f_j - g_j)$ for y_j, we find

$$\phi^*(\bar{f}) - \phi^*(\bar{g}) = \phi'(g_1, \ldots, g_k, f_1 - g_1, \ldots, f_k - g_k).$$

Now, every monomial in the right-hand side has an explicit occurrence of $(f_j - g_j)$ for some j and thus can be derived from Γ using $\Gamma \vdash_d\ (f_j - g_j)$. It is easy to see that in both cases the derivation of $\phi^*(\bar{f}) - \phi^*(\bar{g})$ obtained by summing up over all monomials in ϕ' meets the required degree bound. □

Definition 6.3. *The* degree $deg(\phi)$ *of a formula ϕ is defined inductively as follows:*

1. *$deg(TRUE) := 0$.*

2. *$deg(p_i) := 1$.*

3. *$deg(\neg\phi) := deg(\phi)$.*

4. *$deg(MOD_{p,i}(\phi_1, \ldots, \phi_k)) := (p-1) \cdot \max_j deg(\phi_j)$.*

5. *$deg(\phi \vee \psi) := deg(\phi) + deg(\psi)$.*

The meaning of this definition is that $deg(\phi)$ provides an upper bound on the degree of ϕ^*, and this bound is "effective" in the sense that if ψ is a subformula of ϕ, then $deg(\psi) \leq deg(\phi)$.

The following is a crucial definition.

Definition 6.4. *Let $\bar{g} = g_1, \ldots, g_m$ be polynomials over $\mathbf{F}_p$ and let $h \geq 1$ be a fixed number. Take new variables r_{iu}, $i = 1, \ldots, m$ and $u = 1, \ldots, h$ not occurring in any of g_j.*

The polynomials $E_{i,\bar{g}}$:

$$g_i \cdot \Pi_{u \leq h}(1 - \sum_{j \leq m} r_{ju} g_j),$$

one for each $i \leq m$, are called extension polynomials of accuracy h corresponding to $\bar{g}$. *The variables r_{iu} are called* extension variables corresponding to $\bar{g}$.

Note that in accordance with Definition 6.1 variables r_{iu} are treated differently than variables x_i for which we have $x_i^2 - x_i$ among the initial polynomials.

To motivate this definition consider the polynomial

$$f(\bar{x}, \bar{r}) = 1 - \Pi_{u \leq h}(1 - \sum_{j \leq m} r_{ju} g_j).$$

If $\bigwedge_i (g_i = 0)$ then $f(\bar{x}, \bar{r}) = 0$. On the other hand, if $f(\bar{x}, \bar{r}) = 0$ then $\Pi_{u \leq h}(1 - \sum_{j \leq m} r_{ju} g_j) = 1$ and thus, *assuming* that all extension polynomials corresponding to $\bar{g}$ are zero, necessarily $\bigwedge_i (g_i = 0)$. Hence the equation $f(\bar{x}, \bar{r}) = 0$ of degree at most $h \cdot (1 + \max_i deg(g_i))$ represents the condition $\bigwedge_i (g_i = 0)$ of a possibly much bigger degree $\sum_i deg(g_i)$ modulo the additional assumption $E_{1,\bar{g}} = \ldots = E_{m,\bar{g}} = 0$.

Definition 6.5. *Let $\mathcal{E}$ be a set of extension polynomials. We call the set $\mathcal{E}$* leveled *if:*

1. *All variables occurring in $\mathcal{E}$ are either x-variables or extension variables of some polynomial in $\mathcal{E}$.*

2. *If $\mathcal{E}$ contains some extension polynomial, $E_{i,g_1,\ldots,g_m}$, then $\mathcal{E}$ must contain all companion extension polynomials $E_{1,\bar{g}}, E_{2,\bar{g}}, \ldots, E_{m,\bar{g}}$.*

3. *$\mathcal{E}$ can be decomposed into levels $\mathcal{E} = \mathcal{E}_1 \dot{\cup} \ldots \dot{\cup} \mathcal{E}_\ell$ in such a way that for any polynomial $E_{i,g_1,\ldots,g_m} \in \mathcal{E}_j$ its companion polynomials $E_{1,\bar{g}}, E_{2,\bar{g}}, \ldots, E_{m,\bar{g}}$ also belong to $\mathcal{E}_j$, and extension variables corresponding to $\bar{g}$ do not occur in any other polynomial from $\mathcal{E}_1 \dot{\cup} \ldots \dot{\cup} \mathcal{E}_j$. The minimal ℓ for which such a decomposition is possible is called the* depth *of $\mathcal{E}$.*

THEOREM 6.6. *Let $f_1, \ldots, f_k, g$ be polynomials from $\mathbf{F}_p[x_1, \ldots, x_n]$, and let $\mathcal{E}$ be any leveled set of extension polynomials. If g has a Nullstellensatz proof from $f_1, \ldots, f_k, \mathcal{E}$ then it has one from $f_1, \ldots, f_k$ alone as well.*

PROOF. First we prove a claim:

Claim: *Let $\{E_{1,\bar{g}}, \ldots, E_{m,\bar{g}}\} \subseteq \mathcal{E}$. For any assignment $\bar{a}$ to variables of $g_1, \ldots, g_m$ there is an assignment $\bar{b}$ to the extension variables of $E_{i,\bar{g}}$ such that $E_{i,\bar{g}}(\bar{a}, \bar{b}) = 0$ for all $1 \leq i \leq m$.*

If all $g_i(\bar{a}) = 0$ then set r_{iu} arbitrarily. Otherwise let $g_{i_0}(\bar{a}) \neq 0$ for some i_0. Put $r_{iu} := 0$ if $i \neq i_0$ or $u > 1$, and $r_{i_0 1} := (g_{i_0}(\bar{a}))^{-1}$.

