# LOWER BOUNDS FOR THE POLYNOMIAL CALCULUS AND THE GROEBNER BASIS ALGORITHM

Russell Impagliazzo, Pavel Pudlák and Jiří Sgall

**Abstract.**

Razborov (1996) recently proved that polynomial calculus proofs of the pigeonhole principle $PHP_n^m$ must have degree at least $\lceil n/2 \rceil + 1$ over any field. We present a simplified proof of the same result.

Furthermore we show a matching upper bound on the polynomial calculus proofs of the pigeonhole principle for any field of sufficiently large characteristic, and an $\lceil n/2 \rceil + 1$ lower bound for any subset sum problem over the field of reals.

We show that these degree lower bounds also translate into lower bounds on the number of monomials in any polynomial calculus proof, and hence on the running time of most implementations of the Groebner Basis Algorithm.

**Subject classifications.** 68Q99; 68Q05, 03B05.

## 1. Introduction

**1.1. Motivation.** There is a duality between proof systems and heuristics for $NP$-hard problems. On the one hand, any valid (non-deterministic) algorithm that solves the unsatisfiability problem for Boolean formulas can be viewed as a propositional proof system, since a negative run of the algorithm certifies the formula is unsatisfiable. On the other hand, propositional proof systems are frequently used as the basis for heuristic algorithms that attempt to find small proofs of unsatisfiability to restrict searches. For example, the Putnam-Davis procedure to decide satisfiability of $CNF$ formulas is based on the resolution proof system.

The polynomial calculus is a proof system based on algebraic reasoning, introduced under the name "Groebner proofs" by Clegg *et al.* (1996) and further studied by Buss *et al.* (1997) and Razborov (1996). It generalizes the Nulstellensatz system introduced by Beame *et al.* (1996) and can be simulated by bounded-depth Frege systems with modular gates (Buss *et al.* 1997).

Besides its naturalness as a proof system and connection to bounded-depth Frege systems, the polynomial calculus has the same role for (most) implementations of the widely used Groebner basis algorithm as resolution does for implementations of Davis-Putnam.

A Groebner basis algorithm is a method for producing a normal form for a system of polynomial equations. There are many varieties, but all use the same basic operations, $S$*-remainders* and *reductions*, that can be simulated in the polynomial calculus. Thus the total time for (almost) any implementation of Groebner basis algorithm is bounded from below by the total size of a polynomial calculus proof. (This lower bound would not necessarily pertain to an implementation of Groebner basis algorithm that uses *lazy evaluation* to implement these operations instead of explicitly computing the resulting polynomial. To bound such an implementation, one should prove lower bounds in a richer version of polynomial calculus where algebraic formulas were allowed as lines. Such a proof system is discussed by Buss *et al.* 1997.) As a partial converse, Clegg *et al.* (1996) show that many versions of Groebner Basis algorithm have a running time that can be bounded by a polynomial in $n^d$, where $n$ is the number of variables and $d$ is the *degree* of a Groebner proof. However, there are some examples of (non-Boolean) polynomials where the Groebner basis algorithm is polynomial-time, but the minimum degree of a polynomial calculus refutation is super-constant. Thus bounding the degree does not give the complete answer, but it remains the most important parameter.

Our new proof of the lower bound on the degree of polynomial calculus proofs of the pigeon-hole principle is motivated by the current research in the complexity of propositional proof systems. The pigeonhole principle has been proven to be a hard tautology for other weak propositional proof systems, like the bounded depth Frege system (see Ajtai 1988, Krajíček 1994, Krajíček *et al.* 1995, and Pitassi *et al.* 1993), Nullstellensatz proof system (see Beame *et al.* 1996 and Beame & Riis 1998), therefore it is a natural candidate for the polynomial calculus, too. Since polynomial calculus generalizes Nullstellensatz, the lower bound for polynomial calculus is a generalization of the earlier results, and it is a natural step towards proving lower bounds on the stronger system of bounded depth proofs with a modular connective (Buss *et al.* 1997).

Another important combinatorial principle is the counting principle modulo $m$. Lower bounds for counting principle modulo $q$ in the constant depth Frege system with an axiom for counting principle modulo $m$ (not divisible by a prime (dividing) $q$) have been given by Ajtai (1994). Beame *et al.* (1996) and Buss *et al.* (1997) derived these and stronger (exponential) lower bounds from analogous bounds on the degree of proofs in the Nullstellensatz proof systems, and Beame & Riis (1998) give lower bounds on the pigeonhole principle in the constant depth Frege systems with an axiom for counting principle modulo $m$. Our methods do not seem to give lower bounds for polynomial calculus proofs of counting principles, however such bounds have been recently proved by Krajíček (1997a, 1997b) and improved by Buss, Impagliazzo, Pitassi (personal communication).

**1.2. Our Results.** Razborov (1996) recently proved that polynomial calculus proofs of the pigeonhole principle $PHP_n^m$ must have degree at least $\lceil n/2 \rceil + 1$ over any field. We present a simplified proof of the same result (Section 3). The main idea of our proof is the same as in the original proof of Razborov: we want to describe explicitly the vector space of the polynomials derivable in a low degree polynomial calculus refutation of the pigeonhole principle, and the description uses the pigeon dance as before. We are able to avoid some of the technical machinery, due to the simple counting argument which shows that the set of polynomials, which generates the vector space of consequences, forms its basis.

Furthermore we show a matching upper bound on the polynomial calculus proofs of the pigeonhole principle for any field of sufficiently large characteristic (Section 4), and an $\lceil n/2 \rceil + 1$ lower bound for any subset sum problem over the field of reals (Section 5).

We show that the degree lower bounds also translate into lower bounds on the number of monomials in any polynomial calculus proof, and hence on the running time of most implementations of the Groebner Basis Algorithm (Section 6).

