Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 63 (2020)

On the Existence of Algebraically Natural Proofs

Prerona Chatterjee[*] Mrinal Kumar[†] C. Ramya[‡] Ramprasad Saptharishi[§] Anamay Tengse[¶]

Abstract

For every constant $c > 0$, we show that there is a family $\{P_{N,c}\}$ of polynomials whose degree and algebraic circuit complexity are polynomially bounded in the number of variables, that satisfies the following properties:

- For every family $\{f_n\}$ of polynomials in VP, where f_n is an n variate polynomial of degree at most n^c with bounded integer coefficients and for $N = \binom{n^c+n}{n}$, $P_{N,c}$ *vanishes* on the coefficient vector of f_n.
- There exists a family $\{h_n\}$ of polynomials where h_n is an n variate polynomial of degree at most n^c with bounded integer coefficients such that for $N = \binom{n^c+n}{n}$, $P_{N,c}$ *does not vanish* on the coefficient vector of h_n.

In other words, there are efficiently computable equations for polynomials in VP that have small integer coefficients. In fact, we also prove an analogous statement for the seemingly larger class VNP. Thus, in this setting of polynomials with small integer coefficients, this provides evidence *against* a natural proof like barrier for proving algebraic circuit lower bounds, a framework for which was proposed in the works of Forbes, Shpilka and Volk [FSV18], and Grochow, Kumar, Saks and Saraf [GKSS17].

Our proofs are elementary and rely on the existence of (non-explicit) hitting sets for VP (and VNP) to show that there are efficiently constructible, low degree equations for these classes. Our proofs also extend to finite fields of small size.

*prerona.chatterjee.tifr@gmail.com. School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India. Research supported by a fellowship of the DAE, Government of India.

†mrinal@cse.iitb.ac.in. Department of Computer Science & Engineering, IIT Bombay, Mumbai, India.

‡c.ramya@tifr.res.in. School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India. Research supported by a fellowship of the DAE, Government of India.

§ramprasad@tifr.res.in. School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India. Research supported by Ramanujan Fellowship of DST, and by DAE, Government of India.

¶anamay.tengse@gmail.com. School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India. Research supported by a fellowship of the DAE, Government of India.

ISSN 1433-8092

1 Introduction

The quest for proving strong lower bounds for algebraic circuits is one of the fundamental challenges in algebraic complexity, and maybe the most well studied one. And yet, progress on this problem has been painfully slow and sporadic. Perhaps the only thing more frustrating than the inability to prove such lower bounds is the inability to come up with plausible approaches towards them. This lack of progress on the problem and a dearth of potential approaches towards it has spurred some work towards understanding the viability of some of the current lower bound approaches; the idea being that a good sense of what approaches will *not* work would aid in the search of approaches that *might* work.

In the broader context of lower bounds in computational complexity, there are various results of this flavor which establish that various families of techniques cannot be used for proving very strong lower bounds, e.g., the barrier of *Relativization* due to Baker, Gill and Solovay [BGS75], that of *Algebraization* due to Aaronson and Wigderson [AW09] and that of *Natural Proofs* due to Razborov and Rudich [RR97].[1] While none of these barrier results are directly applicable to the setting of algebraic computation, there have been recent attempts towards generalizing these ideas to the algebraic set up. A key notion in this line of work is the notion of *algebraically natural proofs* alluded to and defined in the works of Aaronson and Drucker [AD08], Forbes, Shpilka and Volk [FSV18], and Grochow, Kumar, Saks and Saraf [GKSS17].

We now discuss this notion, starting with a discussion of *Natural Proofs* which motivated the definition.

1.1 The Natural Proofs framework of Razborov and Rudich

Razborov and Rudich [RR97] noticed that underlying many of the lower bound proofs known in Boolean circuit complexity, there was some common structure. They formalized this common structure via the notion of a *Natural Property*, which we now define.

Definition 1.1. *A subset $\mathcal{P} \subseteq \{f : \{0,1\}^n \to \{0,1\}\}$ of Boolean functions is said to be a natural property useful against a class $\mathcal{C}$ of Boolean circuits if the following are true.*

- **Usefulness.** *Any Boolean function $f : \{0,1\}^n \to \{0,1\}$ that can be computed by a Boolean circuit in $\mathcal{C}$ does not have the property $\mathcal{P}$.*

- **Constructivity.** *Given the truth table of a Boolean function $f : \{0,1\}^n \to \{0,1\}$, whether or not it has the property $\mathcal{P}$ can be decided in time polynomial in the length of the input, i.e. in time $2^{O(n)}$.*

- **Largeness.** *For all large enough n, at least a $2^{-O(n)}$ fraction of all n variate Boolean functions have the property $\mathcal{P}$.* ◊

A proof that a certain family of Boolean functions cannot be computed by circuits in $\mathcal{C}$ is said to be a *natural lower bound proof* if the proof (perhaps implicitly) proceeds via establishing a nat-

[1] Sometimes, these results are conditional, as in [RR97].

ural property useful against $\mathcal{C}$, and showing that the candidate hard function has this property. Razborov and Rudich then showed that most of the Boolean circuit lower bound proofs that we know, e.g., lower bounds for AC^0 circuits [FSS84, Hås86] or lower bounds for $\mathsf{AC}^0[\oplus]$ circuits [Raz87, Smo87] fit into this framework (maybe with some work) and hence are *natural* in this sense. Further, they argue that under standard cryptographic assumptions, the proof of a lower bound against any sufficiently rich circuit class (such as the class $\mathsf{P}/\operatorname{poly}$) cannot be natural! Thus, under standard cryptographic assumptions, most of the current lower bound techniques are not strong enough to show super-polynomial lower bounds for general Boolean circuits.

We now move on to discuss a relatively recent analog of the notion of Natural Proofs, formalized in the context of algebraic computation.

1.2 Algebraically Natural Proofs

Algebraic complexity is the study of computational questions about multivariate polynomials as formal objects. The basic model of computation here, an algebraic circuit, is an algebraic analog of a Boolean circuit with the gates of the circuit being labeled by $+$ (sum) and $\times$ (product) gates as opposed to Boolean functions.[2] The algebraic analog of $\mathsf{P}/\operatorname{poly}$ is the class VP of polynomial families $\{f_n\}$, where f_n is an n variate polynomial of degree and algebraic circuit size $\operatorname{poly}(n)$. A fundamental question in this setting is to come up with explicit families of polynomials, i.e. polynomials in the class VNP (the algebraic analog of $\mathsf{NP}/\operatorname{poly}$), which are not in VP. While the state of the art of lower bounds for algebraic circuits is a bit better than that for Boolean circuits, with slightly super linear lower bounds having been shown by Strassen [Str73] and Baur & Strassen [BS83], this lower bound has seen no improvements for nearly four decades. This absence of progress has led to some research towards understanding the limitations of the current proof techniques in proving strong lower bounds for algebraic circuits.

Considering that algebraic circuits seem like a fairly general and powerful model of computation, it is tempting to think that the *natural proofs barrier* of Razborov and Rudich [RR97] also extends to this setting. This problem turns out to be a non-trivial one, and indeed it is not known whether their results extend to algebraic circuits. This question is closely related to the question of whether cryptographically secure algebraic pseudorandom functions can be computed by small and low degree[3] algebraic circuits and there does not seem to be substantial evidence one way or the other on this. We refer the reader to [AD08] and [FSV18] for a more detailed discussion on this issue.

In the last few years, this question of trying to find an algebraic analog of the barrier results in [RR97] has received substantial attention. It was observed by various authors [AD08, Gro15, FSV18, GKSS17] that most of the currently known proofs of algebraic circuit lower bounds fit into a common unifying framework, not unlike that in [RR97], although of a more algebraic nature.

[2]See Definition 2.1 for a formal definition.

[3]Throughout this paper, by a *low degree* polynomial family, we mean a polynomial family whose degree is polynomially bounded in its number of variables.

Indeed, these proofs also implicitly go via defining a property for the set of all polynomials and using this property to separate the hard polynomial from the easy ones. Moreover, the notions of *largeness* and *constructivity* in Definition 1.1 also seem to extend to these proofs.

We now discuss this framework in a bit more detail. The key notion here is that of an equation of polynomials in a complexity class.

Definition 1.2 (Equations)**.** *For some $n, d \in \mathbb{N}$, let $\mathcal{C}_{n,d}$ be a class of n-variate polynomials of* total *degree at most d; i.e. $\mathcal{C}_{n,d} \subseteq \mathbb{F}[\mathbf{x}]^{\leq d}$.*

Then for $N = \binom{n+d}{n}$, a nonzero polynomial $P_N(\mathbf{Z})$ is said to be an equation *for $\mathcal{C}_{n,d}$ if for all $f(\mathbf{x}) \in \mathcal{C}_{n,d}$, we have that $P_N(\overline{\mathrm{coeff}}(f)) = 0$, where $\overline{\mathrm{coeff}}(f)$ is the coefficient vector of f.* $\lozenge$

The definition naturally extends to a class of polynomial families, as opposed to just a class of polynomials as defined above. In particular, suppose that $\mathcal{C}$ is a class of polynomial families $\{\{f_n\} : f_n \in \mathcal{C}_{n,d_n}\}$, and $\{P_N\}$ is a polynomial family. Then, the family $\{P_N\}$ is said to be a family of equations for $\mathcal{C}$ if there is an n_0, such that for all $n \geq n_0$ the polynomial P_N is an equation for $\mathcal{C}_{n,d_n}$ where $N = \binom{n+d_n}{n}$. That is, P_N is an equation for $\mathcal{C}_{n,d_n}$ for all large enough n.

Intuitively, non-vanishing of an equation (for a class $\mathcal{C}$) on the coefficient vector of a given polynomial f is a proof that f is not in $\mathcal{C}$. We note that the equations for a class $\mathcal{C}$ evaluate to zero not just on the coefficient vectors of polynomials in $\mathcal{C}$ but also on the coefficient vectors of polynomials in the Zariski closure of $\mathcal{C}$. This framework comes up very naturally in the context of algebraic geometry (and geometric complexity theory), where it is often geometrically nicer to work with the variety obtained by taking the Zariski closure of a complexity class.

Getting our hands on an equation of a variety gives us a plausible way to test and certify non-membership in the variety, in other words, to prove a lower bound for the corresponding complexity class. Thus, equations for a class gives an algebraic analog of the notion of *natural properties useful against a class* in [RR97]. Moreover, since a nonzero polynomial does not vanish very often on a random input from a large enough grid, it follows that a nonzero equation for a class $\mathcal{C}$ will be nonzero on the coefficient vector of a "random polynomial". Here by a random polynomial we mean a polynomial whose coefficients are independent and uniformly random elements from some large enough set in the underlying field. With appropriate quantitative bounds, this observation can be formalized to give an appropriate algebraic analog of the notion of *largeness*. Lastly, the algebraic circuit complexity of the equation gives a natural algebraic analog of the notion of *constructivity*. Intuitively, any algebraic circuit lower bound which goes via defining a nonzero proof polynomial of polynomially bounded degree that can be efficiently computed by an algebraic circuit is an *Algebraically Natural Proof* of a lower bound.