From the claim it follows that g is zero on any 0-1 assignment satisfying all f_i (as that assignment can be extended, level by level, to a satisfying assignment for $\mathcal{E}$). By Theorem 5.2, this is equivalent to the existence of a Nullstellensatz proof of g from $\bar{f}$. □

Now we are ready to formulate the main result of this section.

THEOREM 6.7. *Let $\phi_1, \ldots, \phi_k$ be (constant-depth) $F(MOD_p)$-formulas in variables $x_1, \ldots, x_n$. Let d_0 be the maximal degree of ϕ_i's, ℓ be any fixed constant bigger than the depth of any ϕ_i and $h \geq 1$ be an arbitrary parameter.*

(1) *If there exists a depth-ℓ $F(MOD_p)$-refutation of $\phi_1, \ldots, \phi_k$ which has length S, then there exists a leveled set $\mathcal{E}$ of extension polynomials with accuracy h such that $|\mathcal{E}| \leq S^{O(1)}$, the depth of $\mathcal{E}$ is at most $\ell + C$ for some absolute constant C, and $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}$ has a Nullstellensatz refutation of degree $(d_0 + \log S)(h+1)^{O(1)}$.*

(2) *If there exists a leveled set $\mathcal{E}$ of extension polynomials with accuracy h such that $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}$ has a Nullstellensatz refutation of some degree d, then there exists a leveled set $\mathcal{E}'$ of extension polynomials with accuracy 1 whose cardinality and depth are at most the cardinality and depth of $\mathcal{E}$, and such that $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}'$ also has a Nullstellensatz refutation of the same degree d.*

(3) *If there exists a leveled set $\mathcal{E}$ of extension polynomials with accuracy 1 of depth ℓ and cardinality S such that $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}$ has a Nullstellensatz refutation of some degree d, then there exists a depth-ℓ' $F(MOD_p)$-refutation of $\phi_1, \ldots, \phi_k$ that has length $(S+n+k)^{O(d_0+d)}$. Here ℓ' is some absolute constant depending only on ℓ.*

PROOF. We begin with easier parts (2), (3).

(2) Substitute in the Nullstellensatz refutation of $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}$ zeros to all extension variables r_{iu} if $u > 1$. This transforms extension polynomials with accuracy $h \geq 1$ into extension polynomials with accuracy $h = 1$.

(3) What we basically need for this part is a formalization of the proof of Theorem 6.6 in the bounded depth fragment of $F(MOD_p)$. For this purpose, let π be a Nullstellensatz refutation of $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}$ with the specified parameters. For every extension variable

r appearing in π introduce p *propositional* variables $r^0, r^1, \ldots, r^{(p-1)}$, where r^i expresses the fact $r = i$. Let the set of axioms *Ext* say that for any such r exactly one variable among $r^0, r^1, \ldots, r^{(p-1)}$ takes value $TRUE$. For a polynomial $f(x_1, \ldots, x_n, \bar r)$ let us denote by $\tilde f(p_1, \ldots, p_n, \bar r^0, \bar r^1, \ldots, \bar r^{(p-1)})$ an $F(MOD_p)$-formula expressing the fact $f(x_1, \ldots, x_n, \bar r) = 0$. Writing f as a sum of monomials yields $\tilde f$ of constant depth and size $(S+n)^{O(d)}$ for every f appearing in the refutation π.

Now we transform π into an $F(MOD_p)$ refutation of $\widetilde{\phi_1^*}, \ldots, \widetilde{\phi_k^*}, Ext, \tilde{\mathcal{E}}$. This refutation has constant depth (independent of ℓ) and length $(S+n)^{O(d)}$; it is constructed simply by simulating in $F(MOD_p)$ the process of reducing the left-hand side of π to the polynomial form.

Next we observe that $\widetilde{\phi^*}$ has a short $F(MOD_p)$-proof from $\phi(p_1, \ldots, p_n)$ easily constructed by induction on the logical complexity of ϕ. If ϕ is one of ϕ_i's, the depth of this proof is $O(1)$ (as ϕ_i's have constant depth), and its length is at most $n^{O(d_0)}$. Similarly, for any extension polynomial $E_{i,\bar g} = g_i \cdot (1 - \sum_{j \le m} r_j g_j)$ from $\mathcal{E}$, the formula $\widetilde{E_{i,\bar g}}$ has a constant-depth $F(MOD_p)$-proof from *Ext* and

$$E'_{i,\bar g} = \tilde g_i \ \vee\ \left(\bigwedge_{j_1 \neq j_2} (r^0_{j_1} \vee r^0_{j_2}) \ \wedge\ \bigvee_{c=1}^{p-1} \bigvee_{j=1}^{m} (\widetilde{g_j^{(c)}} \wedge r_j^{(c^{-1})}) \right),$$