The results in Sections 4 to 6 were obtained independently of Razborov's work (Razborov 1996), while the degree lower bound for arbitrary field from Section 3 was obtained later, and uses the principal ideas from Razborov's paper. A preliminary version of this paper appeared as Impagliazzo *et al.* (1997).

## 2. Preliminaries

DEFINITION 2.1. (Clegg *et al.* (1996)) *Given a field $K$ and a set of variables,* a polynomial calculus refutation *of the set of axioms $P$ is a sequence of polynomials such that the last line is the polynomial 1 and each line is either an axiom or is derived from the previous lines using the following inference rules:*

$$\frac{f \qquad g}{\alpha f + \beta g}$$

*and*

$$\frac{f}{x \cdot f}$$

*where $\alpha, \beta \in K$ are any scalars and $x$ is any variable. The refutation has degree $d$ if all the polynomials in it have degree at most $d$.*

We assume that polynomials $x^2 - x$ are included in the axioms for all variables $x$. Then the axioms $f_1, \ldots, f_k$ are refutable if and only if the system $f_1 = f_2 = \cdots = f_k = 0$ has no 0-1 solutions. The question we are interested in is whether there exists a refutation of a small degree.

The special case of Nullstellensatz refutations can be viewed as polynomial calculus refutations where all multiplication inferences precede all addition inferences.

The set $\{1, \ldots, i\}$ is denoted by $[i]$. For a polynomial $f$, let $\bar{f}$ be the unique multilinear polynomial equal to it modulo the ideal generated by all the polynomials $x^2 - x$. In this paper, all the lower bounds show that in fact any refutation of some initial polynomials has to contain a polynomial $f$ with large degree of $\bar{f}$.

Now we define the pigeonhole principle, in the usual form used for lower bounds in propositional calculus. The variables are $x_{ij}$, $i \in [m]$, $j \in [n]$. The assignment $x_{ij} = 1$ represents the fact that the pigeon $i$ is sitting in the hole $j$.

DEFINITION 2.2. *Let*

$$Q_i = 1 - \sum_{j \in [n]} x_{ij}.$$

*The (negation of the) pigeonhole principle $\neg PHP_n^m$ is the following set of polynomials:*

$$\begin{array}{rl} Q_i & \text{for } i \in [m] \\ x_{ij}x_{ij'} & \text{for } i \in [m],\ j, j' \in [n],\ j \neq j' \\ x_{ij}x_{i'j} & \text{for } i, i' \in [m],\ j \in [n],\ i \neq i' \end{array}$$

## 3. A lower bound for PHP over any field

For this section we fix $m$, $n$, and an arbitrary field $K$.

DEFINITION 3.1. *Let $T$ be the set of all monomials $x_{i_1 j_1} \cdots x_{i_k j_k}$ such that all $i_l$ are distinct and all $j_l$ are distinct, and let $T_d$ be the set of monomials in $T$ of degree at most $d$.*

Using the identities $x_{ij}x_{ij'} = 0$, $x_{ij}x_{i'j} = 0$, and $x_{ij}^2 = x_{ij}$, we can represent any polynomial as a linear combination of terms from $T$, without increasing the degree. Therefore any polynomial calculus refutation of $\neg PHP_n^m$ can be transformed into a refutation which uses only the axioms $Q_i$ and simplifications by the previous identities (which is equivalent to the use of the corresponding axioms). This is equivalent to computing in the ideal $I$ generated by the polynomials $x_{ij}x_{ij'}$, $x_{ij}x_{i'j}$, and $x_{ij}^2 - x_{ij}$, which has a basis $T$ as a vector space over $K$. From now on we assume that this transformation is performed, hence all the polynomials are in $span(T)$, and all the computations are modulo the ideal $I$.

Our goal is to construct a base $B_d$ of the vector space generated by $T_d$ such that the basis elements are products of some variables and some axioms $Q_i$ (e.g., $x_{3,1}x_{5,3}Q_2Q_4$). If we express all the lines of the proof in this basis, it will become clear that 1 cannot be derived from the initial polynomials $Q_i$. More precisely, we give an explicit subset of $B_d$ such that its linear closure contains (exactly) all low degree consequences of the axioms $Q_i$, but not 1.

The definition of the basis $B_d$ uses the pigeon dance, which was introduced by Razborov (1996). The proof that $B_d$ spans the whole space is based on rewriting the terms using the pigeon dance, as in Razborov (1996), Claim 3.4. To prove that the elements $B_d$ are linearly independent, we exhibit a one-to-one map from $B_d$ into $T_d$, which can be viewed as a generalized pigeon dance. This part of the proof is new, and avoids the more complicated parts of the original proof in Razborov (1996).

We now proceed to define a convenient common notation for the elements of the old and new bases. Elements of $T_d$ correspond to the partial one-to-one mappings from $[m]$ to $[n]$ of size at most $d$. We extend this representation, so that the product of some variables $x_{ij}$ and axioms $Q_i$ is represented by a partial mapping from $[m]$ to $\{0, \ldots, n\}$, where the value 0 on $i$ corresponds to $Q_i$ in the product. Since we allow more axioms $Q_i$ in the product, we require the mapping to be one-to-one only on the part of domain not mapped to 0.

