Contemporary Mathematics

Randomization, Sums of Squares, Near-Circuits, and Faster Real Root Counting

Osbert Bastani, Christopher J. Hillar, Dimitar Popov, and J. Maurice Rojas

ABSTRACT. Suppose that f is a real univariate polynomial of degree D with exactly 4 monomial terms. We present a deterministic algorithm of complexity polynomial in $\log D$ that, for most inputs, counts the number of real roots of f. The best previous algorithms have complexity super-linear in D. We also discuss connections to sums of squares and $\mathcal{A}$-discriminants, including explicit obstructions to expressing positive definite sparse polynomials as sums of squares of few sparse polynomials. Our key theoretical tool is the introduction of efficiently computable *chamber cones*, which bound regions in coefficient space where the number of real roots of f can be computed easily. Much of our theory extends to n-variate $(n+3)$-nomials.

1. Introduction

Counting the number of real solutions of polynomial equations in one variable is a fundamental ingredient behind many deeper tasks involving the topology of real algebraic sets. However, the intrinsic complexity of this basic enumerative problem becomes a mystery as soon as one considers the input representation in a refined way. Such complexity questions become important in many applications such as geometric modeling or the discretization of partial differential equations in physics because one often encounters polynomials that have sparse expansions relative to some basis; i.e., the underlying linear combination has few terms relative to its degree. Our goal here is to provide novel exponential speed-ups for counting the real roots of certain sparse univariate polynomials of high degree.

Sturm sequences [**Stu35**], and their later refinements [**Hab48, BPR06**], have long been a central technique for counting real roots of univariate polynomials. In combination with more advanced algebraic tools such as a Gröbner bases or resultants [**GKZ94, BPR06**], Sturm sequences can even be used to algorithmically study the topology of real algebraic sets in arbitrary dimension (e.g., see [**BPR06**, Chapters 2, 5, 11, and 16]). Unfortunately, Sturm sequences quickly become inefficient for sparse polynomials of large degree (see Examples 1.1 and 1.2 below), and we must therefore seek alternative tools.

1991 *Mathematics Subject Classification.* Primary: 14P25, 14Q20; Secondary: 14M25, 14T05, 65Y20.

Key words and phrases. sparse, sum of squares, $\mathcal{A}$-discriminant, real root counting.

Bastani and Popov were partially supported by NSF REU grant DMS-0552610. Hillar was partially supported by an NSF Postdoctoral Fellowship and an NSA Young Investigator grant. Rojas was partially supported by NSF MCS grant DMS-0915245, DOE ASCR grant DE-SC0002505, a Wenner Gren Foundation grant, Sandia National Laboratories, and MSRI.

More recently, the connection between positive polynomials and sums of squares has been exploited to significantly speed up the optimization of certain real polynomials over semi-algebraic domains [**Par03, Las09**]. However, there are also obstructions to using these techniques to speed up computations with sparse polynomials of large degree (see Theorem 1.6 below).

Discriminants have a history nearly as long as that of Sturm sequences and sums of squares, but their algorithmic power has yet to be fully exploited. Our main result is that $\mathcal{A}$-*discriminants* [**GKZ94**] yield a real root counting algorithm with complexity polynomial in the *logarithm* of the degree, for almost all inputs (see Theorem 1.4 below). The use of randomization is potentially inevitable since even detecting real roots becomes **NP**-hard already for moderately sparse multivariate polynomials [**BRS09, PRT09, PRT11**].

1.1. From Large Sturm Sequences to Fast Probabilistic Counting. The classical technique of Sturm Sequences [**Stu35, BPR06**] reduces counting the roots of a polynomial f in a half-open interval $[a,b)$ to a gcd-like computation, followed by sign evaluations for a sequence of polynomials. A key problem in these methods, however, is their apparent super-linear dependence on the degree of the underlying polynomial. The following examples illustrate some of the technical issues we face (see also [**RY05**, Example 1]).

EXAMPLE 1.1. *Setting* $f(x_1) = x_1^{317811} - 2x_1^{196418} + 1$, *the* `realroot` *command in* `Maple 14` *(which is an implementation of Sturm Sequences) results in an out-of-memory error after* 31 *seconds.*[1] *The polynomials in the underlying computation, while quite sparse, have coefficients with hundreds of thousands of digits, thus causing this failure. On the other hand, via more recent work* [**BRS09**]*, one can show that when* $c > 0$ *and* $g(x_1) = x_1^{317811} - cx_1^{196418} + 1$, *the polynomial* g *has exactly* 0, 1, *or* 2 *positive roots according as* c *is less than, equal to, or greater than* $\frac{317811}{(121393^{121393}196418^{196418})^{1/317811}} \approx 1.94$. *In particular, our* f *has exactly* 2 *positive roots. (We discuss how to efficiently compare monomials in rational numbers with rational exponents in Algorithm 2.18 of Section 2.3.)* $\diamond$

EXAMPLE 1.2. *Moving to tetranomials, consider* $f(x_1) = ax_1^{100008} - x_1^{50005} + bx_1^{50004} - 1$ *with* $a,b > 0$. *The polynomial* f *has exactly* 1 *or* 3 *positive roots (via the classical Descartes' Rule of Signs* [**RS02**, Cor. 10.1.10, pg. 319]*), but the inequalities characterizing which* (a,b) *yield either possibility are much more unwieldy than in our last example. Indeed, there are at least* 2 *such inequalities, involving polynomials in* a *and* b *with tens of thousands of terms. In particular, for* $(a,b) = \left(2, \frac{1}{2}\right)$, *Sturm sequences in* `Maple 14` *result in an out-of-memory error after 122 seconds.* $\diamond$

We have discovered that $\mathcal{A}$-discriminants, reviewed in Section 2, resolve these problems and allow us to construct algorithms with complexity that is polynomial in the logarithm of the degree. We make some definitions before stating our result precisely.

DEFINITION 1.3. *For any subset* $S \subseteq \mathbb{C}^d$, *let*

$$\mathrm{Log}|S| := \{(\log|x_1|, \ldots, \log|x_d|) \mid (x_1, \ldots, x_d) \in S\},$$

where the log base is $e \approx 2.718281828$ *and we use the convention that* $\log(0) = -\infty$. *The* stable log-uniform content *on* $\mathbb{R}^d$ *is defined to be (when the limit below exists)*

$$\nu(S) := \lim_{M \to \infty} \frac{\mu(\mathrm{Log}|S| \cap [-M,M]^d)}{(2M)^d},$$

[1] Running on a 16GB RAM Dell PowerEdge SC1435 departmental server with 2 dual-core Opteron 2212HE 2Ghz processors and OpenSUSE 10.3.

where μ denotes Lebesgue measure on $\mathbb{R}^d$. $\diamond$

The stable log-uniform content satisfies all the axioms of a measure (for the algebra of sets where the limit exists) except for countable additivity, although it is finitely additive. What will be important for us here is that any S with $\log|S|$ a polyhedron always has well-defined $\nu(S)$, ν is invariant under reflection across coordinate hyperplanes, and that $\nu(\mathbb{R}^d)=1$.