We now formally define an algebraically natural proof.

Definition 1.3 (Algebraically natural proofs [FSV18, GKSS17])**.** *Let $\mathcal{C}$ be a class of polynomial families $\{\{f_{n,d}\} : f_{n,d} \in \mathcal{C}_{n,d}\}$.*

Then, for a class $\mathcal{D}$ of polynomial families, we say that $\mathcal{C}$ has $\mathcal{D}$-natural proofs *if there is a family $\{P_N\} \in \mathcal{D}$ which is a non-trivial family of equations for $\mathcal{C}$.* $\lozenge$

In the rest of this paper, whenever we say a natural proof, without specifying the class $\mathcal{D}$, we mean a VP-natural proof.

Analogous to the abstraction of *natural proofs* for Boolean circuit lower bounds, this framework of *algebraically natural proofs* turns out to be rich and general enough that almost all of our current proofs of algebraic circuit lower bounds are in fact algebraically natural, or can be viewed in this framework with a little work [Gro15]. Thus, this definition seems like an important first step towards understanding the strengths and limitations of many of our current lower bound techniques in algebraic complexity.

The immediate next question to ask is whether algebraically natural proofs are rich enough to give strong algebraic circuit lower bounds. This can naturally be worded in terms of the complexity of equations for the class VP as follows.

Question 1.4. *For every constant $c > 0$, does there exist a nonzero polynomial family $\{P_{N,c}\}$ in* VP *such that for all large enough n, the following is true?*

> *For every family of polynomials $\{f_n\}_n$ in* VP*, such that f_n is an n variate polynomial of degree n^c, $P_{N,c}$ vanishes on the coefficient vector of f_n for $N = \binom{n+n^c}{n}$.*

The works of Forbes *et al.* [FSV18] and Grochow *et al.* [GKSS17] argue that under an appropriate (but non-standard) pseudorandomness assumption, the answer to the question above is negative, i.e., algebraically natural proof techniques cannot be used to show strong lower bounds for algebraic circuits. To discuss this pseudorandomness assumption formally, we need the following definition of *succinct hitting sets*.

Definition 1.5 (Succinct hitting sets for a class of polynomials (Informal))**.** *For some $n, d \in \mathbb{N}$, let $\mathcal{C}_{n,d}$ be a class of n-variate polynomials of* total *degree at most d; i.e. $\mathcal{C}_{n,d} \subseteq \mathbb{F}[\mathbf{x}]^{\leq d}$.*

Then for $N = \binom{n+d}{n}$, we say that a class of N variate polynomials $\mathcal{D}_N$ has $\mathcal{C}_{n,d}$-succinct hitting sets *if for all $0 \not\equiv P(\mathbf{Z}) \in \mathcal{D}_N$, there exists some $f \in \mathcal{C}_{n,d}$ such that $P_N(\overline{\mathrm{coeff}}(f)) \neq 0$.* ◊

As with Definition 1.2, this definition naturally extends to polynomial families.

It immediately follows from the definitions that non-existence of $\mathcal{D}$-natural proofs against a class $\mathcal{C}$ is equivalent to the existence of $\mathcal{C}$-succinct hitting sets for the class $\mathcal{D}$. Forbes, Shpilka and Volk [FSV18] showed that for various restricted circuit classes $\mathcal{C}$ and $\mathcal{D}$, the class $\mathcal{D}$ has $\mathcal{C}$ succinct hitting sets. Or equivalently, lower bounds for $\mathcal{C}$ cannot be proved via proof polynomial families in $\mathcal{D}$. However, this question has remained unanswered for more general circuit classes $\mathcal{C}$ and $\mathcal{D}$. In particular, if we take both $\mathcal{C}$ and $\mathcal{D}$ to be VP, we do not seem to have significant evidence on the existence of VP succinct hitting sets for VP. In [FSV18], the authors observed that showing VP succinct hitting sets for VP would immediately imply non-trivial deterministic algorithms for polynomial identity testing, which via well known connections between algebraic hardness and derandomization will in turn imply new lower bounds [HS80, KI04]. Thus, the problem of proving an unconditional barrier result for algebraically natural proof techniques via this route seems as hard as proving new circuit lower bounds! It is, however, conceivable that one can show such a barrier conditionally. And in some more structured settings, such as for the case of matrix completion, such results are indeed known [BIJL18]. However, Question 1.4 continues to remain open. In particular, even though many of the structured subclasses of VP have low degree equations which are very efficiently computable, perhaps hoping that this extends to richer and more general circuit classes is too much to ask for?

We are now ready to state our results.

1.3 Our results

In our main results, we make progress towards answering Question 1.4 in the affirmative. We prove the following theorems.

Equations for polynomials in VP with coefficients of small complexity

Theorem 1.6. *Let $c > 0$ be any constant. There is a polynomial family $\{P_{N,c}\} \in \mathsf{VP}_{\mathbb{Q}}$ such that for all large n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every family $\{f_n\} \in \mathsf{VP}_{\mathbb{C}}$, where f_n is an n variate polynomial of degree at most n^c and coefficients in $\{-1, 0, 1\}$, we have*

$$P_{N,c}(\overline{\mathrm{coeff}}(f_n)) = 0\,,$$

where $\overline{\mathrm{coeff}}(f_n)$ is the coefficient vector of f_n.

- *There exists a family $\{h_n\}$ of n variate polynomials and degree $\leq n^c$ with coefficients in $\{-1, 0, 1\}$ such that*

$$P_{N,c}(\overline{\mathrm{coeff}}(h_n)) \neq 0\,.$$

Here for a field $\mathbb{F}$, $\mathsf{VP}_{\mathbb{F}}$ denotes the class VP where the coefficients of the polynomials are from the field $\mathbb{F}$.

We remark that even though Theorem 1.6 is stated for polynomials with $\{-1, 0, 1\}$ coefficients, the theorem holds for polynomials with coefficients as large as N. However, for brevity, we will confine the discussion in this paper to polynomials with coefficients in $\{-1, 0, 1\}$.

We also prove an analogous theorem for finite fields of small size.

Theorem 1.7. *Let $\mathbb{F}$ be any finite field of constant size and $c > 0$ be any constant. There is a polynomial family $\{P_{N,c}\} \in \mathsf{VP}_{\mathbb{F}}$ such that for all large enough n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every $\{f_n\} \in \mathsf{VP}_{\mathbb{F}}$, where f_n is an n variate polynomial of degree at most n^c, we have*

$$P_{N,c}(\overline{\mathrm{coeff}}(f_n)) = 0\,.$$

- *There exists a family $\{h_n\}$ of n variate polynomials and degree $\leq n^c$ with coefficients in $\mathbb{F}$ such that*

$$P_{N,c}(\overline{\mathrm{coeff}}(h_n)) \neq 0\,.$$

Furthermore, we also prove analogous statements for the larger class VNP, which we now state.

Equations for polynomials in VNP with coefficients of small complexity

Theorem 1.8. *Let $c > 0$ be any constant. There is a polynomial family $\{Q_{N,c}\} \in \mathsf{VP}_{\mathbb{Q}}$ such that for all large n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every family $\{f_n\} \in \mathsf{VNP}_{\mathbb{C}}$, where f_n is an n variate polynomial of degree at most n^c and coefficients in $\{-1, 0, 1\}$, we have*

$$Q_{N,c}(\overline{\text{coeff}}(f_n)) = 0\,.$$

- *There exists a family $\{h_n\}$ of n variate polynomials and $\leq n^c$ with coefficients in $\{-1, 0, 1\}$ such that*

$$Q_{N,c}(\overline{\text{coeff}}(h_n)) \neq 0\,.$$

Theorem 1.9. *Let $\mathbb{F}$ be any finite field of constant size and $c > 0$ be any constant. There is a polynomial family $\{Q_{N,c}\} \in \mathsf{VP}_{\mathbb{F}}$ such that for all large enough n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every family $\{f_n\} \in \mathsf{VNP}_{\mathbb{F}}$, where f_n is an n variate polynomial of degree at most n^c, we have*

$$Q_{N,c}(\overline{\text{coeff}}(f_n)) = 0\,.$$

- *There exists a family $\{h_n\}$ of n variate polynomials and degree $\leq n^c$ with coefficients in $\mathbb{F}$ such that*

$$Q_{N,c}(\overline{\text{coeff}}(h_n)) \neq 0\,.$$

1.4 Discussion and relations to prior work

As is evident from the statements, our main theorems make some progress towards answering Question 1.4 in the affirmative, at least in the setting of small finite fields and for polynomials with small integer coefficients, in a fairly strong sense. In fact, as Theorem 1.8 and Theorem 1.9 show, in the context of polynomials with coefficients of low complexity, not just VP but even the seemingly larger class VNP has efficiently computable low degree equations.

Many of the families of polynomials commonly studied in algebraic complexity have integer coefficients with absolute values bounded by 1, and fall in the setting of the results here. Moreover, the condition of computing polynomials with bounded coefficients is a semantic condition on a model, in the sense that even though the final output of the circuit is required to have bounded coefficients, the circuit is free to use arbitrary constants from $\mathbb{C}$ in the intermediate computation. Thus, it is conceivable that we might be able to prove a super-polynomial lower bound on the algebraic circuit size for the permanent polynomial via an algebraically natural proof constructible in VP, thereby separating VP and VNP. However, since analogs of Theorem 1.6 and Theorem 1.7 are also true for VNP, any such separation of VNP and VP will have to rely on more fine grained information on the equations, and not just their degree and algebraic circuit size. Unfortunately, our proofs are all existential and do not give a sense of what the polynomial families $\{P_{N,c}\}$ (or $\{Q_{N,c}\}$) might look like.

We also note that in the light of some of the prior work, the results here are perhaps a bit surprising. The classes of polynomials in VP and VNP with small coefficients (or over small finite fields), are seemingly rich and complex sets, and the main theorems here show (un-conditionally) that they have equations which are also efficiently computable. As discussed earlier in this introduction, this property is known to be true for many structured subclasses of algebraic circuits (for example, homogeneous circuits of depth 3 and 4, multilinear formulas, polynomials of small Waring rank). However, it is unclear if this property extends to more general circuit classes, in particular VP (or VNP).