where c^{-1} is the multiplicative inverse modulo p and we denoted (for typographical reasons) $g_j - c$ by $g_j^{(c)}$. Indeed, $g_i = 0$ obviously implies $E_{i,\bar g} = 0$. $\bigwedge_{j_1 \neq j_2} (r^0_{j_1} \vee r^0_{j_2})$ says that at most one of r_j's is different from 0, and if we additionally know that $g_j = c,\ r_j = c^{-1}$ for some j, c, then this already suffices to conclude that $E_{i,\bar g} = 0$. It is easy to see that this informal argument can be formalized in $F(MOD_p)$, and the length of the resulting proof is at most $(S+n)^{O(d)}$.

Let $\mathcal{E}'$ be the set of all formulas $E'_{i,\bar g}$ corresponding to extension polynomials from $\mathcal{E}$. Combining things together, we have a constant-depth $F(MOD_p)$-refutation of $\phi_1, \ldots, \phi_k, Ext, \mathcal{E}'$ which has length $(S+n+k)^{O(d_0+d)}$, and we only should get rid of the axioms $Ext, \mathcal{E}'$. For doing that we will show how to substitute constant-depth formulas for propositional variables r^c in such a way that after this substitution $Ext, \mathcal{E}'$ will have short $F(MOD_p)$-proofs. We will proceed level by level, and for extension variables corresponding to different blocks of polynomials from the same level the substitution will be constructed in parallel. Since we have only constantly many (namely, ℓ) levels and after performing substitutions at every particular level the depth of the proof grows at most linearly, and its length grows at most polynomially, the final proof will still have constant depth (depending only on ℓ) and length $(S+n+k)^{O(d_0+d)}$.

Hence, it is sufficient to treat just one group of axioms from $\mathcal{E}'$ that in general has the form

$$\psi_i^0 \ \vee\ \left(\bigwedge_{j_1 \neq j_2} (r^0_{j_1} \vee r^0_{j_2}) \ \wedge\ \bigvee_{c=1}^{p-1} \bigvee_{j=1}^{m} (\psi_j^c \wedge r_j^{(c^{-1})}) \right).$$

Here ψ_j^c are constant-depth formulae obtained from $\widetilde{g_j^{(c)}}$ by substitutions from the previous levels; this implies that for every fixed j there is a short $F(MOD_p)$-proof that exactly one of $\psi_j^0, \psi_j^1, \ldots, \psi_j^{(p-1)}$ is true.

Let

$$r_j^0 := (\neg\psi_1^0) \vee \ldots \vee (\neg\psi_{j-1}^0) \vee \psi_j^0,$$

and

$$r_j^c := (\neg r_j^0) \wedge \psi_j^{(c^{-1})} \ (c \geq 1).$$

Then $F(MOD_p)$ easily proves that $(\neg\psi_i^0)$ implies that exactly one of $r_1^0, \ldots, r_m^0$ is false (namely, the one corresponding to the minimal i with this property). But if r_j^0 (and hence ψ_j^0) is false, the disjunction $\bigvee_{c=1}^{p-1}(\psi_j^c \wedge r_j^{(c^{-1})})$ is clearly satisfied. So, our substitution has all the required properties, and this completes the proof of part (3).

(1) For this part we have to formalize the argument from the motivating remark after Definition 6.4 in the Nullstellensatz system. Since the resulting proof turns out to be somewhat lengthy and technical, this is convenient to split it into a sequence of independent lemmas. We begin with the following crucial definition.

Definition 6.8. *For any $F(MOD_p)$-formula $\phi(\bar{p})$ and $h \geq 1$ define a polynomial $\phi^{\mathbf{ap}}$, called* the approximation of ϕ with accuracy h, *as follows.*

1. *If $\phi = TRUE$ then $\phi^{\mathbf{ap}} := 0$.*

2. *If $\phi = p_i$ then $\phi^{\mathbf{ap}} := x_i$.*

3. *If $\phi = (\neg\psi)$ then $\phi^{\mathbf{ap}} := 1 - \psi^{\mathbf{ap}}$.*

4. *If $\phi = MOD_{p,i}(\psi_1, \ldots, \psi_k)$ then $\phi^{\mathbf{ap}} := (\psi_1^{\mathbf{ap}} + \cdots + \psi_k^{\mathbf{ap}} - (k-i))^{p-1}$.*

5. *If $\phi = \eta(\psi_1, \ldots, \psi_m)$, where $m \geq 2$, η is a formula containing no connective other than the disjunction, and $\psi_1, \ldots, \psi_m$ already do not begin with a disjunction, then*

$$\phi^{\mathbf{ap}} := \Pi_{u \leq h}\left(1 - \sum_{j \leq m} r_{ju}(\neg\psi_j)^{\mathbf{ap}} \right),$$

where r_{ju} are the extension variables corresponding to the polynomials

$$(\neg\psi_1)^{\mathbf{ap}}, \ldots, (\neg\psi_m)^{\mathbf{ap}}.$$

The peculiar occurrences of the negation sign in the approximation of a disjunction are due to the fact that the truth-value $TRUE$ is represented in polynomial rings by zero rather than by one. Note also that the first four items in this definition are identical to the corresponding items in the definition of ϕ^*, the set of all extension polynomials introduced by item 5 is leveled, and its depth is at most the depth of ϕ.