Definition 3.2. *Let*

$$A = \{a \mid a \text{ is a partial function from } [m] \text{ to } \{0,1,\ldots,n\} \text{ such that } \forall i\,(a(i) = a(i') \neq 0 \Rightarrow i = i')\}.$$

$$A_d = \{a \in A \mid |a| \leq d\}$$

*For* $a = \{(i_1,j_1),\ldots,(i_k,j_k),(i'_1,0),\ldots,(i'_l,0)\} \in A$ *such that* $j_1,\ldots,j_k \neq 0$ *we define*

$$\hat{a} = \{(i_1,j_1),\ldots,(i_k,j_k)\}$$

$$x_a = x_{i_1 j_1} \cdots x_{i_k j_k} Q_{i'_1} \cdots Q_{i'_l}.$$

*(If* $a$ *is the empty function,* $x_a = 1$*. Clearly for any* $a \in A$ *the degree of* $x_a$ *is equal to* $|a| = |dom(a)|$*.)*

Given the assignment of pigeons to holes $a \in A$ we define the pigeon dance as the following procedure. We take the first pigeon and move him to some currently unoccupied hole larger than the one he is sitting in, then do the same with the second pigeon, and so on, until the last one. This procedure in general is not unique, and sometimes it is impossible to finish a pigeon dance at all. Intuitively, the best strategy for the pigeon dance is to move always to the closest unoccupied hole. After the formal definition we indeed prove that this is the case.

Definition 3.3. *Let* $a \in A$ *be given. We define* a pigeon dance *on* $a$ *to be any sequence* $b_0 = a$, $b_1$, $\ldots$, $b_m$ *of elements of* $A$ *with the same domain as* $a$ *such that for* $1 \leq t \leq m$, $b_t(i) = b_{t-1}(i)$ *for any* $i \in dom(a) - \{t\}$ *and* $b_t(t) > b_{t-1}(t)$ *if* $t \in dom(a)$*. (In particular,* $b_t = b_{t-1}$ *if* $a(t)$ *is undefined.)*

*Let* $a \in A$, $t \in [m]$*. We define* $D_t(a)$ *to be a function* $b \in A$ *such that* $dom(a) = dom(b)$, $b(i) = a(i)$ *for* $i \in dom(a) - \{t\}$*, and if* $t \in dom(a)$ *then* $b(t) = j$ *where* $j$ *is the smallest number such that* $a(t) < j \leq n$ *and* $j \notin rng(a)$*; if no such* $j$ *exists,* $D_t(a)$ *is not defined. We define* the minimal pigeon dance *on* $a$ *to be*

$$D(a) = D_m(D_{m-1}(\ldots D_1(a)\ldots)).$$

Lemma 3.4. (Razborov (1996)) *If there exists a pigeon dance on* $a \in A$*, then* $D(a)$ *is defined.*

Proof. Let $b_0 = a$, $b_1$, $\ldots$, $b_m$ be a pigeon dance on $a$ such that the first $t-1$ steps are the same as in the minimal dance, i.e., for $1 \leq i < t$, $b_i = D_i(b_{i-1})$. We prove that then $D_t(b_{t-1})$ is also defined and there exists a pigeon dance

$c_0 = a$, $c_1$, ..., $c_m$ such that the first $t$ steps are the same as in the minimal dance. By induction, this is sufficient to conclude that $D(a)$ is defined.

Let $j$ be the minimal $j > b_{t-1}(t)$ such that $j \notin rng(b_{t-1})$, and let $j' = b_t(t)$. Such $j$ exists and satisfies $j \leq j'$, since $b_t$ is a step in a pigeon dance, and thus $b_{t-1}(t) < b_t(t) = j' \notin rng(b_{t-1})$. If $j = j'$, we are done. Otherwise $j < j'$ and we define the new dance as follows. For $i < t$, $c_i = b_i$. For $i \geq t$,

$$c_i(i') = \begin{cases} j & \text{if} \quad b_i(i') = j' \\ j' & \text{if} \quad b_i(i') = j \\ b_i(i') & \text{otherwise.} \end{cases}$$

Note that the first possibility, $b_i(i') = j'$, is true if and only if $i \geq t$ and $i' = t$. It is easy to verify that $c$ is a pigeon dance and the first $t$ steps are the same as in the minimal dance. □

DEFINITION 3.5. (THE BASIS $B_d$)

$$B_d = \{x_a \mid a \in A_d \textit{ and there exists a pigeon dance on } \hat{a}\}$$

From the definition of $B_d$ it is obvious that $B_{d-1} \subseteq B_d$, and $x_a Q_i \in B_d$ if and only if $x_a \in B_{d-1}$ and $i \notin dom(a)$. This monotonicity is very important later in the proof. The fact that a basis with a similar property does not exist for other principles, like the counting principles or onto pigeonhole principle, makes it difficult to prove lower bounds in those cases by similar technique.

Next we proceed to prove that the minimal pigeon dance $D$ is defined on whole $B_d$ for a small degree $d$ (i.e., for every $a \in A_d$, $D$ is defined on $a$ whenever it is defined on $\hat{a}$), and maps it one-to-one into $T_d$ (in fact, its inverse is essentially the minimal dance with the holes numbered in the opposite direction).

LEMMA 3.6. *If $d \leq \lceil n/2 \rceil$ and $a \in A_d$ there exists a pigeon dance on $a$ if and only if there exists a pigeon dance on $\hat{a}$.*

PROOF. Let $x = |\hat{a}|$, let $y$ be the number of pigeons assigned to the hole 0 by $a$. If $y = 0$, there is nothing to prove. Otherwise $y + 2x \leq n$. Fix any pigeon dance on $\hat{a}$. It uses at most $2x$ different holes, hence there are at least $y$ holes among $\{1, \ldots, n\}$ that are not used at any time during the dance on $\hat{a}$. The pigeon dance on $a$ proceeds as follows: on its turn each pigeon sitting in the hole 0 is moved to one of the $y$ unused holes which is still empty; all other pigeons move as in the dance on $\hat{a}$. The other direction is trivial. □