THEOREM 1.4. *Let $0<a_2<a_3<a_4=D$ be positive integers and set*

$$f(x_1)=c_1+c_2x_1^{a_2}+c_3x_1^{a_3}+c_4x_1^{a_4}.$$

There is a set $S\subseteq\mathbb{R}^4$ of coefficients with stable log-uniform content 1, and a deterministic algorithm with arithmetic complexity polynomial in $\log D$ that computes the exact number of real roots of f given $(c_1,c_2,c_3,c_4)\in S$. Furthermore, if we restrict to $S\cap\mathbb{Z}^4$ and set $\sigma:=\log(2+\max_i|c_i|)$, then this algorithm can be modified to instead require a number of bit operations polynomial in $\sigma+\log D$. The underlying computational models for these two complexity bounds are respectively the BSS model over $\mathbb{R}$ and the Turing model.

Although the regions in coefficient space determining polynomials with a constant number of real roots become more complicated as the number of monomial terms increases, nevertheless one can efficiently characterize large subregions — *chamber cones* — where the number of real roots is very easy to compute (see Section 3). This motivates the introduction of probability and average-case complexity, and the $\mathcal{A}$-discriminant allows one to make this approach completely precise and algorithmic. In fact, our framework enables us to transparently extend Theorem 1.4 to n-variate $(n+3)$-nomials (see Theorem 3.19 of Section 3.3).

REMARK 1.5. *The algorithmic underpinning of Theorem 1.4 consists of Algorithms 3.9 and 3.20, respectively of Sections 3.2 and 3.4. As clarified there, and in Section 3.3, one can also sometimes detect when f lies outside S, in which case a different method to count real roots can be used.* $\diamond$

Our focus on the stable log-uniform content simplifies the development of our approach and is motivated by the construction of floating-point numbers as expressions of the form $a\times 10^b$ where $a\in[1,10)\cap\mathbb{Q}$ and $b\in\mathbb{Z}$. Also, the stable log-uniform content, abstracted to more general complete fields, has already been used in work of Avendaño and Ibrahim to study the expected number of roots of sparse polynomial systems over a broad family of fields including $\mathbb{Q}_p$, $\mathbb{R}((t))$, and $\mathbb{C}((t))$ [**AI11**].

It is natural to ask how the success probability in Theorem 1.4 behaves under other well-known measures such as uniform or Gaussian. Unfortunately, the underlying calculations become much more complicated. We hope to address more classical measures in future work. On a deeper level, it is far from clear what a truly "natural" probability measure on the space of tetranomials is. For instance, for non-sparse polynomials, it is popular to use specially weighted independent Gaussian coefficients since the resulting measure becomes invariant under a natural orthogonal group action (e.g., see [**Kos88, SS96, BSZ00**]). However, we are unaware of any study on the types of distributions occurring for the coefficients of polynomials arising in applications.

The speed-ups we achieve here actually hold in far greater generality: see [**BRS09, PRT09, PRT11**] for the case of n-variate $(n+k)$-nomials with $k\leq 2$, Section 3 for connections to n-variate $(n+3)$-nomials, the forthcoming paper [**AAR11**] for the general univariate case, and the forthcoming paper [**PRRT11**] for chamber cone theory of $n\times n$ sparse polynomial systems. A main goal of this paper is to illustrate and clarify the underlying theory in a non-trivial special case.

As for other approaches to this problem, we remark that most well-known algorithms for real root counting lack speed-ups for sparse polynomials. For example, in the notation of Theorem 1.4, [**LM01**] gives an arithmetic complexity bound of $O(D\log^5 D)$ which, via the techniques of [**BPR06**], produces a bit complexity bound super-linear in $\sigma + D$. No algorithm with complexity polynomial in $\log D$ (deterministic, randomized, or high probability) appears to have been known before for tetranomials. (See [**HTZEKM09**] for recent speed benchmarks of univariate real solvers.)

Also, note that while we focus on speed-ups which replace the polynomial degree D by $\log D$ in this paper, other practical speed-ups that combine semidefinite programming and sparsity are certainly possible (e.g., see [**Las06, KM09**]).

1.2. Sparsity and Univariate Sums of Squares. Recent advances in semidefinite programming (SDP) have produced algorithms for finding sum of squares representations of certain nonnegative polynomials [**Par03**], thus enabling efficient polynomial optimization under certain conditions. When the input is a sparse polynomial, it is natural to ask for sum of squares representations that also respect sparseness. Motivation comes from understanding the efficiency of SDP: should such representations exist in general, one could use SDP to speed up real root counting in the spirit of Theorem 1.4.

It is well-known that a nonnegative univariate polynomial can be written as a sum of two squares, although without any guarantee as to the sparsity of the polynomials being squared (see, e.g., [**Pou71**] for refinements). The following result demonstrates that expressing a sparse positive polynomial as a sparse sum of squares of sparse polynomials is likely not possible in general.

THEOREM 1.6. *There do* not *exist absolute constants ℓ and m with the following property: Any trinomial $f \in \mathbb{R}[x_1]$ that is positive on $\mathbb{R}$ can be written as $f = g_1^2 + \cdots + g_\ell^2$, for some $g_1, \ldots, g_\ell \in \mathbb{R}[x_1]$ with g_i having at most m terms for all i.*

Our second main theorem thus reveals an obstruction to using sums of squares techniques for fast real root counting of sparse polynomials. Softening our concept of sparse sum of squares representation, however, may still enable speed-ups similar to Theorem 1.4 via SDP. For instance, one could ask if a positive trinomial of degree D always admits a representation as a sum of $\log^{O(1)} D$ squares of polynomials with $\log^{O(1)} D$ terms. This question appears to be completely open.

EXAMPLE 1.7. *Elementary calculus shows that*

$$f(x_1) = x_1^{2^k} - 2^k x_1 + 2^k - 1$$

attains a unique minimum value of 0 *at $x = 1$ and thus is nonnegative. It is also easily shown by induction that $f(x_1) = 2^{k-1} \sum_{i=0}^{k-1} \frac{1}{2^i} \left(x_1^{2^i} - 1\right)^2$, which gives an expression for f as a sum of $O(\log D)$ bi*nomials *with $D = 2^k$. Note that from this representation one sees immediately that the only real root of f is $x_1 = 1$.* $\diamond$

The outline of this paper is as follows: The necessary background on amoebae and $\mathcal{A}$-Discriminants is discussed in Section 2, including computational results on linear forms of logarithms. Next, Section 3 explains the algorithm evincing Theorem 1.4, proves its correctness, and calculates its overall time complexity. Finally, in Section 4, we give the proof of Theorem 1.6.