Indeed, following the work of Forbes *et al.* [FSV18] and Grochow *et al.* [GKSS17], much of the research on this problem [FSV18, BIJL18, BIL+19] has focused on proving the *non-existence* of efficiently computable equations for VP, and this line of work has made interesting progress in this direction for many structured and special instances of problems of this nature. The results in [BIJL18, BIL+19] draw connections between the existence of efficiently constructible equations of a variety and the problem of testing (non)membership in it and use the conditional hardness of the (non)membership testing problem for certain varieties to rule out the existence of efficiently computable equations for them. More precisely, Bläser *et al.* [BIJL18] show that if all the equations for the variety of matrices with zero permanent are constructible by small constant-free algebraic circuits, then the non-membership problem for this variety can be decided in the class ∃BPP. Thus, unless $\mathsf{P}^{\#\mathsf{P}} \subseteq \exists\mathsf{BPP}$, the equations of this variety do not have small, low degree constant free algebraic circuits. In a subsequent work [BIL+19], the results of [BIJL18] are generalized to *min-rank* or *slice-rank* varieties. However, in the bounded coefficient setting (and over small finite fields), our results show that the contrary is true, and VP does have efficiently computable low degree equations. We also remark that because of the setting of bounded integer coefficients or small finite fields in this work, this natural connection between variety non-membership and equations of varieties discussed in [BIJL18, BIL+19] appears to break down.

A positive result on the complexity of equations of naturally occuring varieties in algebraic complexity appears in a recent work of Kumar and Volk [KV20] where they show polynomial degree bounds on the equations of the Zariski closure of the set of non-rigid matrices and small linear circuits over all large enough fields. However, we do not know if any of these low degree equations can be *efficiently* computed by an algebraic circuit.

As alluded to in the previous paragraphs, most of the prior work related to Question 1.4 has focused on looking for evidence that the answer to it is negative, i.e. VP does not have efficiently computable and low degree equations. We hope that the results in this paper also highlight the possibility of there being interesting upper bounds for the equations for rich and powerful algebraic complexity classes; a line of research that hasn't received much attention so far.

Other related work. Many of the algebraic circuit lower bounds (e.g. lower bounds for depth-3 and depth-4 circuits, and lower bounds for multilinear models) are obtained by considering the rank of certain matrices as a complexity measure. In their recent works, Efremenko, Garg, Oliveira and Wigderson [EGOW18] and Garg, Makam, Oliveira and Wigderson [GMOW19], discuss limitations of some of these rank based methods towards proving lower bounds. In particular, Efremenko *et al.* [EGOW18] show that some of these rank based methods cannot prove lower bounds better than $\Omega_d(n^{\lfloor d/2 \rfloor})$ on tensor rank (resp., Waring rank) for a d-dimensional tensor of side n. Building on [EGOW18], in [GMOW19], the authors demonstrate that one *cannot* hope to significantly improve the known lower bounds for tensor rank for d dimensional tensors by lifting lower bounds on tensors in fewer dimensions. However, we note that a general algebraically natural proof of a lower bound does not necessarily fit into the framework of [EGOW18, GMOW19], and so these limitations for the so called *rank methods* do not seem to immediately extend to algebraically natural proofs in general. As discussed earlier, in the light of the results here, it is conceivable that we might be able to improve the state of the art for general algebraic circuit lower bounds, using techniques that are algebraically natural.

For Boolean circuits, Chow [Cho11] circumvents the natural proofs barrier in [RR97] by providing (under standard cryptographic assumptions) an explicit *almost natural proof* that is useful against P/poly as well as constructive in nearly linear time, but compromises on the largeness condition. Furthermore, Chow [Cho11] shows the unconditional existence of a natural property useful against P/poly (infinitely often) constructive in linear size that has a weakened largeness condition. In some sense, Theorem 1.7 and Theorem 1.6 are analogous to the work of Chow [Cho11], albeit in the algebraic world.

On the largeness criterion. In the definitions of algebraically natural proofs [GKSS17, FSV18], the authors observe that in the algebraic setting, an analog of the *largeness* criterion in Definition 1.1 is often available for free; the reason being that a nonzero equation for any class of polynomials vanishes on a very small fraction of all polynomials over any sufficiently large field. However, this tradeoff becomes a bit subtle when considering polynomials over finite fields of small size, or polynomials with bounded integer coefficients. In particular, as we observe in the course of the proofs of our results, we still have a large number of polynomials whose coefficients will keep $\{P_{N,c}\}$ (and $\{Q_{N,c}\}$) nonzero, although this set is no longer a significant fraction of the set of all polynomials.

1.5 An overview of the proof

At a high level, the idea behind our results is to try and come up with a *non-trivial* property of polynomials which every polynomial with a small circuit satisfies. By a non-trivial property, we mean that there should exist (nonzero) polynomials which do not have this property. The hope is that once we have such a property (which is nice enough), one can try to transform this into an equation via an appropriate *algebraization*. The property that we finally end up using is the existence of (non-explicit) *hitting sets* for polynomials with small circuits.

A hitting set for a class $\mathcal{C}$ of polynomials over a field $\mathbb{F}$ is a set of points $\mathcal{H}$, such that every nonzero polynomial in $\mathcal{C}$ evaluates to a nonzero value on at least one point in $\mathcal{H}$. We then turn this property of *not-vanishing-everywhere on* $\mathcal{H}$ into an equation in some settings to get our main theorems.

To make things a bit more formal, let us consider the map $\Phi_{\mathcal{H}}$, defined by the hitting set $\mathcal{H}$ of $\mathcal{C}$ on the set of all polynomials, that maps any given polynomial f to its evaluations over the points in $\mathcal{H}$. It is clear from the above observation that any nonzero polynomial in the kernel of $\Phi_{\mathcal{H}}$ is guaranteed to be outside $\mathcal{C}$. Thus, if there were a nonzero polynomial that vanishes on all polynomials $f \notin \ker(\Phi_{\mathcal{H}})$, we have an equation for $\mathcal{C}$.

Moreover, if such a polynomial happened to have its degree and circuit complexity polynomially bounded in its number of variables, we would have our main theorems. However, note that *not* being in the kernel of a linear map seems to be a tricky condition to check via a polynomial (as opposed to the complementary property of *being* in the kernel, which can be easily checked via a polynomial). To prove our theorems, we get past this issue in the setting of small finite fields, and for polynomials over $\mathbb{C}$ with bounded integer coefficients.

Over a finite field $\mathbb{F}$, a univariate polynomial that maps every nonzero $x \in \mathbb{F}$ to zero and vice versa, already exists in $q(x) = 1 - x^{|\mathbb{F}|-1}$. Therefore, for a given polynomial f, the equation essentially outputs $\prod_{\mathbf{h}\in\mathcal{H}} q(f(\mathbf{h}))$. Clearly, for a polynomial f, $\prod_{\mathbf{h}\in\mathcal{H}} q(f(\mathbf{h}))$ is zero if and only if f evaluates to a nonzero value on at least one point in $\mathcal{H}$.

To generalize this to other fields, we wish to find a "low-degree" univariate $q(x)$ that maps nonzero values to 0, and zero to a nonzero value. We observe that in the setting when the polynomials in $\mathcal{C}$ have integer coefficients of bounded magnitude , we can still obtain such a univariate polynomial, and in turn a non-trivial equation. Indeed, if q were such a univariate, we essentially output $\prod_{\mathbf{h}\in\mathcal{H}} q(f(\mathbf{h}))$, for a given polynomial f. This step relies on a simple application of the Chinese Remainder Theorem.

In order to show that the equations are non-trivial in the sense that there exist polynomials with bounded integer coefficients which do not pass this test, we need to show that there are nonzero polynomials with bounded integer coefficients which vanish everywhere on the hitting set $\mathcal{H}$. We show this via a well known lemma of Siegel[4], which uses a simple pigeon hole argument to show that an under-determined system of homogeneous linear equations where the constraint matrix has small integer entries has a nonzero solution with small integer entries.

As it turns out, our proofs do not use much about the class VP except for the existence of small hitting sets for polynomials in the class. It is not hard to observe that this property is also true for the seemingly larger class VNP and hence the results here also extend to VNP.

[4] A statement of the lemma can be found here. Refer to [Sie14] for details.

We remark that given the hitting set $\mathcal{H}$ explicitly, the construction of the equation is completely explicit. In other words, the non-explicitness in our construction comes only from the fact that we do not have explicit constructions of hitting sets for algebraic circuits.

Organization of the paper. We begin with some notations and preliminaries in Section 2 before moving on to prove Theorem 1.7 in Section 3 and Theorem 1.6 in Section 4. In Section 5, we observe that these results also generalize to VNP, and finally conclude with some open questions in Section 6.

2 Notation and preliminaries

2.1 Notation

We use $\{f_n\}_{n\in\mathbb{N}}$ to denote families of polynomials. We drop the index set whenever it is clear from context. For a given polynomial f we denote by $\deg(f)$ its degree. For a polynomial $f(\mathbf{x}, \mathbf{y}, \ldots)$ on multiple sets of variables, we use $\deg_{\mathbf{x}}(f)$, $\deg_{\mathbf{y}}(f)$, etc., to denote the degree in the variables from the respective sets.

We use $\mathbb{F}[\mathbf{x}]^{\leq d}$ to denote polynomials over the field $\mathbb{F}$ in variables $\mathbf{x}$ of degree at most d, and use $\mathbf{x}^{\leq d}$ to denote the set of all monomials in variables $\mathbf{x}$ of degree at most d.

For a given polynomial $f \in \mathbb{F}[\mathbf{x}]^{\leq d}$ and a monomial $m \in \mathbf{x}^{\leq d}$, we use $\mathrm{coeff}_m(f)$ to refer to the coefficient of m in f. We further use $\overline{\mathrm{coeff}}(f)$ to denote the vector[5] of coefficients of f.

2.2 Algebraic circuits and complexity classes

Let us first formally define algebraic circuits.

Definition 2.1 (Algebraic circuits)**.** *An* algebraic circuit *is specified by a directed acyclic graph, with leaves (indegree zero; also called* inputs*) labeled by field constants or variables, and internal nodes labeled by $+$ or $\times$. The nodes with outdegree zero are called the* outputs *of the circuit. Computation proceeds in the natural way, where inductively each $+$ gate computes the sum of its children and each $\times$ gate computes the product of its children.*

The size *of the circuit is defined as the number of nodes in the underlying graph.* $\Diamond$

We also define the class VP and its "slices" formally.

Definition 2.2 (VP and $\mathsf{VP}^{[c]}$)**.** *A family of polynomials $\{f_n\}$ over a field $\mathbb{F}$ is said to be in $\mathsf{VP}_{\mathbb{F}}$ (or just* VP *when the field is clear from context) if there exist constants c_1, c_2 such that*

- *f_n is an n-variate polynomial,*
- *$\deg(f_n) \leq n^{c_1}$ for all large enough n,*
- *f_n is computable by an algebraic circuit of size at most n^{c_2}, for all large enough n.*

[5] We do not explicitly mention the monomial ordering used for this vector representation, since all our statements work for any monomial ordering.