LEMMA 3.7. *The minimal pigeon dance $D$ is a one-to-one mapping on its domain.*

PROOF. It is sufficient to prove for any $t \in [m]$ that $D_t$ is one-to-one on its domain, as $D$ is a composition of these mappings. Suppose $D_t(a) = D_t(a') = b$ are both defined. Then according to the definition of $D_t$, $dom(a) = dom(b) = dom(a')$ and $a(i) = b(i) = a'(i)$ for any $i \neq t$ in the domain. It cannot be the case that $a(t) < a'(t)$, since then the hole $a'(t)$ is unoccupied in $a$ and $(D_t(a))(t) \leq a'(t) < (D_t(a'))(t)$. The case $a(t) > a'(t)$ is symmetric. Assuming $D_t(a) = D_t(a')$ we proved that $a = a'$. Hence $D_t$ is one-to-one. □

PROPOSITION 3.8. *For any $d \leq \lceil n/2 \rceil$, the set $B_d$ is a basis of the vector space generated by $T_d$.*

PROOF. By Lemma 3.6, $D(a)$ is defined for any $x_a \in B_d$. Clearly, $x_{D(a)} \in T_d$, since after any pigeon dance no pigeon is sitting in the hole 0. By Lemma 3.7, $D$ is one-to-one, and hence $|B_d| \leq |T_d|$. Since $T_d$ is a basis of $A_d$, its elements are linearly independent. It remains to prove that any $x_a \in T_d$ can be expressed as a linear combination of elements of $B_d$. This is established in Razborov (1996), Claim 3.4; we include the proof for completeness.

We define a variation of the lexicographic ordering on $T_d$: for $x_a, x_b \in T_d$ we put $a \prec b$ if and only if $dom(a) \subset dom(b)$ or ($dom(a) = dom(b)$ and for the largest $i$ such that $a(i) \neq b(i)$ we have $a(i) < b(i)$). (Note that this is not a linear ordering, since we do not compare elements with incomparable domains.)

Suppose that for all $a' \prec a$, $x_{a'} \in span(B_d)$. We want to prove that $x_a \in span(B_d)$. If there exists a pigeon dance on $a$, $x_a \in B_d$ and we are done. Otherwise let $P_t$ be the set of all possible results of the first $t$ steps of the pigeon dance on $a$. We prove by induction on $t$ that

$$x_a \in span(B_d) \qquad \text{iff} \qquad \sum_{b \in P_t} x_b \in span(B_d).$$

This proves that $x_a \in span(B_d)$, since there is no complete dance on $a$ and hence $P_m = \emptyset$. The basis of the induction holds, since $P_0 = \{a\}$. If $t \notin dom(a)$ then $P_t = P_{t-1}$ and the step is trivial. Otherwise for each $b \in P_{t-1}$ we express $x_b$ as $x_{tj} x_c$ and rewrite it as

$$x_b = (1 - Q_t - \sum_{j' \neq j} x_{tj'}) x_c = x_c - x_c Q_t - \sum_{j' < j} x_{tj'} x_c - \sum_{j' > j} x_{tj'} x_c.$$

If $j' \in rng(c)$, then the corresponding term is 0. All the remaining terms except those in the last sum are in $span(B_d)$ by the induction assumption: for the second term we use the fact that $x_c \in span(B_{d-1})$ implies that $x_c Q_i \in span(B_d)$, and for $j' < j$ we use the fact that $b(i) = a(i)$ for $i \geq t$ and hence $x_{tj'} x_c \prec x_a$. The last term corresponds to all possible $t$th steps of the pigeon dance on $a$, and hence summing over all $b \in P_{t-1}$ we obtain that $\sum_{b \in P_t} x_b \in span(B_d)$ if and only if $\sum_{b \in P_{t-1}} x_b \in span(B_d)$. □

THEOREM 3.9. *$PHP_n^m$ has no polynomial calculus refutation of degree $d \leq \lceil n/2 \rceil$.*

PROOF. We prove by induction on the length of the proof that any polynomial derivable from the polynomials $Q_i$ in degree $d$ is a linear combination of polynomials in $B_d - T_d$ (i.e., it is a linear combination of such elements of $B_d$ that are multiples of some axiom $Q_i$). This finishes the proof, since $1 \in T_d$, and hence it cannot be derived (here we use the fact that $B_d$ is a basis and hence the representation of 1 is unique).

For the axioms the claim is true, since $Q_i \in B_d - T_d$ for any $i$. The case of the addition rule is trivial. For the multiplication rule we need to prove that for any $x_{ij}$ and $x_a \in B_d - T_d$, $|a| = d - 1$, the polynomial $x_{ij} x_a$ is a linear combination of elements from $B_d - T_d$. From the definition of $B_d$ it follows that $x_a$ can be written as $x_a = x_b Q_{i'}$ for some $b \in B_{d-2}$ and $i' \in [m]$. Now express $x_{ij} x_b$ as a linear combination of elements $x_{c_\alpha}$ from the basis $B_{d-1}$. Multiply this expression by $Q_{i'}$ to obtain a linear combination of elements $x_{c_\alpha} Q_{i'}$; all these elements are either 0 if $i' \in dom(c_\alpha)$, or are in $B_d$ otherwise. □

## 4. Upper bounds over the reals

In this section we prove a matching upper bound of $\lceil n/2 \rceil + 1$ for the proofs of PHP in all the cases where the field is large enough. First we show that a refutation with this degree exists for the polynomial $m - \sum_{j=1}^{n} y_j$, and then we reduce PHP to it. Note that this formula is also a special case of subset sum, for which we prove a lower bound in the next section.