Lemma 6.9. *Let $\phi_1 = \bigvee_{i \in I_1} \xi_i$ and $\phi_2 = \bigvee_{i \in I_2} \xi_i$ be two $F(MOD_p)$-formulas that are (arbitrarily bracketed) disjunctions of some ξ_i's not beginning with a disjunction. Let $h \geq 1$ and let $d = 3h(1 + \max_i deg(\xi_i^{\mathbf{ap}}))$. Then*

$$\mathcal{E} \vdash_d \ (\phi_1 \vee \phi_2)^{\mathbf{ap}} - \phi_1^{\mathbf{ap}} \cdot \phi_2^{\mathbf{ap}},$$

where $\mathcal{E}$ is the set of extension polynomials corresponding to sets $\{(\neg\xi_i)^{\mathbf{ap}} \mid i \in I_1\}$, $\{(\neg\xi_i)^{\mathbf{ap}} \mid i \in I_2\}$ and $\{(\neg\xi_i)^{\mathbf{ap}} \mid i \in I_1 \cup I_2\}$.

Proof. Put $f_i := (\neg\xi_i)^{\mathbf{ap}}$. We want to find a Nullstellensatz proof of

$$\Pi_u(1 - \sum_{I_1 \cup I_2} r_{iu} f_i) \ - \ \Pi_u(1 - \sum_{I_1} r'_{iu} f_i)(1 - \sum_{I_2} r''_{iu} f_i)$$

from $\mathcal{E}$.

Claim 1: *For any $v \leq h$:*

$$\mathcal{E} \vdash_{d'} \ (\sum_{I_1} r'_{iv} f_i)\Pi_u(1 - \sum_{I_1 \cup I_2} r_{iu} f_i),$$

where $d' = (h+1)\,(1 + \max_i deg(f_i))$.

To prove the claim sum all extension axioms $E_{i,\bar{f}}$, $i \in I_1$, multiplied by r'_{iv}, where $\bar{f} = \{(\neg\xi_i)^{\mathbf{ap}} \mid i \in I_1 \cup I_2\}$.

Claim 2: $\mathcal{E} \vdash_d \ (1 - \phi_1^{\mathbf{ap}}\phi_2^{\mathbf{ap}})(\phi_1 \vee \phi_2)^{\mathbf{ap}}$.

The polynomial to be proved from $\mathcal{E}$ is of the form

$$\sum \, [(\sum_{I_1} r'_{iu_1} f_i) \cdot \ldots \cdot (\sum_{I_1} r'_{iu_\ell} f_i) \ \cdot \ (\sum_{I_2} r''_{iv_1} f_i) \cdot \ldots \cdot (\sum_{I_2} r''_{iv_m} f_i)] \ \cdot \ [\Pi_u(1 - \sum_{I_1 \cup I_2} r_{iu} f_i)],$$

where $\ell + m \geq 1$ and $\ell, m \leq h$. Hence Claim 2 follows from Claim 1.

Claim 3: *For all $v \leq h$*

$$\mathcal{E} \vdash_{d''} \ (\sum_{I_1 \cup I_2} r_{iv} f_i) \cdot \Pi_u[(1 - \sum_{I_1} r'_{iu} f_i)(1 - \sum_{I_2} r''_{iu} f_i)],$$

where $d'' := (2h+1)(1 + \max_i deg(f_i))$.

By (the proof of) Claim 1

$$\mathcal{E} \vdash_{d'} \ (\sum_{I_1} r_{iv} f_i) \cdot \Pi_u(1 - \sum_{I_1} r'_{iu} f_i)$$

and

$$\mathcal{E} \vdash_{d'} \ (\sum_{I_2 \setminus I_1} r_{iv} f_i) \cdot \Pi_u(1 - \sum_{I_2} r''_{iu} f_i).$$

Hence

$$\mathcal{E} \vdash_{d''} \ (\sum_{I_1} r_{iv} f_i) \cdot \Pi_u[(1 - \sum_{I_1} r'_{iu} f_i)(1 - \sum_{I_2} r''_{iu} f_i)]$$

and

$$\mathcal{E} \vdash_{d''} \ (\sum_{I_2 \setminus I_1} r_{iv} f_i) \cdot \Pi_u[(1 - \sum_{I_1} r'_{iu} f_i)(1 - \sum_{I_2} r''_{iu} f_i)].$$

The claim follows by adding the last two equations.

The next claim is proved analogously to Claim 2.

Claim 4: $\mathcal{E} \vdash_d \ (1 - (\phi_1 \vee \phi_2)^{\mathbf{ap}})(\phi_1^{\mathbf{ap}} \cdot \phi_2^{\mathbf{ap}})$.

The lemma now follows from Claims 2 and 4, and from Lemma 6.2(2). □

The next lemma provides an upper bound on $deg(\phi^{\mathbf{ap}})$ and is proved by an easy induction on ℓ.