Even though $f_n \in \mathbb{F}[\mathbf{x}]$, the circuit computing f_n may employ constants from a larger extension field $\mathbb{K} \supset \mathbb{F}$.

For the ease of notation, we also consider "slices" of VP *with $d = n^c$ for a fixed constant c. To this end, we will define* $\mathsf{VP}_{\mathbb{F}}^{[c]}$ *to denote*

$$\mathsf{VP}_{\mathbb{F}}^{[c]} := \left\{ \{f_n\} \in \mathsf{VP}_{\mathbb{F}} \ : \ f_n \in \mathbb{F}[x_1, \ldots, x_n]^{\leq n^c} \right\}.$$

We also consider polynomials $\{f_n\}$ over integers whose coefficients are in $\{-1, 0, 1\}$. However, it is important to note that even in this setting the bound is only on the coefficients of f_n; the circuit computing f_n may use arbitrary constants from the underlying field, or an extension. ♢

Finally, let us formally define the class VNP and its "slices".

Definition 2.3 (VNP and $\mathsf{VNP}^{[c]}$)**.** *A family of polynomials $\{f_n\}$ over a field $\mathbb{F}$ is said to be in $\mathsf{VNP}_{\mathbb{F}}$ (or just VNP when the field is clear from context) if there exist constants c_1, c_2 such that*

- *f_n is an n-variate polynomial,*
- $\deg(f_n) \leq n^{c_1}$ *for all large enough n,*
- *for $m \leq n^{c_2}$ there exists an $(n+m)$-variate polynomial $g_{n+m}(\mathbf{x}, \mathbf{y})$ of degree at most n^{c_2} which has an algebraic circuit of size at most n^{c_2}, that satisfies*

$$f_n(\mathbf{x}) = \sum_{\mathbf{a} \in \{0,1\}^m} g_{n+m}(\mathbf{x}, \mathbf{a})$$

Again, the circuit computing $g_{n+m} \in \mathbb{F}[\mathbf{x}, \mathbf{y}]$ may employ constants from a larger extension field $\mathbb{K} \supset \mathbb{F}$.

Analogous to $\mathsf{VP}_{\mathbb{F}}^{[c]}$*, we will define* $\mathsf{VNP}_{\mathbb{F}}^{[c]}$ *to denote*

$$\mathsf{VNP}_{\mathbb{F}}^{[c]} := \left\{ \{f_n\} \in \mathsf{VNP}_{\mathbb{F}} \ : \ f_n \in \mathbb{F}[x_1, \ldots, x_n]^{\leq n^c} \right\}.$$

Just as with VP*, we also consider polynomials $\{f_n\}$ over integers whose coefficients are in $\{-1, 0, 1\}$. Here again the bound is only on the coefficients of f_n and the corresponding circuits may use arbitrary constants from the underlying field, or an extension.* ♢

2.3 Some Preliminaries

For our proofs, we will need the following notion of *universal circuits* defined by Raz [Raz10]. A universal circuit is such that any polynomial computed by a small circuit is a simple projection of it. For the sake of completeness, we also include a proof sketch.

Lemma 2.4 (Universal circuit, [Raz10])**.** *Let $\mathbb{F}$ be any field and $n, s \geq 1$ and $d \geq 0$. Then there exists an algebraic circuit $\mathcal{U}$ of size* $\mathrm{poly}(n, d, s)$ *computing a polynomial in $\mathbb{F}[x_1, \ldots, x_n, y_1, \ldots, y_r]$ with $r \leq \mathrm{poly}(n, d, s)$ such that:*

- $\deg_{\mathbf{x}}(\mathcal{U}(\mathbf{x},\mathbf{y})), \deg_{\mathbf{y}}(\mathcal{U}(\mathbf{x},\mathbf{y})) \leq \mathrm{poly}(d)$;
- *for any* $f \in \mathbb{F}[x_1, \ldots, x_n]$ *with* $\deg_{\mathbf{x}}(f) \leq d$ *that is computable by an algebraic circuit of size* s*, there exists an* $\mathbf{a} \in \mathbb{F}^r$ *such that* $f(\mathbf{x}) = \mathcal{U}(\mathbf{x}, \mathbf{a})$.

Proof. Let f be an n-variate degree d polynomial computable by a circuit C of size s. Using the classical depth reduction result due to Valiant *et al.* [VSBR83], f has a circuit C' of size $s' = \mathrm{poly}(n, d, s)$ and depth $\ell = O(\log d)$ with the following properties (see e.g. [Sap15] for a complete proof).

- All the product gates have fan-in at most 5.
- C' is *layered*, with alternating layers of sum and product gates.
- The layer above the leaves is of product gates, and the root is an addition gate.

We can therefore construct a *layered* universal circuit $\mathcal{U}$ for the given parameters n, d, s. The circuit will have ℓ layers, with $V_1, V_2, \ldots, V_\ell$ being the layers indexed from leaves to the root. So V_ℓ has a single gate, which is the output gate of the circuit, and V_1 has $n + 1$ gates, labeled with the variables $x_1, \ldots, x_n$ and with the constant 1. All the gates in $\mathcal{U}$ are then connected using auxiliary variables $\mathbf{y}$, as follows.

- V_2 has $\leq (n+1)^5$ product gates, with each gate computing a unique monomial of degree at most 5 in the variables $\mathbf{x}$.
- For every odd i with $2 < i < \ell$, the layer V_i has s' addition gates that are all connected to all the gates in the layer V_{i-1}, with each of the wires being labeled by a fresh $\mathbf{y}$-variable.
- For every even i with $2 < i < \ell$, the layer V_i has $\binom{s'}{5}$ product gates, each one multiplying a unique subset of 5 gates from V_{i-1}.

It is now easy to see that $\mathcal{U}$ has at most $\ell(ns')^5$ gates, which is $\mathrm{poly}(n, d, s)$. Also, $\deg(\mathcal{U}) \leq 5^\ell$, which is $\mathrm{poly}(d)$; and $|\mathbf{y}| = r \leq \ell \cdot (ns')^6$, which is $\mathrm{poly}(n, d, s)$. Further, by the depth reduction result [VSBR83], the circuit C' for f can be obtained by setting the auxiliary variables $\mathbf{y}$ appropriately. Since the choice of f was arbitrary, this finishes the proof. $\square$

We will also be using the well-known Polynomial Identity Lemma.

Lemma 2.5 (Polynomial Identity Lemma, [Ore22, DL78, Sch80, Zip79])**.** *Let* $f \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ *be a nonzero polynomial of degree at most* d *and let* S *be a subset of* $\mathbb{F}$ *(or an extension of* $\mathbb{F}$*). Then, the number of zeroes of* f *on the grid* S^n *is at most* $d\,|S|^{n-1}$.

3 Constructible equations for VP over small finite fields

In this section, we prove our main theorem for finite fields. As mentioned in the introduction, our proof uses the existence of non-explicit hitting sets for small circuits. This fact appears to be folklore but we state below the version that can be found in Forbes' thesis [For14].

Lemma 3.1 (Folklore (cf. Lemma 3.2.14 in [For14])). *Let $\mathbb{F}$ be a finite field with $|\mathbb{F}| \geq d^2$. Let $\mathcal{C}(n,d,s)$ be the class of polynomials in $\mathbb{F}[x_1,\ldots,x_n]$ of degree at most d that are computable by fan-in 2 algebraic circuits of size at most s. Then, there is a non-explicit hitting set for $\mathcal{C}$ of size at most $\lceil 2s \cdot (\log n + 2\log s + 4)\rceil$.*

The above lemma shows that over large enough finite fields, there are non-explicit hitting sets of size $O(s^2)$ (when $n, d \leq s$). We now use this to prove Theorem 1.7 which we first restate below.

Theorem 1.7. *Let $\mathbb{F}$ be any finite field of constant size and $c > 0$ be any constant. There is a polynomial family $\{P_{N,c}\} \in \mathsf{VP}_{\mathbb{F}}$ such that for all large enough n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every $\{f_n\} \in \mathsf{VP}_{\mathbb{F}}$, where f_n is an n variate polynomial of degree at most n^c, we have*

$$P_{N,c}(\overline{\mathrm{coeff}}(f_n)) = 0\,.$$

- *There exists a family $\{h_n\}$ of n variate polynomials and degree $\leq n^c$ with coefficients in $\mathbb{F}$ such that*

$$P_{N,c}(\overline{\mathrm{coeff}}(h_n)) \neq 0\,.$$

Proof. Let $d_n = n^c$ and $s_n = n^{\log n}$ (in fact, s_n can be any function that is barely super-polynomial in n). Since $\mathbb{F}$ has constant size, and we need fields of sufficiently large size for invoking Lemma 3.1, we work over an extension $\mathbb{K}_n$ of $\mathbb{F}$ of size at least n^{2c} and at most $O(n^{2c})$. Let $r_n = [\mathbb{K}_n : \mathbb{F}] = O(\log n)$. Note that the elements of $\mathbb{K}_n$ can also be interpreted as vectors over $\mathbb{F}$ via an $\mathbb{F}$-linear map $\Phi : \mathbb{K}_n \to \mathbb{F}^{r_n}$. We can then define for any $i \in [r_n]$, $\Phi_i : \mathbb{K}_n \to \mathbb{F}$ to be its projection to the i-th co-ordinate. That is, $\Phi_i : \alpha \mapsto (\Phi(\alpha))_i$ for every $i \in [r_n]$.

By Lemma 3.1, there are hitting sets in $\mathbb{K}_n^n$ for $\mathcal{C}(n, d_n, s_n)$ of size at most $O(s_n^2)$; let $\mathcal{H}_n$ be such a hitting set.

For $N = \binom{n+d_n}{n}$, let us index the set $[N]$ by the set $\mathbf{x}^{\leq d_n}$ of n-variate monomials of degree at most d_n. For a point $\mathbf{a} \in \mathcal{H}_n$, we define the vector $\mathrm{eval}(\mathbf{a}) \in \mathbb{K}_n^N$ as $\mathrm{eval}(\mathbf{a})_m = m(\mathbf{a})$ where $m \in \mathbf{x}^{\leq d_n}$ (that is, the m-th coordinate is the evaluation of the monomial m at $\mathbf{a}$). To get vectors over $\mathbb{F}$ instead, for each $i \in [r_n]$, we shall define $\mathrm{eval}(\mathbf{a})^{(i)} \in \mathbb{F}^N$ as $\mathrm{eval}(\mathbf{a})_m^{(i)} = \Phi_i(m(\mathbf{a}))$.