LEMMA 4.1. *Suppose that there exists a derivation of $g$ from a set of polynomials $P$ such that every non-zero monomial in the derivation has at most $b$ distinct variables. Then there exists a derivation of $g$ from $P$ of degree* $\max(d, b + 1)$, *where $d$ is the maximal degree of $g$ and $P$. In particular, if $g$ is any consequence of $P$ and $N$ is the total number of variables, there exists a derivation of $g$ from $P$ of degree* $\max(d, N + 1)$.

Proof. For any polynomial $f$, we can derive $\bar{f} - f$ using the axioms $x^2 - x$ in degree equal to the degree of $f$. We modify the original refutation so that each line $f$ is replaced by $\bar{f}$. The inferences by linear combinations are still valid. When $f$ is multiplied by $x$ in the original proof, we first multiply $x \cdot \bar{f}$, and then adding a result of an auxiliary derivation of $\overline{x \cdot \bar{f}} - x \cdot \bar{f}$ we obtain $\overline{x \cdot \bar{f}} = \overline{x \cdot f}$. Also, at the axioms and at the end of the proof we add auxiliary derivations of $\bar{f} - f$ as needed. The degree is as claimed, since the degree of any multilinear polynomial in the new derivation is at most $b$, and the largest degree of a polynomial $x \cdot \bar{f}$ is thus $b + 1$. The second assertion follows since polynomial calculus is complete, and clearly every monomial has at most $N$ variables. □

Theorem 4.2. *Assuming $m > n$ and $m$ is larger that the characteristic of the given field, the polynomial $m - \sum_{j=1}^{n} y_j$ has a refutation of degree $\lceil n/2 \rceil + 1$.*

Proof. Let $Y = \sum_{j=1}^{\lceil n/2 \rceil} y_j$, and $X = \sum_{j=\lceil n/2 \rceil + 1}^{j=n} y_j$; the polynomials $X$ and $Y$ are only shorthands for the corresponding experessions, not new variables. The refutation starts by deriving the polynomials $p$ and $q$ and then uses $q$ and $X + Y = m$ to derive $r$, where $p$, $q$, and $r$ are defined below. Intuitively $p$ asserts that the sum of the first half of the variables is between 0 and $\lceil n/2 \rceil$, $q$ asserts the same for the second half, and $r$ expresses that the sum is between $m - \lfloor n/2 \rfloor$ and $m$:

$$\begin{aligned} p &= Y(Y-1)(Y-2)\cdots(Y - \lceil n/2 \rceil) \\ q &= X(X-1)(X-2)\cdots(X - \lfloor n/2 \rfloor) \\ r &= (Y-m)(Y-m+1)\cdots(Y - m + \lfloor n/2 \rfloor) \end{aligned}$$

The polynomial $p$ is a consequence of the axioms $y_j^2 - y_j$ for the first $\lceil n/2 \rceil$ variables, and hence by Lemma 4.1 it has a proof of degree $\lceil n/2 \rceil + 1$. Similarly, $q$ is a consequence of the axioms for the second half of the variables, so is derivable by a proof of the same degree. The polynomial $r$ is derived by substituting $X = m - Y$ term by term into $q$, since $X + Y = m$ is an axiom. More precisely, we derive $X^t - (m - Y)^t$ for $t = 1, \ldots, \lfloor n/2 \rfloor + 1$ inductively as $X^{t-1}(X + Y - m) + (m - Y)(X^{t-1} - (m - Y)^{t-1})$. Then we multiply out $q$ as a degree $\lfloor n/2 \rfloor + 1$ polynomial in $X$, and replace each power of $X$ by the corresponding power of $m - Y$. The result is either $r$ or $-r$.

From the assumption on $m$ it follows that $m \notin \{0, 1, \ldots, n\}$ in the given field, and hence the polynomials $p$ and $r$ in the first $\lceil n/2 \rceil$ variables are contradictory. Therefore there exists a refutation from them, and by Lemma 4.1 it has degree $\lceil n/2 \rceil + 1$. □

THEOREM 4.3. *Assuming $m > n$ and $m$ is larger that the characteristic of the given field, then $\neg PHP^m_n$ has a refutation of degree $\lceil n/2 \rceil + 1$. In particular, $\neg PHP^m_n$ has a refutation of degree $\lceil n/2 \rceil + 1$ over the reals for any $m > n$.*

PROOF. Denote $y_j = \sum_{i=1}^m x_{ij}$, $j \in [n]$. The refutation of PHP is constructed as follows. First, using the axioms $x_{ij}x_{i'j}$ and $x_{ij}^2 - x_{ij}$ we derive $y_j^2 - y_j$ for all $j \in [n]$; this part has degree 2. Then we derive $m - \sum_{j=1}^n y_j$ as the sum of all the axioms $Q_i$, $i \in [m]$; this part has degree 1. Finally, using Theorem 4.2 we derive the polynomial 1. The last part has degree $\lceil n/2 \rceil + 1$ even after the substitution $y_j = \sum_{i=1}^m x_{ij}$, since the substitution is linear. □

## 5. A lower bound for subset sum over the reals

In this section we prove a simple lower bound on the degree of polynomial refutations in the case of the field of real numbers. A special case of this lower bound for $c_1 = \ldots = c_n = 1$ gives a matching bound to Theorem 4.2 for the field of reals.

THEOREM 5.1. *Let $c_1, \ldots, c_n$ be nonzero reals, $m$ an arbitrary real. Then $m - \sum_{i=1}^n c_i x_i$ has no refutation of degree $\lceil n/2 \rceil$ in the field of real numbers. (If $m$ is the sum of a subset of $c_1, \ldots, c_n$, then, of course, there is no such refutation of any degree.)*

PROOF. The combinatorial content of the proof is a well-known result about incidence matrices. For every $k < n$ define a matrix $\mathbf{D}_k^n$ as follows. The rows of $\mathbf{D}_k^n$ are indexed by the $k$ element sets $A \subseteq [n]$, the columns are indexed by $k+1$ element sets $B \subseteq [n]$, and the entry $(A, B)$ is 1 if $A \subseteq B$ and 0 otherwise. It has been shown that the matrix has full rank, see Gottlieb (1966).[1] For $k < n/2$ this means that the rows of the matrix are linearly independent, as the number of rows is at most the number of columns.