2. Background

2.1. Amoebae and Efficient $\mathcal{A}$-Discriminant Parametrization. We begin by briefly reviewing two important constructions by Gelfand, Kapranov, and Zelevinsky [**GKZ94**].

DEFINITION 2.1. *Given a set of m integer vectors $\mathcal{A} = \{a_1, \ldots, a_m\} \subset \mathbb{Z}^n$, define the following family of (Laurent) polynomials:*

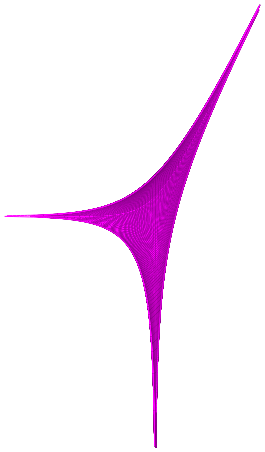
$$\mathcal{F}_{\mathcal{A}} := \{c_1 x^{a_1} + \cdots + c_m x^{a_m} \mid c \in \mathbb{C}^m\},$$

where the notation $x^{a_i} := x_1^{a_{1,i}} \cdots x_n^{a_{n,i}}$ is understood. When $c_i \neq 0$ for all $i \in \{1, \ldots, m\}$, we call $\mathcal{A}$ the support *of $f(x) = \sum_{i=1}^m c_i x^{a_i}$ and we write* $\mathrm{Supp}(f) = \mathcal{A}$. $\diamond$

DEFINITION 2.2. *For a field K, set $K^* := K \setminus \{0\}$. Given any Laurent polynomial $g \in \mathbb{C}\left[x_1^{\pm 1}, \ldots, x_n^{\pm 1}\right]$, its* amoeba *is*

$$\mathrm{Amoeba}(g) := \{\mathrm{Log}|c| \mid c \in (\mathbb{C}^*)^n \text{ and } g(c) = 0\}.$$

Recall that the convex hull *of a set $S \subseteq \mathbb{R}^n$, denoted* Conv$S$*, is the smallest (with respect to containment) convex set containing S. We then define the* (standard) Newton polytope *of g to be* $\mathrm{Newt}(g) := \mathrm{Conv}(\mathrm{Supp}(g))$. $\diamond$



ARCHIMEDEAN AMOEBA THEOREM. (see [**GKZ94**, Cor. 1.8, pg. 196 & Prop. 1.9, pg. 197]) *Given any $g \in \mathbb{C}\left[x_1^{\pm 1}, \ldots, x_n^{\pm 1}\right]$, the complement of* $\mathrm{Amoeba}(g)$ *in $\mathbb{R}^n$ is a finite disjoint union of open convex sets. Also, the vertices of* $\mathrm{Newt}(g)$ *are in bijective correspondence with the unbounded connected components of $\mathbb{R}^n \setminus \mathrm{Amoeba}(g)$.* ■

An example of an amoeba appears above (see also Example 2.7 below). While the complement of the amoeba (in white) appears to have 3 convex connected components, there are in fact 4: the fourth component is a thin sliver emerging further below from the downward pointing tentacle.

DEFINITION 2.3. [**GKZ94**, Chs. 1 & 9–11] *Letting $\mathcal{A} = \{a_1, \ldots, a_m\} \subset \mathbb{Z}^n$ have cardinality m and $f(x) = c_1 x^{a_1} + \cdots + c_m x^{a_m}$, the* $\mathcal{A}$-discriminant variety $\nabla_{\mathcal{A}}$ *is the closure of the set of all points $[c_1 : \cdots : c_m] \in \mathbb{P}^{m-1}_{\mathbb{C}}$ such that*

$$f = \frac{\partial f}{\partial x_1} = \cdots = \frac{\partial f}{\partial x_n} = 0$$

has a solution in $(\mathbb{C}^*)^n$. We also let $\nabla_{\mathcal{A}}(\mathbb{R})$ denote the real part of $\nabla_{\mathcal{A}}$. Finally, when $\nabla_{\mathcal{A}}$ is a hypersurface, the $\mathcal{A}$-discriminant $\Delta_{\mathcal{A}} \in \mathbb{Z}[c_1, \ldots, c_m]$ *is defined to be, up to sign, the irreducible defining polynomial of $\nabla_{\mathcal{A}}$.* $\diamond$

DEFINITION 2.4. *When $\mathcal{A} \subset \mathbb{R}^n$ contains a point a such that $1 + \dim \mathrm{Conv}(\mathcal{A} \setminus \{a\}) = \dim \mathrm{Conv}\mathcal{A}$, we say that* Conv$\mathcal{A}$ *is a* pyramid. *Also, we say that $\mathcal{A}$ is a* near-circuit *when $\mathcal{A}$ has cardinality $n+3$, $\dim \mathrm{Conv}\mathcal{A} = n$, and $\mathcal{A}$ is not a pyramid.* $\diamond$

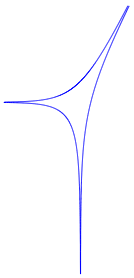
REMARK 2.5. *Our illustrations were drawn via* `Matlab`. *The key program,* `nearckthkplot.m`*, was written by Rojas and is downloadable from*

`www.math.tamu.edu/~rojas/nearcircuits.html`

The program `nearckthkplot.m` *is an implementation of the near-circuit case of the Horn-Kapranov Uniformization (quoted below). Note also that our mention of circuits in this paper alludes to matroid theory, not electrical networks.* $\diamond$

EXAMPLE 2.6. *If $\mathcal{A}=\{0,1,2\}$, then $\mathcal{F}_{\mathcal{A}}$ consists of univariate polynomials of degree ≤ 2 and $\Delta_{\mathcal{A}}$ is the well-known quadratic discriminant $c_2^2-4c_1c_3$. More generally, if $\mathcal{A}\subset\mathbb{Z}^n$ has cardinality $n+2$, $\dim \mathrm{Conv}\mathcal{A}=n$, and $\mathrm{Conv}\mathcal{A}$ is not a pyramid, then $\Delta_{\mathcal{A}}$ is a binomial (see* [**GKZ94**, pp. 217–218 & Prop. 1.8, pg. 274] *or* [**BRS09**, Lemma 2.12]*). This setting, also known as the* circuit *case, is studied from an algorithmic point of view in* [**BRS09, PRT09, PRT11**]. $\diamond$