We are now ready to define the polynomial family $\{P_N\}$.

$$P_N(z_m \;:\; m \in \mathbf{x}^{\leq d_n}) := \mathrm{OR}(\mathbf{z}) \cdot \prod_{\mathbf{a} \in \mathcal{H}_n} \left(\prod_{i=1}^{r_n} \left(1 - \left(\sum_m z_m \cdot \mathrm{eval}(\mathbf{a})_m^{(i)} \right)^{|\mathbb{F}|-1} \right) \right),$$
$$\text{where } \mathrm{OR}(\mathbf{z}) = \left(1 - \prod_{m \in \mathbf{x}^{\leq d_n}} \left(1 - z_m^{|\mathbb{F}|-1} \right) \right)$$

Constructivity: Note that $\deg(P_N) \leq |\mathbb{F}| \cdot (N + (|\mathcal{H}_n| \cdot r_n)) = O(N + s_n^2 \cdot \log n) = O(N)$ and the above expression immediately yields an $O(N^2)$-sized circuit for P_N. Therefore, the above family $P_N \in \mathsf{VP}_{\mathbb{F}}$.

Usefulness: Now consider any family $\{f_n\} \in \mathsf{VP}_{\mathbb{F}}^{[c]}$; let k be an integer such that for all large enough n we have that f_n is computable by size n^k circuits. We need to show that $P_N(\overline{\mathrm{coeff}}(f_n)) = 0$ for all large enough n.

For any polynomial $g \in \mathbb{F}[x_1, \ldots, x_n]$ with $\deg g \leq n^c$, we have

$$\begin{aligned} P(\overline{\mathrm{coeff}}(g)) &= \mathrm{OR}(\overline{\mathrm{coeff}}(g)) \cdot \prod_{\mathbf{a} \in \mathcal{H}_n} \left(\prod_{i=1}^{r_n} \left(1 - \left(\sum_m \overline{\mathrm{coeff}}(g)_m \cdot \mathrm{eval}(\mathbf{a})_m^{(i)} \right)^{|\mathbb{F}|-1} \right) \right), \\ &= \mathrm{OR}(\overline{\mathrm{coeff}}(g)) \cdot \prod_{\mathbf{a} \in \mathcal{H}_n} \left(\prod_{i=1}^{r_n} \left(1 - (\Phi_i(g(\mathbf{a})))^{|\mathbb{F}|-1} \right) \right), \\ &= \begin{cases} 1 & \text{if } g \neq 0 \text{ and } g(\mathbf{a}) = 0 \text{ for all } \mathbf{a} \in \mathcal{H}_n, \\ 0 & \text{if } g = 0 \text{ or } g(\mathbf{a}) \neq 0 \text{ for some } \mathbf{a} \in \mathcal{H}_n. \end{cases} \end{aligned}$$

If $f_n = 0$, then $\mathrm{OR}(\overline{\mathrm{coeff}}(f_n)) = 0$. Else, if n is chosen large enough, then f_n is computable by circuits of size at most $s_n = n^{\log n}$ and the set $\mathcal{H}_n$ is a hitting set for f_n. Therefore, there is some point $\mathbf{a} \in \mathcal{H}_n$ such that $f_n(\mathbf{a}) \neq 0$. Hence, $\{P_N\}$ vanishes on the coefficient vector of every polynomial in $\mathsf{VP}_{\mathbb{F}}^{[c]}$.

A remark on the largeness: From the definition of P_N, any nonzero $g \in \mathbb{F}[x_1, \ldots, x_n]^{\leq d_n}$ such that $g(\mathbf{a}) = 0$ for all $\mathbf{a} \in \mathcal{H}_n$ will satisfy $P_N(\overline{\mathrm{coeff}}(g)) \neq 0$. If we interpret the coefficients of g as indeterminates, each equation of the form $g(\mathbf{a}) = 0$ introduces one homogeneous linear constraint in these N indeterminates, over the extension $\mathbb{K}_n$. Each such constraint can be interpreted as $r_n = O(\log n)$ homogeneous linear constraints, over $\mathbb{F}$. Since $|\mathcal{H}_n| \ll N$, the set of g's that are not annihilated by P_N form a subspace of dimension at least $N - O(|\mathcal{H}_n| \log n)$. Thus, there are at least $\left(|\mathbb{F}|^{N - O(|\mathcal{H}_n| \log n)} - 1\right)$ many g's such that $P_N(\overline{\mathrm{coeff}}(g)) \neq 0$. $\square$

4 Constructible equations for VP with coefficients in $\{-1, 0, 1\}$

In this section, we prove Theorem 1.6. As before, our proof uses the existence of non-explicit hitting sets for circuits of small size. When the underlying field is $\mathbb{C}$, their existence is known due to the results of Heintz and Schnorr [HS80] as stated below.

Lemma 4.1 (Hitting sets for efficiently computable polynomials [HS80])**.** *There are (non-explicit) hitting sets $\mathcal{H}$ for $\mathcal{C}(n, d, s)$ (the set of all n-variate polynomials with degree at most d that are computable by algebraic circuits of size at most s), such that $\mathcal{H} \subset [(sd)^2]^n$ and $|\mathcal{H}| = \mathrm{poly}(s)$.*

We now prove our main theorem which we restate below.

Theorem 1.6. *Let $c > 0$ be any constant. There is a polynomial family $\{P_{N,c}\} \in \mathsf{VP}_{\mathbb{Q}}$ such that for all large n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every family* $\{f_n\} \in \mathsf{VP}_{\mathbb{C}}$, *where* f_n *is an* n *variate polynomial of degree at most* n^c *and coefficients in* $\{-1,0,1\}$, *we have*

$$P_{N,c}(\overline{\mathrm{coeff}}(f_n)) = 0\,,$$

where $\overline{\mathrm{coeff}}(f_n)$ *is the coefficient vector of* f_n.

- *There exists a family* $\{h_n\}$ *of* n *variate polynomials and degree* $\leq n^c$ *with coefficients in* $\{-1,0,1\}$ *such that*

$$P_{N,c}(\overline{\mathrm{coeff}}(h_n)) \neq 0\,.$$

As mentioned earlier, the proof would also generalise in a straightforward manner for polynomial families $\{f_n\} \in \mathsf{VP}_{\mathbb{Z}}^{[c]}$ whose coefficients are bounded by N. We state this for polynomials whose coefficients are in $\{-1,0,1\}$ just to avoid cumbersome notation.

Proof. The proof will proceed similar to the proof of Theorem 1.7, with a careful use of the Chinese Remainder Theorem.

Let $d_n = n^c$ and $s_n = n^{\log n}$ (again, s_n needs to be barely super-polynomial in n). For $N = \binom{n+d_n}{n}$, let us index the set $[N]$ by the set $\mathbf{x}^{\leq d_n}$ of n-variate monomials of degree at most d_n. For a point $\mathbf{a} \in \mathbb{Z}^n$, we define the vector $\mathrm{eval}(\mathbf{a}) \in \mathbb{Q}^N$ as $\mathrm{eval}(\mathbf{a})_m = m(\mathbf{a})$ where $m \in \mathbf{x}^{\leq d_n}$ (that is, the m-th coordinate is the evaluation of the monomial m at $\mathbf{a}$). Therefore, for any n-variate polynomial f of degree at most d_n, we have $f(\mathbf{a}) = \left\langle \overline{\mathrm{coeff}}(f), \mathrm{eval}(\mathbf{a}) \right\rangle$.

Let $B_n = (s_n \cdot d_n)^2$. By Lemma 4.1, there are hitting sets in $[B]^n$ of size $\mathrm{poly}(s_n)$ for the class $\mathcal{C}(n, d_n, s_n)$ (of n-variate polynomials, of degree at most d_n that are computable by circuits of size s_n) with coefficients in $\Delta = \{-1,0,1\}$. Let $\mathcal{H}_n$ be one such set. Note that for any n-variate polynomial f of degree at most d_n and coefficients in Δ, and any $\mathbf{a} \in \mathcal{H}_n$, we have $|f(\mathbf{a})| \leq N \cdot B^{d_n}$, which unfortunately is not $\mathrm{poly}(N)$. However, we can work with some "proxy evaluations" by simulating Chinese Remaindering.

For any $\mathbf{a} \in \mathcal{H}_n$ and a positive integer r, define the vector $\widetilde{\mathrm{eval}}_r(\mathbf{a})$ as follows:

$$\widetilde{\mathrm{eval}}_r(\mathbf{a})_m := (m(\mathbf{a}) \bmod r) \quad \text{for all } m \in \mathbf{x}^{\leq d_n}.$$

It is to be stressed that $\widetilde{\mathrm{eval}}_r(\mathbf{a})$ is a vector over $\mathbb{Q}$, whose entries are integers between 0 and $r-1$.

Claim 4.2. *Suppose* f *is a polynomial with integer coefficients, and* $\mathbf{a} \in \mathbb{Z}^n$. *If* $f(\mathbf{a}) \neq 0$ *and* $|f(\mathbf{a})| \leq M$, *then there is some* $r \leq O((\log M)^2)$ *such that*

$$\left\langle \overline{\mathrm{coeff}}(f), \widetilde{\mathrm{eval}}_r(\mathbf{a}) \right\rangle \neq 0 \bmod r.$$

Proof of claim. Let $\ell = \log(M+1)$, note that the LCM of the set $[\ell^2]$ is at least $2^\ell > M$. Since $f(\mathbf{a})$

is a nonzero integer with $|f(\mathbf{a})| \le M$, by the Chinese Remainder Theorem there is some prime $r \le \ell^2$ such that $f(\mathbf{a}) \ne 0 \bmod r$.

$$\begin{aligned}\left\langle \overline{\mathrm{coeff}}(f), \widetilde{\mathrm{eval}}_r(\mathbf{a}) \right\rangle &= \left\langle \overline{\mathrm{coeff}}(f), \mathrm{eval}_r(\mathbf{a}) \right\rangle \bmod r \\ &= f(\mathbf{a}) \bmod r \\ &\ne 0 \bmod r \qquad \square\end{aligned}$$

Let $M = N \cdot B^{d_n}$ and $\ell = \log(M+1)$. For any $r \in [\ell^2]$, any $\mathbf{a} \in \mathcal{H}_n$ and n-variate polynomial f of degree at most d_n and coefficients from Δ, we have

$$\left| \left\langle \overline{\mathrm{coeff}}(f), \widetilde{\mathrm{eval}}_r(\mathbf{a}) \right\rangle \right| \le N \cdot \ell^2 =: R.$$

We are now ready to define the polynomial family $\{P_N\}$.

$$\begin{aligned} P_N(z_m \;:\; m \in \mathbf{x}^{\le n}) &= \mathrm{OR}(\mathbf{z}) \cdot \prod_{\mathbf{a} \in \mathcal{H}_n} \prod_{r=2}^{\ell^2} Q_r\left(\left\langle \mathbf{z}, \widetilde{\mathrm{eval}}_r(\mathbf{a}) \right\rangle \right), \\ \text{where } Q_r(x) &= \prod_{\substack{i \in [-R,\ldots,R] \\ i \bmod r \ne 0}} (x - i), \\ \mathrm{OR}(\mathbf{z}) &= 1 - \prod_{m \in \mathbf{x}^{\le d_n}} (1 - z_m) \end{aligned}$$

Constructivity: For our setting of the underlying parameters, $|\mathcal{H}_n| \le n^{O(\log n)}, B_n \le n^{O(\log n)}$, $M \le N \cdot n^{O(d_n \log n)}$ and $\ell = \mathrm{poly}(n)$; and $R \le O(N \,\mathrm{poly}(n)) = \tilde{O}(N)$. Therefore, P_N is a polynomial of degree at most $\tilde{O}(N^2)$. Moreover, the above expression also shows that P_N is computable by a circuit of size $\tilde{O}(N^3)$ and hence $\{P_N\} \in \mathsf{VP}$.