LEMMA 5.2. *Let $p(x_1, \ldots, x_n)$ be a nonzero real polynomial of degree less than $\lceil n/2 \rceil$. Then the degree of $\overline{(m - \sum_{i=1}^n c_i x_i) \cdot p(x_1, \ldots, x_n)}$ is $deg(p) + 1$.*

PROOF. Let the degree of $p$ be $k$. Let $\mathbf{C}$ be a matrix indexed as $\mathbf{D}_k^n$ such that the entry $\mathbf{C}_{(A,B)}$ is $c_i$ if $A \cup \{i\} = B$ and 0 otherwise. The matrix $\mathbf{C}$ can be obtained from $\mathbf{D}_k^n$ by multiplying each row $A$ by $\prod_{i \in A} c_i^{-1}$ and multiplying

[1] The result has been rediscovered by several several people, including the authors of this paper. We are indebted to Nati Linial for the reference.

each column $B$ by $\prod_{i\in B} c_i$. Since all the numbers $c_i$ are nonzero, the rows of $\mathbf{C}$ are linearly independent.

Suppose that the degree of $\overline{(m - \sum_{i=1}^{n} c_i x_i) \cdot p(x_1, \ldots, x_n)}$ is less than $k+1$. Thus the monomials of degree $k+1$ cancel in this product. These monomials result from multiplying the monomials of $p$ of degree $k$ by monomials $c_i x_i$. Consider a monomial $\prod_{i\in A} x_i$, where $|A| = k$. If we multiply $\sum_{i=1}^{n} c_i x_i$ by this monomial, the resulting monomials of degree $k+1$ are $\mathbf{C}_{(A,B)} \prod i \in B x_i$, over all $B$. Thus if all monomials of degree $k+1$ cancel, we have a nontrivial linear dependency among the rows of $\mathbf{C}$, a contradiction. $\square$

The proof of the theorem now follows easily from the above lemma. First we get by induction that every polynomial derivable by a proof of degree $\leq \lceil n/2 \rceil$ is (modulo the ideal generated by $x_i^2 - x_i$) of the form $(m - \sum_{i=1}^{n} c_i x_i) \cdot p(x_1, \ldots, x_n)$ with $deg(p) \leq \lceil n/2 \rceil - 1$. Applying the lemma once again, we see that 1, which is a polynomial of degree 0, is not derivable by proofs of degree $\leq \lceil n/2 \rceil$. $\square$

## 6. Lower bounds on the number of monomials

In this section, we show that for any instance of subset sum over the reals, and the pigeonhole principle over any field, an exponential number of monomials appear in any polynomial calculus refutation. Since most implementations of the Groebner basis algorithm represent polynomials as a data structure with monomials as basic elements, this yields an exponential lower bound on the time these implementations of the Groebner basis algorithm take on any unsolvable instance of subset sum, or on the pigeonhole principle.

Our argument uses the outline of the simulation of resolution by the polynomial calculus from Clegg *et al.* (1996). We simulate any polynomial calculus proof with a sub-exponential size by a different proof with sub-linear degree. Thus, since we have linear degree bounds, this yields an exponential size bound. The argument for subset sum applies to any set of polynomials with a linear degree bound. The proof has to be tailored for the pigeon-hole principle, since the number of variables is actually the square of the degree bound.

Let $p|_{x=c}$ be the restriction of a polynomial $p$ obtained by replacing variable $x$ by the constant $c$; let $P|_{x=c}$ be the set of all $p|_{x=c}$, $p \in P$. We use the following definitions and claim from Clegg *et al.* (1996) (it is stated there for multilinear polynomials, but it holds with no change for arbitrary polynomials).

LEMMA 6.1. (Clegg *et al.* (1996), Lemma 9) *Let $x$ be a variable, $c \in \{0,1\}$ and $P$ a set of polynomials of degree at most $d$. Suppose that $P|_{x=c}$ has* *a polynomial calculus refutation of degree $d$ and $P|_{x=1-c}$ has a refutation of degree $d+1$. Then $P$ has a refutation of degree $d+1$.*

We use this lemma to prove:

THEOREM 6.2. *If $P$ is a set of polynomials of degree at most $d$ in $n$ variables, and $P$ has a polynomial calculus refutation with $M$ non-zero monomials, then $P$ has a polynomial calculus refutation of degree* $\max(d, 2\lceil(n\log M)^{1/2}\rceil+1)$.

PROOF. Let $b = \lceil(n\ln M)^{1/2}\rceil$. If $b \geq n$, the conclusion follows from Lemma 4.1. Otherwise, for a refutation of $P$ with at most $M$ monomials, let $S$ be the number of monomials containing at least $b$ distinct variables.

If $S = 0$ then all the monomials have less than $b$ distinct variables, hence $P$ has a refutation of degree $b$ by Lemma 4.1.

We prove that, if $S \geq 1$, then there exists a refutation of $P$ of degree at most $b - \log_{(1-b/n)} S + 1$. The conclusion of the theorem then follows, since $S \leq M$ and $\ln(1-b/n) \leq -b/n$, which implies that $b - \log_{(1-b/n)} S + 1 = b - (\ln S)/(\ln(1-b/n)) + 1 \leq b + n/b \ln S \leq b + (n/b)\ln M + 1 \leq b + (n\ln M)/(n\ln M)^{1/2} + 1 = b + (n\ln M)^{1/2} + 1 \leq 2b+1$.