EXAMPLE 2.7. *When $\mathcal{A}=\{0,404,405,808\}$, the set $\mathcal{F}_{\mathcal{A}}$ consists of polynomials of the form $f(x_1)=c_1+c_2x_1^{404}+c_3x_1^{405}+c_4x_1^{808}$. The underlying $\mathcal{A}$-discriminant is a polynomial in the c_i having* 609 *monomial terms and degree* 1604*. Even though $\Delta_{\mathcal{A}}$ is unwieldy, we can still easily sketch* $\mathrm{Log}|\cdot|$ *of a slice of the real part of its zero set $\nabla_{\mathcal{A}}(\mathbb{R})$ via the* Horn-Kapranov Uniformization *(see its statement below, and the illustration to the right).* $\diamond$



The curve plotted above is the image of the real roots of $\overline{\Delta}_{\mathcal{A}}(c_2,c_4):=\Delta_{\mathcal{A}}(1,c_2,1,c_4)$ under the $\mathrm{Log}|\cdot|$ map; i.e., part of the amoeba of $\overline{\Delta}_{\mathcal{A}}$. Note in particular that the boundary of $\mathrm{Amoeba}(\overline{\Delta}_{\mathcal{A}})$ is contained in the curve above. The connection to amoebae naturally introduces methods from polyhedral and tropical geometry into our setting.

Part of what we accomplish in our paper is to set the stage for fast algorithms that compute the topology of real zero sets of polynomials supported on near-circuits. A key step is understanding the real discriminant complement $\mathbb{P}^{m-1}_{\mathbb{R}}\setminus\nabla_{\mathcal{A}}(\mathbb{R})$.

EXAMPLE 2.8. *Elaborating a folkloric example (see, e.g.,* [**DR06**, Ex. 1.2]*), consider the subset*

$$\mathcal{A}=\left\{\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}1\\0\\1\end{bmatrix}\right\}.$$

One can check that, in addition to satisfying our definition here, $\nabla_{\mathcal{A}}$ coincides with the set of all points $[a_0:a_1:b_0:b_1:c_0:c_1]\in\mathbb{P}^5_{\mathbb{C}}$ such that the homogeneous 3×2 linear system

$$a_0x_0+a_1x_1=b_0x_0+b_1x_1=c_0x_0+c_1x_1=0$$

has a root $[x_0:x_1]\in\mathbb{P}^1_{\mathbb{C}}$. In particular,

$$\nabla_{\mathcal{A}}=\left\{[a_0:a_1:b_0:b_1:c_0:c_1]\;\middle|\;\begin{bmatrix}a_0&a_1\\b_0&b_1\end{bmatrix}=\begin{bmatrix}b_0&b_1\\c_0&c_1\end{bmatrix}=0\right\}$$

and has codimension 2*. Most importantly, the real zero set of any polynomial $f\in\mathcal{F}_{\mathcal{A}}\cap\mathbb{R}[x_1,x_2,x_3]$ is always a connected, doubly ruled quadric surface (possibly a plane) when $f\notin\nabla_{\mathcal{A}}$, and thus the topology of the real zero set of f is constant away from $\nabla_{\mathcal{A}}$.* $\diamond$

When $\nabla_{\mathcal{A}}$ is a hypersurface, the topology of the real zero set of an $f\in\mathcal{F}_{\mathcal{A}}\cap\mathbb{R}[x_1,\ldots,x_n]$ need not be constant away from the discriminant variety (see Section 2.2). Characterizing when $\nabla_{\mathcal{A}}$ has codimension ≥ 2 (for general $\mathcal{A}$) is a subtle problem addressed in [**DS02, DR06, CC07**]. A necessary and sufficient condition for $\mathrm{codim}\nabla_{\mathcal{A}}=1$ when $\mathcal{A}\subset\mathbb{Z}^n$ has cardinality $n+3$ appears in Corollary 3.7 of Section 3. In particular, $\nabla_{\mathcal{A}}$ is always a hypersurface when $\mathcal{A}\subset\mathbb{Z}$ has cardinality 4.

In all but a few restricted settings $\mathcal{A}$-discriminant polynomials are large. For instance, the polynomial $\overline{\Delta}_{\{0,404,405,808\}}$ after Example 2.7 has the following coefficient for $c_2^{808}c_4$:

$$90394708657670090944 84\ldots[2142\text{ digits omitted}]\ldots 08170311749217550336.$$

Fortunately, the following theorem describes an efficient parametrization of $\nabla_{\mathcal{A}}$.

THE HORN-KAPRANOV UNIFORMIZATION. *(See* [**Kap91**], [**PT05**], *and* [**DFS07**, Prop. 4.1]*.) Given $\mathcal{A}=\{a_1,\ldots,a_m\}\subset\mathbb{Z}^n$ with $\nabla_{\mathcal{A}}$ a hypersurface, the discriminant locus $\nabla_{\mathcal{A}}$ is the closure of*

$$\left\{[u_1\lambda^{a_1}:\cdots:u_m\lambda^{a_m}]\;\middle|\;u\in\mathbb{C}^m,\ \mathcal{A}u=\mathbf{O},\ \sum_{i=1}^m u_i=0,\ \lambda\in(\mathbb{C}^*)^n\right\}.\qquad\blacksquare$$

Thus, the null-space of a particular $(n+1)\times m$ matrix provides a parametrization of $\nabla_{\mathcal{A}}$.

Recall that for any two subsets $U,V\subseteq\mathbb{R}^N$, their *Minkowski sum* $U+V$ is the set $\{u+v\mid u\in U,\ v\in V\}$. Also, for any matrix M, we let $M^\top$ denote its transpose.

COROLLARY 2.9. *With the notation above, let $\hat{\mathcal{A}}$ denote the $(n+1)\times m$ matrix whose $i^{\underline{th}}$ column has coordinates corresponding to $1\times a_i$, and let $B\in\mathbb{R}^{m\times p}$ be any real matrix whose columns are a basis for the right null-space of $\hat{\mathcal{A}}$. Also, define $\varphi:\mathbb{C}^p\longrightarrow\mathbb{R}^m$ via $\varphi(t):=\log\left|tB^\top\right|$. Then* Amoeba$(\Delta_{\mathcal{A}})$ *is the Minkowski sum of the row space of $\hat{\mathcal{A}}$ and $\varphi(\mathbb{C}^p)$.* $\blacksquare$

For those familiar with elimination theory, it is evident from the Horn-Kapranov Uniformization that discriminant amoebae are subspace bundles over a lower-dimensional amoeba. This is a geometric reformulation of the homogeneities satisfied by the polynomial $\Delta_{\mathcal{A}}$.