Lemma 6.10. *Let ϕ be an $F(MOD_p)$-formula of depth ℓ, and let $\phi^{\mathbf{ap}}$ be its approximating polynomial of accuracy $h \geq 1$. Then*

$$deg(\phi^{\mathbf{ap}}) \leq (\ell + 1) \cdot (\max\{p-1, h\}^{\ell}).$$

Lemma 6.11. *For any $F(MOD_p)$-formula $\phi(p_1, \dots, p_k)$ of depth ℓ and any formulas $\psi_1, \dots, \psi_k$,*

$$\mathcal{E} \vdash_d \ \phi^*(\psi_1^{\mathbf{ap}}, \dots, \psi_k^{\mathbf{ap}}) - (\phi(\psi_1, \dots, \psi_k))^{\mathbf{ap}},$$

where $d = deg(\phi) \cdot 3h \left(1 + (\ell+1) \cdot (\max\{p-1, h\}^{\ell}) \cdot \max_i deg(\psi_i^{\mathbf{ap}})\right)$ and $\mathcal{E}$ is the set of extension polynomials with accuracy h corresponding to all disjunctions in $\phi(\psi_1, \dots, \psi_k)$.

Proof. By induction on the complexity of ϕ, using Lemma 6.2(3,4) and Lemma 6.9 for the inductive step. For the application of the latter lemma notice that if $\eta(\psi_1, \dots, \psi_k) = \bigvee_{i \in I} \xi_i$, where η is a subformula of ϕ, then $deg(\xi_i^{\mathbf{ap}}) \leq (\ell+1) \cdot (\max\{p-1, h\}^{\ell}) \cdot \max_i deg(\psi_i^{\mathbf{ap}})$ by Lemma 6.10. □

Lemma 6.12. *For any axiom scheme $\phi(p_1, \dots, p_k)$ of $F(MOD_p)$ and any formulas $\psi_1, \dots, \psi_k$,*

$$\mathcal{E} \vdash_d \ (\phi(\psi_1, \dots, \psi_k))^{\mathbf{ap}},$$

where $d := (h+1)^{O(1)} \cdot \max_i deg(\psi_i^{\mathbf{ap}})$ and $\mathcal{E}$ is the set of extension polynomials with accuracy h corresponding to all disjunctions in $\phi(\psi_1, \dots, \psi_k)$. The constant assumed in the term $(h+1)^{O(1)}$ depends only on the chosen formalization of $F(MOD_p)$.

Proof. An easy inspection shows that both the degree and the depth of all axiom schemes of $F(MOD_p)$ are bounded by some absolute constant (notice that the system F itself has only finitely many axiom schemes). By Lemma 6.11,

$$\mathcal{E} \vdash_d \ \phi^*(\psi_1^{\mathbf{ap}}, \dots, \psi_k^{\mathbf{ap}}) - (\phi(\psi_1, \dots, \psi_k))^{\mathbf{ap}},$$

where $d = (h+1)^{O(1)} \cdot \max_i deg(\psi_i^{\mathbf{ap}})$. On the other hand, ϕ is a tautology, hence ϕ^* is identically 0 on $\{0,1\}^k$ and has therefore a Nullstellensatz proof $\vdash_{d_\phi} \phi^*$ of some degree d_ϕ. Again, it is easy to see that all d_ϕ are bounded by some absolute constant which gives us

$$\mathcal{E}' \vdash_{(h+1)^{O(1)}} \phi^*,$$

where $\mathcal{E}'$ is the set of extension polynomials corresponding to all $\eta(p_1, \ldots, p_k)$. Now we only have to substitute $\psi_i^{\mathbf{ap}}$ for p_i and substract the previous proof. □

After all this preliminary work is done, we can finish the proof of part (1) in Theorem 6.7. Firstly, it is well-known that every Frege proof of length S can be transformed into a tree-like Frege proof of length $S^{O(1)}$ and height $O(\log S)$ (see e.g. Krajíček (1995), Lemma 8.4.8). This transformation works also for $F(MOD_p)$-formulas, and, moreover, the depth of the proof (as defined in Section 1) gets increased only by an absolute additive constant under this transformation. Let π be a tree-like $F(MOD_p)$-refutation of $\phi_1, \ldots, \phi_k$ that has depth $\ell + O(1)$, length $S^{O(1)}$ and height $O(\log S)$.

Let $\mathcal{E}$ be all extension polynomials corresponding to approximations with accuracy $h \geq 1$ of all disjunctions occurring (as subformulas) in π. Clearly, $\mathcal{E}$ is leveled, its depth is at most $\ell + O(1)$ and $|\mathcal{E}| \leq S^{O(1)}$.

By induction on the number of inferences above a formula ϕ in π show that

$$\phi_1^*, \ldots, \phi_k^*, \mathcal{E} \vdash_{(d_0+h(\phi))(h+1)^{O(1)}} \phi^{\mathbf{ap}},$$

where $h(\phi)$ is the height of the subderivation ending with ϕ. For ϕ one of ϕ_i this follows from Lemma 6.11 (with $\psi_i = p_i$). If ϕ is an axiom of $F(MOD_p)$ then the claim follows from Lemmas 6.12 and 6.10. Finally, the claim is preserved for ϕ's obtained by modus ponens: use Lemmas 6.9, 6.10.