We prove the statement by induction on $S$ and $n$. Choose a variable $x$ that is in at least $(b/n)S$ monomials of degree larger than $b$; it exists by an averaging argument, since each of $S$ monomials contains at least $b$ of the $n$ variables. For any $c \in \{0,1\}$, we can form a refutation of $P|_{x=c}$ by replacing each line $q$ of the refutation of $P$ by $q|_{x=c}$. Since $(p+q)|_{x=c} = p|_{x=c} + q|_{x=c}$, $(pq)|_{x=c} = p|_{x=c}q|_{x=c}$, and $1|_{x=c} = 1$, this remains a valid refutation.

For $c = 0$ the restriction described above removes all the monomials containing $x$, and hence it removes at least $(b/n)S$ monomials of degree larger than $b$. Thus it produces a refutation of $P|_{x=0}$ with less than $(1-b/n)S$ monomials with at least $b$ variables. If $(1-b/n)S \geq 1$ then by induction on $S$ the set $P|_{x=0}$ has a refutation of degree $b - \log_{(1-b/n)}((1-b/n)S) + 1 = b - \log_{(1-b/n)} S - 1 + 1 = b - \log_{(1-b/n)} S$. Otherwise all monomials of $P|_{x=0}$ have less than $b$ variables and hence by Lemma 4.1, $P|_{x=0}$ has a refutation of degree $b \leq b - \log_{(1-b/n)} S$.

For $c = 1$ we obtain a refutation of $P|_{x=1}$ with at most $S$ monomials with at least $b$ variables. Using the induction on the number of variables in the proof, $P|_{x=1}$ has a refutation of degree $b - \log_{(1-b/n)} S + 1$. Thus, applying Lemma 6.1, there is a refutation of $P$ of degree at most $b - \log_{(1-b/n)} S + 1$. □

As an immediate consequence we obtain:

COROLLARY 6.3. *For any set of inconsistent constant degree polynomials of degree at most $d$ in $n$ variables, if the minimum degree of a polynomial calculus* *refutation is at least $d$, then the minimal number $M$ of non-zero monomials in such a refutation is $M \geq 2^{\Omega(d^2/n)}$.*

COROLLARY 6.4. *Any polynomial calculus proof over the reals of any subset-sum problem requires $2^{\Omega(n)}$ monomials.*

PROOF. By Theorem 5.1 the minimum degree is $\Omega(n)$. The bound $M \geq 2^{\Omega(n^2/n)} = 2^{\Omega(n)}$ then follows from Corollary 6.3 □

The Groebner basis algorithm is often suggested as a heuristic for satisfiability of polynomials or to count the number of solutions. The ease of representing Subset Sum problems as polynomials over the reals makes this a tempting application for Groebner basis techniques. The above shows that, for at least this one domain, Groebner basis algorithms never do much better than exhaustive search! So the appropriateness of using the Groebner basis algorithm for $NP$-hard problems is at best problem specific.

COROLLARY 6.5. *For any implementation of the Groebner basis algorithm that stores polynomials as vectors or lists of their non-zero monomials, the algorithm will take exponential time, i.e., time $2^{\Omega(n)}$ when run on any unsolvable instance of subset sum over the real numbers.*

The lower bound of Corollary 6.3 is meaningless if the degree bound is less than the square root of the number of variables. For example, $PHP_n^m$ has $nm$ variables, and can be refuted in degree $n$, so we cannot immediately apply Corollary 6.3 to get a size bound. However, we can prove a stronger simulation of degree vs. size when the variables are in mutually exclusive groups, like in the pigeon-hole principle and in the mod $p$ partition principles. This stronger simulation will give a size lower bound for the pigeon-hole principle.[2]

DEFINITION 6.6. *Let $C \subseteq \{x_1, \ldots, x_n\}$. Then $E_C$ is the set of polynomials $\{xy \mid x, y \in C, x \neq y\}$. If $P$ is a set of polynomials in $x_1, \ldots, x_n$, and $C_1, \ldots, C_k$ are sets with $\bigcup_{i=1}^{i=k} C_i = \{x_1, \ldots, x_n\}$ and $E_{C_i} \subseteq P$ for $i = 1, \ldots, k$, then we say $P$ has a $k$-division.*

For example, $C_i = \{x_{i,j} \mid j \in [m]\}$ is an $n$-division for $PHP_n^m$.

[2] As one of the referees points out, the size lower bound for the pigeon-hole principle can also be obtained by a random-restriction argument.

THEOREM 6.7. *Let $P$ be a set of degree at most $d$ polynomials which has a $k$-division. Then if $P$ has a polynomial calculus refutation with $M$ non-zero monomials, then $P$ has a polynomial calculus refutation of degree at most* $\max(d, 3(k \log M)^{1/2} + 1)$,

PROOF. Let $b = \lceil (k \ln M)^{1/2} \rceil$. Call a monomial $b$-big if it intersects with at least $b$ different $C_i$'s. Let $S$ be the number of $b$-big monomials in a refutation of $P$ with $M$ monomials. If $S = 0$, we claim there is a refutation of degree $\max(d, b+1)$. In this case, every monomial of degree greater than $b$ must contain a squared variable $x^2$ or two variables $x, y$ from the same set $C_i$. Such monomials can be removed from the proof immediately after they arise, using either the axiom $x^2 = x$ or the axiom $xy = 0$ in $P$.

We prove that, if $S \geq 1$, then there exists a refutation of $P$ of degree at most $b - \log_{(1-b/2k)} S + 1$. The conclusion of the theorem then follows, since $S \leq M$ and $\ln(1 - b/2k) \leq -b/2k$, which implies that $b - \log_{(1-b/2k)} S + 1 = b - (\ln S)/(\ln(1 - b/2k)) + 1 \leq b + (2k/b) \ln S + 1 \leq b + (2k/b) \ln M + 1 \leq b + (2k \ln M)/(k \ln M)^{1/2} + 1 = b + 2(k \ln M)^{1/2} + 1 \leq 3b + 1$.