Usefulness: Fix a polynomial family $\{f_n\} \in \mathsf{VP}^{[c]}$ such that the coefficients of f_n are in $\{-1, 0, 1\}$ for all n. Let k be an integer such that for all large enough n we have that f_n is computable by size n^k circuits. We need to show that $P_N(\overline{\mathrm{coeff}}(f_n)) = 0$ for all large enough n. Note that we have $\mathrm{OR}(\overline{\mathrm{coeff}}(f_n)) \ne 0$ if f_n is nonzero, and 0 if $f_n = 0$. Hence, it suffices to show that $P_N(\overline{\mathrm{coeff}}(f_n)) = 0$ for nonzero f_n.

For any large enough n so that $0 \ne f_n$ is computable by circuits of size at most $s_n = n^{\log n}$ and the set $\mathcal{H}_n$ is a hitting set for f_n, we know that $f_n(\mathbf{a}) \ne 0$ for some $\mathbf{a} \in \mathcal{H}_n$. Therefore, for some $r \in [\ell^2]$, we have that $\left\langle \overline{\mathrm{coeff}}(f), \widetilde{\mathrm{eval}}_r(\mathbf{a}) \right\rangle$ is a nonzero integer in $\{-R, \ldots, R\}$ that is not divisible by r. Hence, we have

$$\begin{aligned} Q_r\left(\left\langle \overline{\mathrm{coeff}}(f), \widetilde{\mathrm{eval}}_r(\mathbf{a}) \right\rangle \right) &= 0, \\ \implies P(\overline{\mathrm{coeff}}(f)) &= 0. \end{aligned}$$

A remark on the largeness: From the definition of P_N, any nonzero $g \in \mathbb{F}[x_1, \ldots, x_n]^{\leq d_n}$ such that $g(\mathbf{a}) = \left\langle \overline{\mathrm{coeff}}(g), \mathrm{eval}(\mathbf{a}) \right\rangle = 0$ for all $\mathbf{a} \in \mathcal{H}_n$ will satisfy $P_N(\overline{\mathrm{coeff}}(g)) \neq 0$. In order to show that there are many such g's with coefficients in $\{-1, 0, 1\}$, we use a pigeon-hole argument, which is essentially an instance of a lemma of Siegel [Sie14]. For completeness, we include a sketch of the argument here.

Consider the map $\Gamma : \mathbb{Z}^N \to \mathbb{Z}^{|\mathcal{H}_n|}$ defined as

$$\Gamma(z_m \ : \ m \in \mathbf{x}^{\leq d_n}) := (\langle \mathbf{z}, \mathrm{eval}(\mathbf{a}) \rangle \ : \ \mathbf{a} \in \mathcal{H}_n)$$

The map Γ is linear in the sense that $\Gamma(\mathbf{z} + \mathbf{z}') = \Gamma(\mathbf{z}) + \Gamma(\mathbf{z}')$. Consider the restriction of Γ on just $\{0,1\}^N$; the range of Γ under this restriction is $\{-M, \ldots, M\}^{|\mathcal{H}_n|}$. Hence, by the pigeon-hole-principle there must be some $\mathbf{b} \in \{-M, \ldots, M\}^{|\mathcal{H}_n|}$ with at least $2^N / (2M+1)^{|\mathcal{H}_n|}$ pre-images inside $\{0,1\}^N$. If $\mathbf{h}_0$ is any fixed preimage, then

$$\left\{ \mathbf{h} - \mathbf{h}_0 \in \{-1, 0, 1\}^N \ : \ \mathbf{h} \in \Gamma^{-1}(\mathbf{b}) \cap \{0,1\}^N \right\}$$

are all coefficient vectors of polynomials $g \in \mathbb{Z}[x_1, \ldots, x_n]^{\leq d_n}$ with coefficients in $\{-1, 0, 1\}$ whose coefficient vectors are not zeroes of P_N. $\square$

It is worth mentioning that there are 3^N possible polynomials in $\mathbb{Z}[x_1, \ldots, x_n]^{\leq d_n}$ with coefficients in $\{-1, 0, 1\}$. The above remark on the largeness shows that there are $2^{N - q(n)}$ many polynomials g such that $P_N(\overline{\mathrm{coeff}}(g)) \neq 0$; for some $q(n) = n^{O(\log n)}$.

5 Equations for VNP

We shall now state and prove the VNP analogs of Theorem 1.6 and Theorem 1.7. First, we have the following definition.

Definition 5.1 (Definability of Polynomials). *For $s \geq 1$, a polynomial f_n is said to be s-*definable *if there exists a polynomial $g_s \in \mathcal{C}(s, s, s)$ such that for $m = s - n$,*

$$f_n(\mathbf{x}) = \sum_{\alpha \in \{0,1\}^m} g_s(\mathbf{x}, \alpha) .$$

Further, let us denote by $\mathcal{D}(n, d, s)$ the class of all n-variate polynomials of degree d that are s-definable. $\diamondsuit$

Remark 5.2. *Note that for every family $\{f_n\} \in$ VNP, there is a polynomially bounded function $s(n) > n, d(n)$ such that f_n is $s(n)$-definable, for all large n.* $\diamondsuit$

5.1 VNP over Small Finite Fields

As in the VP case, we will need the existence of non-explicit hitting sets. A slight modification to the proof of Lemma 3.2.14 in [For14]) gives us the following statement.

Lemma 5.3. *Let $\mathbb{F}$ be a finite field with $|\mathbb{F}| \geq d^2$. Let $\mathcal{D}(n, d, s)$ be the class of polynomials in $\mathbb{F}[x_1, \ldots, x_n]$ of degree at most d that are s-definable. Then, there is a non-explicit hitting set $\mathcal{H}$ for $\mathcal{D}(n, d, s)$ of size at most $\lceil 2s \cdot (3 \log s + 4) \rceil$.*

Proof. In order to prove the existence of a hitting set for the class $\mathcal{D}(n, d, s)$, we will need a bound on the number of polynomials in the class $\mathcal{D}(n, d, s)$ as well as a bound on the size of an explicit hitting set for the class of n-variate degree at most d polynomials. These two bounds are summarized in the following claims, proofs of which can be found in [For14].

Claim 5.4 (Lemma 3.1.6 in [For14])**.** *Let $\mathbb{F}$ be a finite field and $n, s \geq 1$. There are at most $(8n\,|\mathbb{F}|\,s^2)^s$ n-variate polynomials in $\mathbb{F}[\mathbf{x}]$ computable by (single-output) algebraic circuits of size $\leq s$ and fan-in ≤ 2.*

Claim 5.5 (Lemma 3.2.13 in [For14])**.** *Let $\mathbb{F}$ be a finite field with $|\mathbb{F}| \geq (1+\varepsilon)d$. Let $\mathcal{C} \subseteq \mathbb{F}[\mathbf{x}]$ be a finite set of n-variate polynomials of degree $< d$. Then there is a non-explicit hitting set for $\mathcal{C}$ of size $\leq \lceil \log_{1+\varepsilon} |\mathcal{C}| \rceil$.*

Note that by definition, the number of n-variate polynomials that are s-definable is at most the number of polynomials in $\mathcal{C}(s, s, s)$; the class of s-variate polynomials of degree $\leq s$ computable by size s algebraic circuits of fan-in ≤ 2. Thus, by Claim 5.4, $|\mathcal{D}(n, d, s)| \leq (8\,|\mathbb{F}|\,s^3)^s$.

The rest of the proof follows exactly along the lines of the proof of Lemma 3.2.14 in [For14].

As $|\mathbb{F}| \geq d^2$, we have $d \leq |\mathbb{F}|$, and so $|\mathbb{F}| \geq (1+\varepsilon)d$ for $(1+\varepsilon) = \sqrt{|\mathbb{F}|}$. Thus, using $\varepsilon = \sqrt{|\mathbb{F}|} - 1$ in Claim 5.5, we get that there is a non-explicit hitting set $\mathcal{H}$ for $\mathcal{D}(n, d, s)$ of size at most

$$\left\lceil \log_{\sqrt{|\mathbb{F}|}} |\mathcal{D}(n, d, s)| \right\rceil \leq \left\lceil \log_{\sqrt{|\mathbb{F}|}} (8\,|\mathbb{F}|\,s^3)^s \right\rceil = \left\lceil s(2 + 2\log_{|\mathbb{F}|}(8s^3)) \right\rceil = \left\lceil s(2 + 6\log_{|\mathbb{F}|}(2s)) \right\rceil$$

Finally, as $|\mathbb{F}| \geq 2$, we have

$$|\mathcal{H}| \leq \lceil s \cdot (2 + 6\log(2s)) \rceil = \lceil 2s \cdot (1 + 3\log(2s)) \rceil = \lceil 2s \cdot (3 \log s + 4) \rceil .$$

This completes the proof. ☐

By Remark 5.2, we have that over large enough finite fields, there are non-explicit hitting sets of size $O(s^2)$ for VNP. Since the proof of Theorem 1.7 does not use any property of VP except for the existence of non-trivial hitting sets, we get the following theorem. The proof is omitted, since it is exactly the same except we use Lemma 5.3 instead of Lemma 3.1.

Theorem 1.9. *Let $\mathbb{F}$ be any finite field of constant size and $c > 0$ be any constant. There is a polynomial family $\{Q_{N,c}\} \in \mathsf{VP}_{\mathbb{F}}$ such that for all large enough n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every family $\{f_n\} \in \mathsf{VNP}_{\mathbb{F}}$, where f_n is an n variate polynomial of degree at most n^c, we have*

$$Q_{N,c}(\overline{\mathrm{coeff}}(f_n)) = 0 .$$

- *There exists a family $\{h_n\}$ of n variate polynomials and degree $\leq n^c$ with coefficients in $\mathbb{F}$ such that*

$$Q_{N,c}(\overline{\mathsf{coeff}}(h_n)) \neq 0\,.$$

5.2 Polynomials in VNP with Small Integer Coefficients

Our argument will be identical to that in Section 4, for which we will need a statement analogous to Lemma 4.1 showing the existence of non-explicit hitting sets for VNP with small bit-complexity. We will first give a universal map for the polynomials in VNP, analogous to Lemma 2.4; for which we need the following lemma.