The proof of Theorem 6.7 is thus complete. □

Consider in particular typical (and the most interesting) case when $k = n^{O(1)}$ and $d_0 = O(1)$. Then $\phi_1, \ldots, \phi_k$ have constant-depth quasipolynomial size $F(MOD_p)$-refutation if and only if there exists a Nullstellensatz refutation of $\phi_1^*, \ldots, \phi_k^*, \mathcal{E}$ that has degree $(\log n)^{O(1)}$, where $\mathcal{E}$ is a leveled constant-depth quasipolynomial size set of extension polynomials with any prescribed accuracy h chosen from the interval $1 \leq h \leq (\log n)^{O(1)}$.

Note that the simulation of Theorem 6.7 is valid also for unbounded depth Frege systems vs. unbounded depth leveled sets of extension polynomials. Parts (1) and (2) apply literally. For part (3) note that the extension variables can be defined using the extension rule of Extended Frege systems. It is well-known (cf. Cook & Reckhow (1979)) that the presence of the extension rule affects the minimal number of *lines* in a Frege proof by at most a constant factor.

Our next result shows that it is possible to eliminate one block of extension polynomials at the expense of a polynomial increase in degree.

Theorem 6.13. *Assume*

$$\Gamma, \{g_i(1-(\sum_j r_j g_j)) \mid i\} \ \vdash_d \ 1,$$

where Γ *is a set of polynomials. Then*

$$\Gamma \ \vdash_{p(p-2)d^2+3d} \ 1.$$

Proof. Assume

$$\sum_{f\in\Gamma} P_f f \ + \ \sum_i P'_i g_i(1-(\sum_j r_j g_j)) \ + \ \sum_i R_i(x_i^2 - x_i) \ + \ \sum_j R'_j(r_j^p - r_j) = \ 1$$

is a degree d Nullstellensatz refutation. Denote by S the sum $\sum_j r_j g_j$ and by T the sum $S + S^2 + \ldots S^{p-1}$. Multiplying the identity above by T yields a degree pd Nullstellensatz proof of T:

$$\sum_{f\in\Gamma}(P_f T)f \ + \ \sum_i (R_i T)(x_i^2 - x_i) \ + \ \sum_j (R'_j T)(r_j^p - r_j) \ = \ T,$$

where the second term cancels out by Lemma 6.2(1).

Substitute 0 for r_j with $j \neq i$ and g_i^{p-2} for r_i. This increases the degree by a factor of $(p-2)d$ at most. Then S becomes g_i^{p-1} which is an idempotent and thus T becomes just $(p-1)g_i^{p-1}$. Multiplying the resulting equation by $(p-1)g_i$ and using Lemma 6.2(1) again yields

$$\Gamma \ \vdash_{p(p-2)d^2+d} \ g_i$$

for all i. Hence

$$\Gamma \ \vdash_{p(p-2)d^2+2d} \ g_i(1-\sum_j r_j g_j)).$$

Substituting this into the original Nullstellensatz proof yields

$$\Gamma \ \vdash_{p(p-2)d^2+3d} \ 1.$$

□

Since there are examples showing that bounded depth Frege with MOD_p gates is strictly stronger than Nullstellensatz, it is not possible in general to eliminate all extension axioms without too big increase of the degree. One such example is the weak pigeonhole principle PHP_n^{2n} Paris *et al.* (1988), Beame *et al.* (1995), and another example, the "house-sitting tautologies" of Clegg *et al.* (1996) was already mentioned in Section 5: it separates the polynomial calculus from the Nullstellensatz system.

We conclude by observing that it is possible to introduce extension polynomials in an unstructured way, i.e., with no implicit stratification into levels. This appears to be more akin to the fusion method in Boolean complexity, cf. Razborov (1989), Wigderson (1993).

Define an *unstructured extension polynomial of accuracy h* corresponding to $g_1, \ldots, g_h$ to be the polynomial

$$\prod_{u \le h} (g_u - r_u),$$

where all r_u are different and none of them occurs in any g_v (but they may occur arbitrarily in other extension polynomials). The next theorem shows that, on the one hand, this unstructured version is at least as strong as the structured one, and, on the other hand, the analogue of Theorem 6.6 still holds true provided the set $\mathcal{E}$ is not too big.

Theorem 6.14.

(1) *For every extension polynomial $E_{i,\bar{g}}$ of accuracy h there exists an unstructured extension polynomial $E'_{i,\bar{g}}$ of the same accuracy h such that*

$$E'_{i,\bar{g}} \vdash_d \ E_{i,\bar{g}},$$

where $d = (ph+1) \cdot \max_j deg(g_j) + h$.