EXAMPLE 2.10. *Continuing Example 2.7, the matrix $\hat{\mathcal{A}}=\begin{bmatrix}1&1&1&1\\0&404&405&808\end{bmatrix}$ has right null-space generated by $(1,-405,404,0)^\top$ and $(1,-2,0,1)^\top$. Thus, from the Horn-Kapranov Uniformization, the set $\nabla_{\mathcal{A}}$ is the closure of the rational surface*

$$\left\{\left[t_1+t_2:(-405t_1-2t_2)\lambda^{404}:404t_1\lambda^{405}:t_2\lambda^{808}\right]\;\middle|\;t_1,t_2\in\mathbb{C},\ \lambda\in\mathbb{C}^*\right\}\subset\mathbb{P}^3_{\mathbb{C}}.$$

Note that f and $\frac{1}{c_1}f$ have the same roots and that $u\mapsto u^{1/405}$ is a well-defined bijection on $\mathbb{R}$ that preserves sign. Note also that the roots of f and $\overline{f}(y):=\frac{1}{c_1}f\left(\left(\frac{c_1}{c_3}\right)^{1/405}y\right)$ differ only by a real scaling when f has real coefficients, and that $\overline{f}$ can be written $1+c_2'y^{404}+y^{405}+c_4'y^{808}$. It follows that the study of $\nabla_{\mathcal{A}}(\mathbb{R})$ reduces to a lower-dimensional slice; the intersection of $\nabla_{\mathcal{A}}$ with the plane defined by $c_1=c_3=1$ is the parametrized curve in $\mathbb{C}^2$:

$$\overline{\nabla}_{\mathcal{A}}=\left\{\left(\frac{-405t_1-2t_2}{t_1+t_2}\left(\frac{404t_1}{t_1+t_2}\right)^{-404/405},\ \frac{t_2}{t_1+t_2}\left(\frac{404t_1}{t_1+t_2}\right)^{-808/405}\right)\;\middle|\;t_1,t_2\in\mathbb{C}\right\}.$$

In other words, $\overline{\nabla}_{\mathcal{A}}$ is the closure of the set of all $(c_2',c_4')\in(\mathbb{C}^*)^2$ such that $1+c_2'x^{404}+x^{405}+c_4'x^{808}$ has a degenerate root in $\mathbb{C}^*$.

Our preceding illustration of the image of $\overline{\nabla}_{\mathcal{A}}(\mathbb{R})$ within Amoeba$(\overline{\Delta}_{\mathcal{A}})$ *(after taking log absolute values of coordinates) thus has the explicit parametrization with $[t_1:t_2]\in\mathbb{P}^1_{\mathbb{R}}\setminus\{[1:0],[0:1],[-2:405],[1:-1]\}$:*

$$\left(\log|405t_1+2t_2|-\frac{1}{405}\log|t_1+t_2|-\frac{404}{405}\log|404t_1|,\log|t_2|+\frac{403}{405}\log|t_1+t_2|-\frac{808}{405}\log|404t_1|\right).$$

In particular, the image of $\mathbb{P}^1_{\mathbb{R}}$ *under this parametrization is the curve from Example 2.7, and it contains all non-isolated points of the boundary of* $\text{Amoeba}(\overline{\Delta}_{\mathcal{A}})$. *See* [**DRRS07**, Lemma 3.3] *and the illustration before that paper's appendix for an example where* $\overline{\nabla}_{\mathcal{A}}(\mathbb{R})$ *contains isolated points (lying in the interior of* $\text{Amoeba}(\overline{\Delta}_{\mathcal{A}})$*).* $\diamond$

A geometric fact about amoebae that will prove useful is the following elegant quantitative result of Passare and Rullgård.

PASSARE-RULLGÅRD THEOREM. [**PR04**, Cor. 1] *Suppose* $g \in \mathbb{C}\left[x_1^{\pm 1}, x_2^{\pm 1}\right]$ *has Newton polygon* P. *Then* $\text{Area}(\text{Amoeba}(g)) \leq \pi^2 \text{Area}(P)$. ■

2.2. Discriminant Chambers and Cones. $\mathcal{A}$-discriminants are central in real root counting because the real part of $\nabla_{\mathcal{A}}$ determines where in coefficient space the real zero set of a polynomial changes topology. Recall that a *(convex) cone* in $\mathbb{R}^m$ is any subset closed under nonnegative linear combinations. (All cones throughout this paper are convex.) Recall also that a *flat* in $\mathbb{R}^n$ is a translated subspace. The dimension of a cone C is then the dimension of the smallest flat containing C.

DEFINITION 2.11. *Suppose* $\mathcal{A} = \{a_1, \ldots, a_m\} \subset \mathbb{Z}^n$ *and* $\nabla_{\mathcal{A}}$ *is a hypersurface. Any connected component* $\mathcal{C}$ *of the complement of* $\nabla_{\mathcal{A}}$ *in* $\mathbb{P}^{m-1}_{\mathbb{R}} \setminus \{c_1 \cdots c_m = 0\}$ *is called a* (real) discriminant chamber. *Let* $\hat{\mathcal{A}}$ *denote the* $(n+1) \times m$ *matrix whose* $i^{\underline{th}}$ *column has coordinates* $1 \times a_i$, *and let* $B = [b_{i,j}] \in \mathbb{R}^{m \times p}$ *be any real matrix with* $\begin{bmatrix} \hat{\mathcal{A}} \\ B^\top \end{bmatrix}$ *invertible. If* $\log|\mathcal{C}|B$ *contains an* m*-dimensional cone, we call* $\mathcal{C}$ *an* outer *chamber (of* $\nabla_{\mathcal{A}}$*). All other chambers of* $\nabla_{\mathcal{A}}$ *are called* inner *chambers (of* $\nabla_{\mathcal{A}}$*). Finally, the formal expression*

$$(c_1, \ldots, c_m)^B := \left(c_1^{b_{1,1}} \cdots c_m^{b_{m,1}}, \ldots, c_1^{b_{1,p}} \cdots c_m^{b_{m,p}}\right)$$

is called a monomial change of variables, *and we say that images of the form* $\mathcal{C}^B$ *(with* $\mathcal{C}$ *an inner or outer chamber) are* reduced *chambers.* $\diamond$

It is easily verified that $\log|\mathcal{C}^B| = \log|\mathcal{C}|B$, where the second expression simply means the image of $\log|\mathcal{C}|$ under right multiplication by the matrix B.

EXAMPLE 2.12. *The illustration from Example 2.7 shows* $\mathbb{R}^2$ *partitioned into what appear to be* 3 *convex and unbounded regions, and* 1 *non-convex unbounded region. There are in fact* 4 *convex and unbounded regions with the fourth visible only if the downward pointing spike were allowed to extend much farther down (see Example 3.2). Thus,* $\mathcal{A} = \{0, 404, 405, 808\}$ *results in exactly* 4 *reduced outer chambers.* $\diamond$

Note that exponentiation by a matrix B gives a well-defined multiplicative homomorphism from $(\mathbb{R}^*)^m$ to $(\mathbb{R}^*)^p$ when B has rational entries with all denominators odd. In particular, thanks to the Archimedean Amoeba Theorem, the definition of outer chamber is independent of B since (for the B above) $\text{Log}\left|\mathcal{C}^B\right|$ is unbounded and convex iff $\text{Log}\left|\mathcal{C}^{B^*}\right|$ is unbounded and convex, where B^* is any matrix whose columns are a basis for the orthogonal complement of the row space of $\hat{\mathcal{A}}$.