Lemma 5.6 (Coefficient Vectors of Definable Polynomials)**.** *Let $f \in \mathbb{C}[\mathbf{x}]$ be an n-variate polynomial of degree d that is s-definable. Then there exists an s-variate polynomial g and a* linear *map $L_{n,d,s}$, such that $\overline{\mathsf{coeff}}(f) = L(\overline{\mathsf{coeff}}(g))$. Furthermore, the map L depends* solely *on n, d and s.*

Proof. Let $m = s - n$. Since f is s-definable, there is an s-variate polynomial $g(\mathbf{x}, \mathbf{w})$ of degree at most s as follows.

$$f(\mathbf{x}) = \sum_{\alpha \in \{0,1\}^m} g(\mathbf{x}, \mathbf{w} = \alpha)$$

Now observe that for any monomial $\mathbf{x}^{\mathbf{e}} \in \mathbf{x}^{\leq d}$,

$$\begin{aligned}
\mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}}(f) &= \mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}}\left(\sum_{\alpha \in \{0,1\}^m} g(\mathbf{x}, \alpha)\right) \\
&= \sum_{\alpha \in \{0,1\}^m} \mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}}\left(g(\mathbf{x}, \alpha)\right) \\
&= \sum_{\alpha \in \{0,1\}^m} \mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}}\left(\sum_{\mathbf{w}^{\mathbf{a}} \in \mathbf{w}^{\leq s}} \alpha^{\mathbf{a}}\, \mathsf{coeff}_{\mathbf{w}^{\mathbf{a}}}(g(\mathbf{x}, \mathbf{w}))\right) \\
&= \sum_{\mathbf{w}^{\mathbf{a}} \in \mathbf{w}^{\leq s}} \left(\sum_{\alpha \in \{0,1\}^m} \alpha^{\mathbf{a}}\right) \mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}\mathbf{w}^{\mathbf{a}}}(g) \\
&= \sum_{\mathbf{w}^{\mathbf{a}} \in \mathbf{w}^{\leq s}} 2^{(m - |\mathrm{supp}(\mathbf{a})|)}\, \mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}\mathbf{w}^{\mathbf{a}}}(g)
\end{aligned}$$

Now we can define the desired map $L : \mathbb{C}^M \to \mathbb{C}^N$ for $M = \binom{s+s}{s}$ and $N = \binom{n+d}{n}$, as follows.

$$L_{\mathbf{e}}(\overline{\mathsf{coeff}}(g)) = \sum_{\mathbf{w}^{\mathbf{a}} \in \mathbf{w}^{\leq s}} 2^{(m - |\mathrm{supp}(\mathbf{a})|)}\, \mathsf{coeff}_{\mathbf{x}^{\mathbf{e}}\mathbf{w}^{\mathbf{a}}}(g) \qquad \forall \mathbf{e} \in [N]$$

□

Lemma 5.7 (Universal Map for Definable Polynomials)**.** *Let $s \geq n \geq 1$ and $d \geq 0$. Then for $N = \binom{n+d}{n}$ there exists a polynomial map $\mathcal{U}(\mathbf{y}) : \mathbb{C}^r \to \mathbb{C}^N$ with $r \leq \mathrm{poly}(n, d, s)$ such that:*

- $\deg(\mathcal{U}(\mathbf{y})) \leq \mathrm{poly}(s)$*;*

- *for any $f \in \mathbb{C}[x_1, \ldots, x_n]$ with $\deg_{\mathbf{x}}(f) \leq d$ that is s-definable, there exists an $\mathbf{a} \in \mathbb{C}^r$ such that $\overline{\text{coeff}}(f) = \mathcal{U}(\mathbf{a})$.*

Proof. Let $\mathcal{D}(n, d, s)$ be the class of all n-variate, degree d polynomials that are s-definable and suppose $f_n(\mathbf{v}) \in \mathcal{D}(n, d, s)$. Then by Lemma 5.6 there exists an s-variate, degree s polynomial $g_s \in \mathcal{C}(s, s, s)$ such that the coefficients of f_n are obtained by taking suitable *linear* combinations of the coefficients of g_s. Therefore we will now shift our focus to the coefficient vectors of polynomials from $\mathcal{C}(s, s, s)$.

Using Lemma 2.4 for number of variables, degree and size, all bounded by s, we get a universal circuit $\mathcal{U}(\mathbf{x}, \mathbf{y})$ for $\mathcal{C}(s, s, s)$ with $|\mathbf{y}| = r \leq s^k$ for some constant k. We will assume without loss of generality that $\deg_{\mathbf{y}}(\mathcal{U}) \leq s^k$. Now for $M = \binom{s+s}{s}$, we can view $\mathcal{U}(\mathbf{x}, \mathbf{y})$ as a polynomial map $\mathcal{U} : \mathbb{C}^r \to \mathbb{C}^M$ given by $\mathcal{U}(\mathbf{y}) = (\mathcal{U}(\mathbf{y}_1), \ldots, \mathcal{U}(\mathbf{y}_M))$, where $\mathcal{U}_m(\mathbf{y})$ is the coefficient of the monomial $m \in \mathbf{x}^{\leq s}$ in the polynomial $\mathcal{U}(\mathbf{x}, \mathbf{y})$. Note that the degree of every such $\mathcal{U}_m(\mathbf{y})$ is at most s^k.

Now by Lemma 2.4, for every $g_s \in \mathcal{C}(s, s, s)$ the coefficient vector of g_s is in the image of $\mathcal{U}$. Therefore, for $N = \binom{n+d}{n}$, let $L : \mathbb{C}^M \to \mathbb{C}^N$ be the linear map given by Lemma 5.6. Then for every $f_n \in \mathcal{D}(n, d, s)$ the coefficient vector of f_n is in the image of $(L \circ \mathcal{U}) : \mathbb{C}^r \to \mathbb{C}^N$. Further, since L is a linear map, the degree of $(L \circ \mathcal{U})$ is also bounded by $s^k = \text{poly}(s)$. $\square$

We can then prove the existence of non-explicit hitting sets of small bit-complexity even for the class of efficiently definable polynomials with small integer coefficients.

Number of efficiently definable polynomials with small coefficients

We first need to bound the number of definable polynomials with small coefficients. The lemma below is a slight modification of a result of Hrubeš and Yehudayoff [HY11, Claim 3.6]. The proof uses some basic algebraic geometry notions such as *dimension and degree of varieties* and also employs Bézout's theorem, which may be found in most algebraic geometry texts (e.g. [DS13]).

Lemma 5.8 ([HY11]). *Let $V \in \mathbb{C}^n$ be an irreducible algebraic variety of dimension k and degree r. Suppose $F = (F_1, \ldots, F_m)$ with $F_i \in \mathbb{F}[x_1, \ldots, x_n]^{\leq d}$ is a polynomial map. Then, for $\Delta \subset \mathbb{Z}$,*

$$|F(V) \cap \Delta^m| \leq r \cdot (|\Delta| \cdot d)^k.$$

Proof. The proof is by induction on the dimension k. For the base case of $k = 0$, we would have $|V| = 1$ as V is irreducible and hence $|F(V)| = 1$.

For each $i \in [m]$ and $b \in \Delta$, define $V_{i,b} = V \cap F_i^{-1}(b)$. Suppose for every $i \in [m]$ there is just a single $b \in \Delta$ such that $V_{i,b} \neq \emptyset$, then clearly $|F(V) \cap \Delta^m| \leq 1$. Otherwise, let i be such that at least two of $\{V_{i,b} \ : \ b \in \Delta\}$ are non-empty. Since at least two of them are non-empty, each $V_{i,b}$ is a proper subvariety of V and hence $\dim(V_{i,b}) < \dim(V)$. Let the non-empty varieties be decomposed into irreducible varieties as

$$V_{i,b} = V_{i,b}^{(1)} \cup \cdots \cup V_{i,b}^{(t_b)}.$$

By Bézout's theorem (see e.g., [DS13]), we also have $\sum_j \deg(V_{i,b}^{(j)}) \leq d \cdot \deg(V_{i,b})$. Then,

$$\begin{aligned}
F(V) \cap \Delta^m &\subseteq \bigcup_{b \in \Delta} F(V_{i,b}) \cap \Delta^m \\
&= \bigcup_{b \in \Delta} \bigcup_{j \in [t_b]} F(V_{i,b}^{(t)}) \cap \Delta^m \\
\implies |F(V) \cap \Delta^m| &\leq \sum_{b \in \Delta} \sum_{j \in [t_b]} \deg(V_{i,b}^{(j)})(|\Delta| \cdot d)^{k-1} \\
&\leq \deg(V) \cdot (|\Delta| \cdot d)^k.
\end{aligned}$$

□

Corollary 5.9. *The number of polynomials with coefficients in $\Delta \subset \mathbb{Z}$ that are s-definable is at most $(|\Delta| \cdot s)^{\mathrm{poly}(s)}$.*

Proof. By Lemma 5.7, we know that any s-definable polynomial can be seen as an image of a universal map $\mathcal{U} : \mathbb{F}[\mathbf{x}, \mathbf{y}] \to \mathbb{F}[\mathbf{x}]$. Thus, if we view $\mathcal{U}$ as a polynomial map of the form $\mathcal{U} = (\mathcal{U}_1(\mathbf{y}), \ldots, \mathcal{U}_N(\mathbf{y}))$, then any polynomial of the type we wish to count is contained in the set $(\mathcal{U}(\mathbb{C}^{|\mathbf{y}|}) \cap \Delta^N)$. Here $\mathcal{U}_m$ computes the coefficient of the m-th monomial in $\mathbf{x}$, and by Lemma 5.7, $|\mathbf{y}| = \mathrm{poly}(s)$ and $\deg(\mathcal{U}_m) = \mathrm{poly}(s)$ for every $m \in [N]$.