(2) *Let $f_1, \ldots, f_k, g$ be polynomials from $\mathbf{F}_p[x_1, \ldots, x_n]$, and let $\mathcal{E}$ be any set of unstructured extension polynomials with accuracy h such that $|\mathcal{E}| < e^{h/p}$. If g has a Nullstellensatz proof from $f_1, \ldots, f_k, \mathcal{E}$ then it has one from $f_1, \ldots, f_k$ alone as well.*

Proof. (1) Let

$$f_{iu} = 1 - \sum_{\substack{j \le m \\ j \ne i}} r_{ju} g_j \ (u \le h),$$

then $E_{i,\bar{g}} = g_i \cdot \prod_{u \le h} (f_{iu} - r_{iu} g_i)$. Consider the unstructured extension polynomial

$$E'_{i,\bar{g}} = \prod_{u \le h} (f_{iu} g_i^{p-2} - r_{iu}).$$

Then, clearly,

$$E'_{i,\bar{g}} \vdash_d \ g_i \cdot \prod_{u \le h} (f_{iu} g_i^{p-1} - r_{iu} g_i).$$

On the other hand, $g_i \cdot \prod_{u \le h} (f_{iu} g_i^{p-1} - r_{iu} g_i)$ represents the same function as $E_{i,\bar{g}}$. Hence we can apply Lemma 6.2(1) to finish the proof.

(2) Assign to all r-variables random values from $\mathbf{F}_p$. Then for every *fixed* assignment to x-variables, every extension polynomial from $\mathcal{E}$ gets a non-zero value with probability $\left(1 - \frac{1}{p}\right)^h \le e^{-h/p}$. Thus, every fixed assignment to x-variables can be extended to an assignment reducing all polynomials from $\mathcal{E}$ to zero. Therefore, if this assignment makes all $f_1, \ldots, f_k$ zero, then g gets the value zero as well. □

Note that a bound to the number of unstructured extension polynomials must be a priori present as otherwise the proofs could be unsound. For example, 2^h unstructured extension polynomials $\prod_{u\leq h}(\epsilon_u - r_u)$, $(\epsilon_1, \ldots, \epsilon_h) \in \{0,1\}^h$ do not have common 0-1 zeros and thus entail $1 = 0$ (in degree h). Define a proof system UENS (Unstructured Extended NS) as a system whose proofs are NS-proofs with less than $e^{h/p}$ unstructured extension polynomials. UENS is complete (as NS is so) and sound by the previous theorem. By the remark after Theorem 6.7, every Extended Frege (EF) proof of size S can be transformed into a degree $(\log S)^{O(1)}$ UENS-proof. Amazingly, it is not known whether the converse simulation of UENS by EF takes place, and the reason is that we do not know how to derandomize the soundness proof of UENS (Theorem 6.14(2)). In fact, UENS is one of the extremely rare examples of a propositional proof system that does not allow a "straightforward" simulation by EF, and this makes understanding its exact power an interesting open problem.

Acknowledgements

We are indebted to both anonymous referees of this paper for many useful comments and remarks. S. Buss, R. Impagliazzo, J. Krajíček, P. Pudlák and J. Sgall were partially supported by grant 93 025 of the US-Czechoslovak Science and Technology Program; S. Buss was partially supported by NSF grant DMS 95-03247; R. Impagliazzo was partially supported by NSF YI grant CCR-92-570979, Sloan grant BR-3311 and US-Israel BSF grant 92-00043; J. Krajíček, P. Pudlák and J. Sgall were partially supported by grants A119107 and A1019602 of the AV ČR; and A. A. Razborov was partially supported by RBRF grant 93-011-16015 and by an AMS-FSU grant.

References

M. Ajtai, The complexity of the pigeonhole principle. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, 1988, 346–355.

M. Ajtai, Parity and the pigeonhole principle. In *Feasible Mathematics*, ed. S. R. Buss and P. J. Scott, 1–24. Birkhauser, 1990.

M. Ajtai, The independence of the modulo p counting principle. In *Proceedings of the 26th ACM STOC*, 1994, 402–411.

P. Beame, S. Cook, J. Edmonds, R. Impagliazzo, and T. Pitassi, The relative complexity of NP search problems. In *Proceedings of the 27th ACM STOC*, 1995, 303–314.

P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, and P. Pudlák, Lower bounds on Hilbert's Nullstellensatz and propositional proofs. In *Proceedings of the 35th IEEE FOCS*, 1994, 794–806. Journal version to appear in *Proc. of the London Math. Soc.*

P. Beame and T. Pitassi, Exponential separation between the matching principles and the pigeonhole principle. Submitted to *Annals of Pure and Applied Logic*, 1993.

P. Beame and S. Riis, More on the relative strength of counting principles. To appear in *Proceedings of the DIMACS workshop on Feasible Arithmetic and Complexity of Proofs*, 1996.

S. Bellantoni, T. Pitassi, and A. Urquhart, Approximation of small depth Frege proofs. *SIAM Journal on Computing* **21**(6) (1992), 1161–1179.

M. Bonet, T. Pitassi, and R. Raz, Lower bounds for cutting planes proofs with small coefficients. In *Proceedings of the 27th ACM STOC*, 1995, 575–584.

Samuel R. Buss, A new exposition of the design for the housesitting principle. Manuscript, 1995a.

Samuel R. Buss, Some remarks on lengths of propositional proofs. *Archive for Mathematical Logic* **34** (1995b), 377–394.

V. Chvátal and E. Szemerédi, Many hard examples for resolution. *Journal of the ACM* **35**(4) (1988), 759–768.

M. Clegg, J. Edmonds, and R. Impagliazzo, Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the 28th ACM STOC*, 1996, 174–183.