One can reduce the study of the topology of the real zero set of a sparse polynomial to that of a representative in a reduced discriminant chamber. A special case of this reduction is contained in the following result.

LEMMA 2.13. [**DRRS07**, Prop. 2.17]. *Suppose that* $\mathcal{A} \subset \mathbb{Z}^n$ *is a near circuit,* $\mathcal{A} \cap Q$ *has cardinality* n *for all facets* Q *of* $\text{Conv}\mathcal{A}$, *all the entries of* $B \in \mathbb{Q}^{(n+3)\times 2}$ *have odd*

denominator, and $\begin{bmatrix} \hat{\mathcal{A}} \\ B^\top \end{bmatrix}$ *is invertible. Also let* $f, g \in \mathcal{F}_{\mathcal{A}} \setminus \nabla_{\mathcal{A}}$ *have respective real coefficient vectors* c *and* c' *with* c^B *and* c'^B *lying in the same reduced discriminant chamber. Then all the complex roots of* f *and* g *are non-singular, and the respective zero sets of* f *and* g *in* $(\mathbb{R}^*)^n$ *are diffeotopic. In particular, when* $n=1$, f *and* g *have the same number of positive roots.* ■

2.3. Integer Linear Algebra and Linear Forms in Logarithms. We now review the quantitative results on integer matrix factorizations and linear forms in logarithms which are crucial for proving our main algorithmic results. Recall that any $n \times m$ matrix $[u_{i,j}]$ with $u_{i,j} = 0$ for all $i > j$ is called *upper triangular*.

DEFINITION 2.14. *Let* $\mathbf{GL}_n(\mathbb{Z})$ *denote the set of all matrices in* $\mathbb{Z}^{n\times n}$ *with determinant* ± 1 *(the set of* unimodular *matrices). Given any* $M \in \mathbb{Z}^{n\times m}$*, an identity of the form* $UM = H$*, with* $H = [h_{i,j}] \in \mathbb{Z}^{n\times m}$ *upper triangular and* $U \in \mathbf{GL}_n(\mathbb{Z})$ *is called a* Hermite factorization *of* M*. In addition, if the following conditions are met:*

(1) *the left-most nonzero entry in each row of* H *is positive,*
(2) *if* $h_{i,j}$ *is the left-most nonzero entry of row* i*, then* $0 \le h_{i',j} < h_{i,j}$ *for all* $i' < i$*,*

then we call H *the* Hermite normal form *of* M*.* ⋄

PROPOSITION 2.15. *Let* K *be any field. We have* $x^{AB} = (x^A)^B$ *for any* $A, B \in \mathbb{Z}^{n\times n}$ *and* $x \in (K^*)^n$*. Moreover, when* $U \in \mathbb{Z}^{n\times n}$ *is unimodular, the map defined by* $m(x) := x^U$ *is an automorphism of* $(K^*)^n$*.* ■

THEOREM 2.16. [**Sto00**, Ch. 6, Table 6.2, pg. 94]. *Given any* $A = [a_{i,j}] \in \mathbb{Z}^{n\times m}$ *with* $m \ge n$*, a Hermite factorization of* A *can be computed within*

$$O\left(nm^{2.376} \log^2(m \max_{i,j} |a_{i,j}|) \right)$$

bit operations. Furthermore, the entries of all matrices involved in the Hermite factorization have bit-size $O(m \log(m \max_{i,j} |a_{i,j}|))$*.* ■

The following result is a very special case of work of Nesterenko that dramatically refines Baker's famous theorem on linear forms in logarithms [**Bak77**].

THEOREM 2.17. [**Nes03**, Thm. 2.1, Pg. 55]. *Given integers* $\gamma_1, \ldots, \gamma_N$ *and* $\alpha_1, \ldots, \alpha_N$ *with* $\alpha_i \ge 2$ *for all* i*, define*

$$\Lambda(\gamma, \alpha) := \gamma_1 \log \alpha_1 + \cdots + \gamma_N \log \alpha_N.$$

If $\Lambda(\gamma, \alpha) \neq 0$*, then the following bound holds:*

$$\log \left| \frac{1}{\Lambda(\gamma, \alpha)} \right| \le 2.9(N+2)^{9/2}(2e)^{2N+6}(2 + \log \max_j |\gamma_j|) \prod_{j=1}^{N} \log \alpha_j. \qquad \blacksquare$$

An obvious consequence of lower bounds for linear forms in logarithms is an efficient method to determine the signs of monomials in integers.

ALGORITHM 2.18.
Input: *Positive integers* $\alpha_1, u_1, \ldots, \alpha_M, u_M$ *and* $\beta_1, v_1, \ldots, \beta_N, v_N$ *with* $\alpha_i, \beta_i \ge 2$ *for all* i*.*
Output: *The sign of* $\alpha_1^{u_1} \cdots \alpha_M^{u_M} - \beta_1^{v_1} \cdots \beta_N^{v_N}$*.*
Description:

(0) *Check via gcd-free bases (see, e.g.,* [**BS96**, Sec. 8.4]*) whether* $\alpha_1^{u_1} \cdots \alpha_M^{u_M} = \beta_1^{v_1} \cdots \beta_N^{v_N}$*. If so, output "*`They are equal.`*" and* `STOP`*.*

(1) *Let* $\gamma = \max\{u_1, \ldots, u_M, v_1, \ldots, v_N\}$ *and set*

$$\delta = \frac{2.9}{\log 2}(2e)^{2M+2N+6}(1+\log\gamma) \times \left(\prod_{i=1}^{M} \log\alpha_i\right)\left(\prod_{i=1}^{N} \log\beta_i\right).$$

(2) *For all* $i \in [M]$ *(resp.* $i \in [N]$*), let* A_i *(resp.* B_i*) be a rational number agreeing with* $\log\alpha_i$ *(resp.* $\log\beta_i$*) in its first* $2+\delta+\log_2 M$ *(resp.* $2+\delta+\log_2 N$*) leading bits.*[2]

(3) *Output the sign of* $\sum_{i=1}^{M} u_i A_i - \sum_{i=1}^{N} v_i B_i$ *and* STOP.