Finally, note that $\mathbb{C}^{|\mathbf{y}|}$ is an irreducible variety which has degree 1 and dimension $|\mathbf{y}|$. Thus using Lemma 5.8, we have that the number of polynomials with coefficients in Δ that are s-definable is at most $(|\Delta| \cdot \mathrm{poly}(s))^{\mathrm{poly}(s)} \leq (|\Delta| \cdot s)^{\mathrm{poly}(s)}$. □

Existence of hitting sets with low bit-complexity

Lemma 5.10 (Hitting sets for efficiently definable polynomials)**.** *Let $\Delta \subset \mathbb{Z}$. There are (non-explicit) hitting sets $\mathcal{H}$ for $\mathcal{D}(n, d, s)$ (the set of all n-variate polynomials with degree at most d that are s-definable) with coefficients in Δ, such that $\mathcal{H} \subset [ds\,|\Delta|]^n$ and $|\mathcal{H}| = \mathrm{poly}(s)$.*

Proof. Let $\mathcal{H}$ be a uniformly random subset of size $t = \mathrm{poly}(s)$ of the grid $[ds\,|\Delta|]^n$. For any nonzero polynomial $f(x_1, \ldots, x_n) \in \mathcal{D}(n, d, s)$, by the Polynomial Identity Lemma (Lemma 2.5) we know that the number of zeroes of any n-variate degree d polynomial f on the grid $[ds\,|\Delta|]^n$ is upper bounded by $d(ds\,|\Delta|)^{n-1} = \frac{1}{s|\Delta|}(ds\,|\Delta|)^n$. Thus, the probability that $\mathcal{H}$ is not a hitting set for a fixed $f \in \mathcal{D}(n, d, s)$ is equal to $\left(\binom{(ds|\Delta|)^n / s|\Delta|}{t} / \binom{(ds|\Delta|)^n}{t}\right)$, which can be upper bounded by $(1/s\,|\Delta|)^{\Omega(t)}$.

Let $\mathcal{D}'$ be the set of all polynomials in $\mathcal{D}(n, d, s)$ whose coefficients are from Δ. Therefore, the probability that $\mathcal{H}$ is *not* a hitting set for $\mathcal{D}'$ is upper bounded by:

$$\begin{aligned}
\Pr_{\mathbf{a}_1, \ldots, \mathbf{a}_t \in [ds|\Delta|]^n} \left[\{\mathbf{a}_1, \ldots, \mathbf{a}_t\} \text{ is not a hitting set for } \mathcal{C}' \right] &\leq |\mathcal{C}'| \cdot \left(\frac{1}{s\,|\Delta|}\right)^{\Omega(t)} \\
&\leq (s\,|\Delta|)^{\mathrm{poly}(s) - \Omega(t)} \quad \text{(Corollary 5.9)} \\
&\ll 1. \quad \text{(if } t = \mathrm{poly}(s) \text{ large enough)}
\end{aligned}$$

Hence, there exist poly(s)-sized hitting sets $\mathcal{H} \subset [ds\,|\Delta|]^n$ for $\mathcal{D}'$. $\square$

We can now prove the following theorem along the lines of Theorem 1.6. The proof of Theorem 1.6 almost directly extends here as it does not assume anything about VP except for the existence of non-explicit hitting sets of small bit-complexity, which here is given by Lemma 5.10. We omit the proof to avoid repetition.

Theorem 1.8. *Let $c > 0$ be any constant. There is a polynomial family $\{Q_{N,c}\} \in \mathsf{VP}_{\mathbb{Q}}$ such that for all large n and $N = \binom{n+n^c}{n}$, the following are true.*

- *For every family $\{f_n\} \in \mathsf{VNP}_{\mathbb{C}}$, where f_n is an n variate polynomial of degree at most n^c and coefficients in $\{-1, 0, 1\}$, we have*

$$Q_{N,c}(\overline{\mathrm{coeff}}(f_n)) = 0\,.$$

- *There exists a family $\{h_n\}$ of n variate polynomials and $\leq n^c$ with coefficients in $\{-1, 0, 1\}$ such that*

$$Q_{N,c}(\overline{\mathrm{coeff}}(h_n)) \neq 0\,.$$

Remark 5.11. *In the setting of small integer coefficients (or over small finite fields), there exist constructible low degree equations for both* VP *and* VNP*. However, this does not mean that the framework of algebraically natural proofs cannot be used for separating* VP *and* VNP*. It is worth noting that the equations for* VP *and* VNP *that are constructible in* VP *seem to be different from each other as they use different universal map constructions. This also highlights the fact that any separation of* VP *and* VNP *(in the bounded coefficient setting) cannot rely solely on the degree and circuit size of their equations, but might need to look more carefully at the structure and properties of these equations.* $\Diamond$

6 Open problems

We conclude with some open questions.

- The most natural question here is to extend the results in this paper to the entire class VP over all fields. Our proofs crucially use the complexity of the coefficients and it is not clear if ideas from this paper can be used for such an extension. A first step towards this generalization would be to understand the complexity of equations for constant-free versions of the classes VP and VNP, namely VP^0 and VNP^0.
- In general, proving non-trivial upper (and lower) bounds on the circuit complexity and degree of equations for varieties associated with natural algebraic models is an interesting question. In addition to proving such bounds for VP as mentioned in the item above, it is also of great interest to prove such bounds for other models, like formulas or algebraic branching programs.

Acknowledgements

The authors from TIFR acknowledge support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.01-0500.

Mrinal thanks Rahul Santhanam and Ben Lee Volk for many insightful conversations about algebraic natural proofs and succinct hitting sets.

We also thank Ben Lee Volk for a careful reading of the paper and insightful comments which helped improve the presentation.

We thank the reviewers of FOCS 2020 for their valuable suggestions and comments on an earlier version of this paper.

References

[AD08] Scott Aaranson and Andrew Drucker. Arithmetic natural proofs theory is sought. Shtetl Optimized: Scott Aaranson's Blog, 2008.

[AW09] Scott Aaronson and Avi Wigderson. Algebrization: A New Barrier in Complexity Theory. *ACM Trans. Comput. Theory*, 1(1), February 2009.

[BGS75] Theodore Baker, John Gill, and Robert Solovay. Relativizations of the $\mathcal{P} = ?\mathcal{NP}$ Question. *SIAM Journal on Computing*, 4(4):431–442, 1975.

[BIJL18] Markus Bläser, Christian Ikenmeyer, Gorav Jindal, and Vladimir Lysikov. Generalized matrix completion and algebraic natural proofs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1193–1206. ACM, 2018.

[BIL+19] Markus Bläser, Christian Ikenmeyer, Vladimir Lysikov, Anurag Pandey, and Frank-Olaf Schreyer. Variety Membership Testing, Algebraic Natural Proofs, and Geometric Complexity Theory. *CoRR*, abs/1911.02534, 2019.

[BS83] Walter Baur and Volker Strassen. The Complexity of Partial Derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[Cho11] Timothy Y. Chow. Almost-natural proofs. *J. Comput. Syst. Sci.*, 77(4):728–737, 2011.

[DL78] Richard A. DeMillo and Richard J. Lipton. A Probabilistic Remark on Algebraic Program Testing. *Information Processing Letters*, 7(4):193–195, 1978.

[DS13] Vladimir Ivanovich Danilov and Vyacheslav V Shokurov. *Algebraic Geometry I: Algebraic Curves, Algebraic Manifolds and Schemes*, volume 23. Springer Science & Business Media, 2013.

[EGOW18] Klim Efremenko, Ankit Garg, Rafael Oliveira, and Avi Wigderson. Barriers for Rank Methods in Arithmetic Complexity. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pages 1:1–1:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[For14] Michael Forbes. *Polynomial Identity Testing of Read-Once Oblivious Algebraic Branching Programs*. PhD thesis, Massachusetts Institute of Technology, 2014.

[FSS84] Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical systems theory*, 17(1):13–27, 1984.

[FSV18] Michael A. Forbes, Amir Shpilka, and Ben Lee Volk. Succinct Hitting Sets and Barriers to Proving Lower Bounds for Algebraic Circuits. *Theory of Computing*, 14(1):1–45, 2018.

[GKSS17] Joshua A. Grochow, Mrinal Kumar, Michael E. Saks, and Shubhangi Saraf. Towards an algebraic natural proofs barrier via polynomial identity testing. *CoRR*, abs/1701.01717, 2017. Pre-print available at `arXiv:1701.01717`.

[GMOW19] Ankit Garg, Visu Makam, Rafael Oliveira, and Avi Wigderson. More Barriers for Rank Methods, via a "numeric to Symbolic" Transfer. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 824–844. IEEE Computer Society, 2019.

[Gro15] Joshua A. Grochow. Unifying Known Lower Bounds via Geometric Complexity Theory. *Comput. Complex.*, 24(2):393–475, 2015.

[Hås86] Johan Håstad. Almost Optimal Lower Bounds for Small Depth Circuits. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, STOC '86, page 6–20, New York, NY, USA, 1986. Association for Computing Machinery.

[HS80] Joos Heintz and Claus-Peter Schnorr. Testing Polynomials which Are Easy to Compute (Extended Abstract). In *Proceedings of the 12th Annual ACM Symposium on Theory of Computing (STOC 1980)*, pages 262–272. ACM, 1980.

[HY11] Pavel Hrubeš and Amir Yehudayoff. Arithmetic Complexity in Ring Extensions. *Theory of Computing*, 7(8):119–129, 2011.

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing Polynomial Identity Tests Means Proving Circuit Lower Bounds. *Computational Complexity*, 13(1-2):1–46, 2004. Preliminary version in the *35th Annual ACM Symposium on Theory of Computing (STOC 2003)*.

[KV20] Mrinal Kumar and Ben Lee Volk. A Polynomial Degree Bound on Defining Equations of Non-rigid Matrices and Small Linear Circuits. *CoRR*, abs/2003.12938, 2020. Preprint available at `arXiv:2003.12938`.

[Ore22] Øystein Ore. Über höhere Kongruenzen. *Norsk Mat. Forenings Skrifter*, 1(7):15, 1922.

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Raz10] Ran Raz. Elusive Functions and Lower Bounds for Arithmetic Circuits. *Theory of Computing*, 6(1):135–177, 2010.

[RR97] Alexander A. Razborov and Steven Rudich. Natural Proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

[Sap15] Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015.

[Sch80] Jacob T. Schwartz. Fast Probabilistic Algorithms for Verification of Polynomial Identities. *Journal of the ACM*, 27(4):701–717, 1980.

[Sie14] Carl L Siegel. Über einige anwendungen diophantischer approximationen. In *On Some Applications of Diophantine Approximations*, pages 81–138. Springer, 2014.

[Smo87] Roman Smolensky. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC 1987)*, pages 77–82, 1987.

[Str73] Volker Strassen. Die Berechnungskomplexität Von Elementarsymmetrischen Funktionen Und Von Interpolationskoeffizienten. *Numerische Mathematik*, 20(3):238–251, 1973.

[VSBR83] Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM Journal of Computing*, 12(4):641–644, 1983. Preliminary version in the *6th Internationl Symposium on the Mathematical Foundations of Computer Science (MFCS 1981)*.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation*, volume 72 of *Lecture Notes in Computer Science*, pages 216–226. Springer, 1979.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il