S. A. Cook and A. R. Reckhow, The relative efficiency of propositional proof systems. *Journal of Symbolic Logic* **44**(1) (1979), 36–50.

A. Haken, The intractability or fesolution. *Theoretical Computer Science* **39** (1985), 297–308.

Johan Hastad, Almost optimal lower bounds for small depth circuits. In *Randomness and Computation* (*Advances in Computing Research, Vol. 5*), ed. S. Micali, 143–170. JAI Press, 1989.

R. Impagliazzo, The sequential ideal generation proof system. Unpublished note, 1995.

J. Krajíček, Lower bounds to the size of constant-depth propositional proofs. *Journal of Symbolic Logic* **59**(1) (1994a), 73–86.

J. Krajíček, Interpolation theorems, lower bounds for proof systems and independence results for bounded arithmetic. To appear in *Journal of Symbolic Logic*, 1994b.

J. Krajíček, *Bounded arithmetic, propositional logic and complexity theory*. Encyclopedia of Mathematics and Its Applications, Vol. **60**, Cambridge University Press, 1995.

J. Krajíček, A fundamental problem of mathematical logic. *Annals of the Kurt Gödel Society*, Collegium Logicum, Vol. **2**, Springer-Verlag, 1996, 56–64.

J. Krajíček, On methods for proving lower bounds in propositional logic. In: *Logic and Scientific Methods* Eds. M. L. Dalla Chiara et al., (Vol. 1 of Proc. of the Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence (August 19-25, 1995)), Synthese Library, Vol. **259**, Kluwer Academic Publ., Dordrecht, 1997, pp.69-83.

J. Krajíček, P. Pudlák, and A. R. Woods, Exponential lower bounds to the size of bounded depth Frege proofs of the pigeonhole principle. *Random Structures and Algorithms* **7**(1) (1995), 15–39.

J. B. Paris, A. J. Wilkie, and A. R. Woods, Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic* **53**(4) (1988), 1235–1244.

T. Pitassi, Algebraic Propositional Proof Systems. To appear in the proceedings volume of the DIMACS workshop on *Finite Models and Descriptive Complexity*, held January 14-17, 1996.

T. Pitassi, P. Beame, and R. Impagliazzo, Exponential lower bounds for the pigeonhole principle. *Computational Complexity* **3** (1993), 97–140.

P. Pudlák, The lengths of proofs. To appear in *Handbook of Proof Theory*, 1995a.

P. Pudlák, Lower bounds for resolution and cutting planes proofs and monotone computations. To appear in *Journal of Symbolic Logic*, 1995b.

А. А. Разборов, Нижние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения. *Матем. Зам.* **41**(4) (1987), 598–607. A. A. Razborov, Lower bounds on the size of bounded-depth networks over a complete basis with logical addition, *Mathem. Notes of the Academy of Sci. of the USSR*, 41(4):333-338, 1987.

A. A. Razborov, On the method of approximation. In *Proceedings of the* 21*st ACM Symposium on Theory of Computing*, 1989, 167–176.

A. Razborov, Lower bounds for the polynomial calculus. Submitted to *Computational Complexity*, 1996.

R. A. Reckhow, On the lengths of proofs in the propositional calculus. Technical Report 87, University of Toronto, 1976.

S. Riis, *Independence in Bounded Arithmetic.* PhD thesis, Oxford University, 1993.

S. Riis, Count(q) does not imply Count(p). Technical Report RS-94-21, Basic Research in Computer Science Center, Aarhus, Denmark, 1994.

R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th ACM Symposium on Theory of Computing*, 1987, 77–82.

Г. С. Цейтин, О сложности вывода в исчислении высказываний. In *Исследования по конструктивной математике и математической логике,* II*; Записки научных семинаров ЛОМИ, т. 8*, ed. А. О. Слисенко, 234–259. Наука, Ленинград, 1968. Engl. translation: G. C. Tseitin, On the complexity of derivations in propositional calculus, in: *Studies in mathematics and mathematical logic, Part II*, ed. A. O. Slissenko, pp. 115-125.

A. Urquhart, Hard examples for resolution. *Journal of the ACM* **34**(1) (1987), 209–219.

A. Urquhart, The complexity of propositional proofs. *Bulletin of Symbolic Logic* **1** (1995), 425–467.

A. Wigderson, The fusion method for lower bounds in circuit complexity. In *Combinatorics, Paul Erdos is Eighty (Vol. 1)*, 453–468. 1993.

Samuel R. Buss
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
sbuss@ucsd.edu

Russell Impagliazzo
Computer Science Engineering
University of California, San Diego
La Jolla, CA 92093-0112, USA
russell@cs.ucsd.edu

Jan Krajíček, Pavel Pudlák
and Jiří Sgall
Mathematical Institute, Academy of Sciences
Žitná 25, 115 67, Prague, Czech Republic
{krajicek,pudlak,sgallj}@math.cas.cz

Alexander A. Razborov
Steklov Mathematical Institute
Academy of Sciences
Gubkina 8, 117966, Moscow, Russia
razborov@genesis.mi.ras.ru