LEMMA 2.19. *Algorithm 2.18 is correct and terminates within a number of bit operations asymptotically linear in*

$$(M+N)(30)^{M+N} L(\log\gamma)\left(\prod_{i=1}^{M} L(\log\alpha_i)\right)\left(\prod_{i=1}^{N} L(\log\beta_i)\right),$$

where $L(x) := x(\log x)^2 \log\log x$. ■

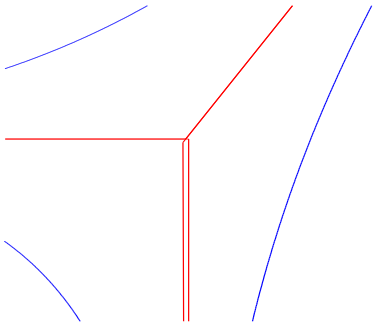
Lemma 2.19 follows directly from Theorem 2.17, the well-known fast iterations for approximating log (see [**Bre76, Sal76, Ber03**]), and the known refined bit complexity estimates for fast multiplication (see, e.g., [**BS96**, Table 3.1, pg. 43]).

3. Chamber Cones and Polyhedral Models

3.1. Defining and Describing Chamber Cones.

DEFINITION 3.1. *Suppose that* $X \subset \mathbb{R}^m$ *is convex and* $Q \supseteq X$ *is the polyhedral cone consisting of all* $c \in \mathbb{R}^m$ *with* $c + X \subseteq X$*. We call* Q *the* recession cone *for* X *and, if* $p \in \mathbb{R}^m$ *satisfies (1)* $p+Q \supseteq X$ *and (2)* $p+c+Q \not\supseteq X$ *for any* $c \in Q \setminus \{\mathbf{O}\}$*, then we call* $p+Q$ *the* placed *recession cone. In particular, the placed recession cone for* $\mathrm{Log}|\mathcal{C}|$ *with* $\mathcal{C}$ *an outer chamber (resp. reduced outer chamber) is called a* chamber cone *(resp.* reduced chamber cone*) of* $\nabla_{\mathcal{A}}$*. We call the facets of the (reduced) chamber cones of* $\nabla_{\mathcal{A}}$ (reduced) walls of $\nabla_{\mathcal{A}}$*. We also refer to walls of dimension* 1 *as* rays. ◇

EXAMPLE 3.2. *Returning to Example 2.7, we draw the rays that are the boundaries of the* 4 *reduced chamber cones. The fourth (slender) reduced chamber cone is now visually exposed. (The magnified illustration to the right actually shows two close and nearly parallel rays going downward.) Note also that reduced chamber cones need not share vertices.* ◇



Chamber cones are well-defined since chambers are *log-convex*, being the domains of convergence of a particular class of hypergeometric series (see, e.g., [**GKZ94**, Ch. 6]). A useful corollary of the Horn-Kapranov Uniformization is a surprisingly simple and explicit description of chamber cones.

[2]For definiteness, we use Arithmetic-Geometric Mean Iteration as in [**Ber03**] to find these approximations. (See also [**Bre76, Sal76**].) In speaking of leading bits, we assume our rational numbers are written in base 2; e.g., 1011.11010011.

DEFINITION 3.3. *Suppose $\mathcal{A}\subset\mathbb{Z}^n$ is a near-circuit. Also let B be any real $(n+3)\times 2$ matrix whose columns are a basis for the right null space of $\hat{\mathcal{A}}$, and let $\beta_1,\ldots,\beta_{n+3}$ be the rows of B. Any set of indices $\mathcal{I}\subset\{1,\ldots,n+3\}$ satisfying the two conditions:*

(a) $[\beta_i]_{i\in\mathcal{I}}$ is a maximal rank 1 submatrix of B,

(b) $\sum_{i\in\mathcal{I}}\beta_i$ is not *the zero vector,*

is called a radiant subset corresponding to $\mathcal{A}$. $\diamond$

THEOREM 3.4. *Suppose that $\mathcal{A}\subset\mathbb{Z}^n$ is a near-circuit and $\nabla_{\mathcal{A}}$ is a hypersurface. Also let B be any real $(n+3)\times 2$ matrix whose columns are a basis for the right null space of $\hat{\mathcal{A}}$, and let $\beta_1,\ldots,\beta_{n+3}$ be the rows of B. Finally, let $S=[s_{i,j}]=B\begin{bmatrix}0&-1\\1&0\end{bmatrix}B^{\top}$, and let s_i denote the row vector whose $j^{\underline{th}}$ coordinate is 0 or $\log|s_{i,j}|$ according as $s_{i,j}$ is 0 or not. Then each wall of $\nabla_{\mathcal{A}}$ is the Minkowski sum of the row-space of $\hat{\mathcal{A}}$ and a ray of the form $s_i-\mathbb{R}_+\sum_{j\in\mathcal{I}}e_j$ for some unique radiant subset $\mathcal{I}$ of $\mathcal{A}$ and any $i\in\mathcal{I}$. In particular, the number of walls of $\nabla_{\mathcal{A}}$, the number of chamber cones of $\nabla_{\mathcal{A}}$, and the number of radiant subsets corresponding to $\mathcal{A}$ are all identical, and lies in $\{3,\ldots,n+3\}$.*

Note that the definition of a radiant subset corresponding to $\mathcal{A}$ is independent of the chosen basis B, since the definition is invariant under column operations on B.

REMARK 3.5. *Theorem 3.4 refines an earlier result of Dickenstein, Feichtner, and Sturmfels* [**DFS07**, Thm. 1.2] *where* unshifted *variants of chamber cones (all going through the origin) were computed for non-pyramidal $\mathcal{A}\subset\mathbb{Z}^n$ with arbitrary cardinality and $\nabla_{\mathcal{A}}$ a hypersurface. A version of Theorem 3.4 for general $\mathcal{A}$ will appear in* [**PRRT11**]. $\diamond$

EXAMPLE 3.6. *It is easy to show that a generic $\mathcal{A}$ satisfying the hypotheses of Theorem 3.4 will have exactly $n+3$ chamber cones, as in Example 2.7. It is also almost as easy to construct examples having fewer chamber cones. For instance, taking*

$$\mathcal{A}=\left\{\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}2\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix}\right\}\ \textit{and}\ \ B=\begin{bmatrix}-1&2&-2&0&1\\0&1&-2&1&0\end{bmatrix}^{\top},$$

we see that the hypotheses of Theorem 3.4 are satisfied and that $\{1,5\}$ is a non-*radiant subset. Thus, the underlying discriminant variety $\nabla_{\mathcal{A}}$ has only 3 chamber cones.* $\diamond$

Proof of Theorem 3.4: First note that by Corollary 2.9, the set $\text{Amoeba}(\Delta_{\mathcal{A}})$ is the Minkowski sum of $\varphi(\mathbb{C}^2)$ and the row space of $\hat{\mathcal{A}}$, where $\varphi(t)=\text{Log}\left|tB^{\top}\right|$. Determining the walls therefore reduces to determining the directions orthogonal to the row space of $\hat{\mathcal{A}}$ in which $\varphi(t)$ becomes unbounded.