We prove the statement by induction on $S$ and $n$. Choose a class $C_i$ that intersects with at least $(b/k)S$ $b$-big monomials; it exists by an averaging argument, since each $b$-big monomials intersects at least $b$ of the $k$ classes. Let $x \in C_i$. We can form a refutation of $P|_{x=0}$ by replacing each line $q$ of the refutation of $P$ by $q|_{x=0}$. We can form a refutation of $P|_{x=1}$ by restricting $x = 1$ and then restricting $y = 0$ for every $y \in C_i$, $y \neq x$, since $y = 0$ is the axiom $xy = 0$ restricted by $x = 1$.

Either $x$ appears in at least $(b/2k)S$ big monomials, or there are $(b/2k)S$ big monomials intersecting $C_i - \{x\}$. For $x = 0$ the restriction described above removes all the monomials containing $x$. For $x = 1$, it removes all monomials intersecting $C_i - \{x\}$. Thus, at least one of the two restrictions has at most $(1 - b/2k)S$ $b$-big monomials, and both involve fewer variables. For the former restriction, we inductively get a refutation of degree at most $b - \log_{(1-b/2k)}((1 - b/2k)S) + 1 = b - \log_{(1-b/2k)} S$. For the other restriction, by induction on $n$, we get a refutation of degree at most $b - \log_{(1-b/2k)} S + 1$. Thus, applying Lemma 6.1, there is a refutation of $P$ of degree at most $b - \log_{(1-b/2k)} S + 1$.

COROLLARY 6.8. *Any polynomial calculus proof over any field of $PHP_n^m$ has $2^{\Omega(n)}$ monomials.*

## 7. Conclusions and open problems

The lower bound can be translated in a design-based lower bound on Nullstellensatz refutations (see Buss *et al.* 1997); the value of the design on a monomial is the coefficient at 1 if the monomial is expressed in the basis $B_d$. Perhaps if we would have a more explicit description of this design, we could modify it to work also for other principles. However, the current proof shows that any polynomial calculus refutation can be converted to a Nullstellensatz proof of the same degree, which is not necessarily true for other principles than PHP.

Of course, as mentioned by Razborov (1996), the most interesting problem is to extend the lower bounds to stronger systems like the bounded depth Frege systems with modular gates, perhaps using the machinery of extension polynomials from Buss *et al.* (1997).

## Acknowledgements

This work was supported by cooperative research grant INT-9600919/ME-103 from the NSF (USA) and the MŠMT (Czech republic). The first author was partially supported by NSF YI Award CCR-92-570979 and Sloan Research Fellowship BR-3311. The second and third authors were partially supported by grant A1019602 of GA AV ČR, and the third author also by postdoctoral grant 201/97/P038 of GA ČR.

## References

M. Ajtai, The complexity of the pigeonhole principle. In *Proc. of the 29th Ann. IEEE Symp. on Foundations of Computer Sci.* IEEE, 1988, 346–355.

M. Ajtai, The independence of the modulo $p$ counting principle. In *Proc. of the 26th Ann. ACM Symp. on Theory of Computing*. ACM, 1994, 402–411.

P. Beame and S. Riis, More on the relative strength of counting principles. In *Proof Complexity and Feasible Arithmetics*, ed. P. W. Beame and S. R. Buss, vol. 39 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 13–35. AMS, 1998.

P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, and P. Pudlák, Lower bounds on Hilbert's Nullstellensatz and propositional proofs. *Proc. London Math. Soc.* **73** (1996), 1–26.

S. Buss, R. Impagliazzo, J. Krajíček, P. Pudlák, A. A. Razborov, and J. Sgall, Proof complexity in algebraic systems and bounded depth Frege systems with modular counting. *Comput. Complexity* **6**(3) (1997), 256–298.

M. Clegg, J. Edmonds, and R. Impagliazzo, Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proc. of the 28th Ann. ACM Symp. on Theory of Computing*. ACM, 1996, 174–183.

D. H. Gottlieb, A certain class of incidence matrices. *Proc. Amer. Math. Soc.* **17** (1966), 1233–1237.

R. Impagliazzo, P. Pudlák, and J. Sgall, Lower bounds for the polynomial calculus and the Groebner basis algorithm. Technical Report TR97-042, Electronic Colloquium on Computational Complexity, `http://www.eccc.uni-trier.de/eccc/`, 1997.

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

J. Krajíček, On the degree of ideal membership proofs from uniform families of polynomials over a finite field. Manuscript, 1997a.

J. Krajíček, Uniform families of polynomial equations over a finite field and structures admitting an Euler characteristic of definable sets. Manuscript, 1997b.

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

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

A. A. Razborov, Lower bounds for the polynomial calculus. Manuscript (to appear in Comput. Complexity, 1998), 1996.

Manuscript received

Russell Impagliazzo
Computer Science and Engineering
UC San Diego
9500 Gilman Drive
La Jolla, CA 92093-0114
`russell@cs.ucsd.edu`

Pavel Pudlák
Mathematical Institute, AS CR
Žitná 25
CZ-11567 Praha 1, Czech Republic
`pudlak@math.cas.cz`
`http://www.math.cas.cz/~pudlak`

Jiří Sgall
Mathematical Institute, AS CR
Žitná 25
CZ-11567 Praha 1, Czech Republic
and
Dept. of Applied Mathematics
Faculty of Mathematics and Physics
Charles University, Prague
sgall@math.cas.cz
http://www.math.cas.cz/~sgall/