Since $\mathbb{1}:=(1,\ldots,1)$ is in the row space of $\hat{\mathcal{A}}$, we have $\mathbb{1}B=\mathbf{O}$ and thus $\varphi(t)=\varphi(t/M)$ for all $M>0$. Thus, we can restrict to the compact subset $\{(t_1,t_2)\mid|t_1|^2+|t_2|^2=1\}$, and we observe that $\varphi(t)$ becomes unbounded iff $t\beta_i^{\top}$ goes to zero for some i. In particular, there are no more than $n+3$ reduced walls. Note also that $t\beta_i^{\top}\to 0$ iff t tends to a suitable (nonzero) multiple of $\beta_i\begin{bmatrix}0&-1\\1&0\end{bmatrix}$, in which case those coordinates of $\varphi(t)$ which become unbounded are precisely those with index $j\in\mathcal{I}$, in which $\mathcal{I}$ is the unique radiant subset corresponding to those rows of B that are nonzero multiples of β_i. (The assumption that $\mathcal{A}$ not be a pyramid implies that B can have no zero rows.) Furthermore, the coordinates of $\varphi(t)$ that become unbounded each tend to $-\infty$. Note that radiance condition (b) comes into play since we are looking for directions *orthogonal to the row-space of $\hat{\mathcal{A}}$* for which $\varphi(t)$ becomes unbounded.

It follows that each wall is of the asserted form. However, we still need to account for the coordinates of $\varphi(t)$ that remain bounded. If t tends to a suitable (nonzero) multiple of $\beta_i \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then it is clear that any coordinate of $\varphi(t)$ with index $j \notin \mathcal{I}$ tends to $s_{i,j}$ (modulo a multiple of $\mathbb{1}$ added to $\varphi(t)$). Thus, we have provided a bijection between radiant subsets corresponding to $\mathcal{A}$ and the walls of $\nabla_{\mathcal{A}}$.

To conclude, note that the row space of $\hat{\mathcal{A}}$ has dimension $n+1$ by construction, so the walls are all actually (parallel) n-plane bundles over rays. By the Archimedean Amoeba Theorem, each outer chamber of $\nabla_{\mathcal{A}}$ must be bounded by 2 walls, and the walls have a natural cyclic ordering. It follows that the number of chamber cones is the same as the number of rays. The upper bound of $n+3$ on the number of rays is thus clear. To see the lower bound of 3, first note that having one or two radiant subsets is impossible: this is because $\mathbb{1}B = \mathbf{O}$. Since $\nabla_{\mathcal{A}}$ is a hypersurface, the Horn-Kapranov Uniformization implies that there must be at least one radiant subset, there must therefore be at least 3, so we are done. ■

A simple consequence of our proof, combined with an earlier observation of Dickenstein and Sturmfels [**DS02**, Cor. 4.5], is the following characterization of near-circuits yielding $\mathcal{A}$-discriminants that are hypersurfaces.

COROLLARY 3.7. *Suppose $\mathcal{A}$ is a near-circuit. Then $\nabla_{\mathcal{A}}$ is a hypersurface iff $\mathcal{A}$ has a radiant subset. In particular, if $\mathcal{A}$ has a radiant subset then it has at least* 3 *radiant subsets.* ■

Note in particular that when $\mathcal{A} \subset \mathbb{Z}$ has cardinality 4, $\nabla_{\mathcal{A}}$ is always a hypersurface: it is easy to show that the right null-space of such an $\mathcal{A}$ always has at least 2 linearly independent rows, thus implying at least 2 (and thus at least 3) radiant subsets.

3.2. Which Chamber Cone Contains Your Problem? An important consequence of Theorem 3.4 is that while the underlying $\mathcal{A}$-discriminant polynomial $\Delta_{\mathcal{A}}$ may have huge coefficients, the *rays* of a linear projection of $\text{Amoeba}(\Delta_{\mathcal{A}})$ admit a concise description involving few bits, save for the transcendental coordinates coming from the "shifts" s_i. Applying our quantitative estimates from Section 2.3, we can then quickly find which chamber cone contains a given n-variate $(n+3)$-nomial.

THEOREM 3.8. *With the notation of Theorem 3.4, suppose that $f \in \mathcal{F}_{\mathcal{A}} \cap \mathbb{R}[x_1, \ldots, x_n]$, and let τ denote the maximum bit-size of any coordinate of $\mathcal{A}$. Then we can determine the unique chamber cone containing f — or correctly decide if f is contained in* 2 *or more chamber cones — within a number of arithmetic operations that is polynomial in $n+\tau$. Furthermore, if $f \in \mathcal{F}_{\mathcal{A}} \cap \mathbb{Z}[x_1, \ldots, x_n]$, σ is the maximum bit-size of any coefficient of f, and n is fixed, we can also obtain a* bit *complexity bound polynomial in $\tau+\sigma$.* ■

Theorem 3.8 is the central tool behind our complexity results and follows from the correctness of (and giving suitable complexity bounds for) the following algorithm:

ALGORITHM 3.9.

Input: *A near-circuit $\mathcal{A} \subset \mathbb{Z}^n$ of cardinality $n+3$ and the coefficient vector c of a polynomial $f \in \mathcal{F}_{\mathcal{A}} \cap \mathbb{R}[x_1, \ldots, x_n]$.*

Output: *Radiant subsets $\mathcal{I}$ and $\mathcal{I}'$ (corresponding to $\mathcal{A}$) generating the walls of the unique chamber cone containing f, or a true declaration that f is contained in at least* 2 *chamber cones, or a true declaration that $\nabla_{\mathcal{A}}$ is not a hypersurface.*

Description:

(-5) *(Preprocessing) Compute the Hermite Factorization* $H^\top = U^\top \hat{\mathcal{A}}$ *and let* B *be the submatrix defined by the rightmost* 2 *columns of* U.

(-4) *(Preprocessing) Let* $\beta_1, \ldots, \beta_{n+3}$ *be the rows of* B, *set* $S = [s_{i,j}] = B \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} B^\top$, *and let* s_i *denote the row vector whose* $j^{\underline{th}}$ *coordinate is* 0 *or* $\log|s_{i,j}|$ *according as* $s_{i,j}$ *is* 0 *or not.*

(-3) *(Preprocessing) Find all radiant subsets* $\mathcal{I} \subset \{1, \ldots, n+3\}$ *corresponding to* $\mathcal{A}$. *If there are none, then output* "$\nabla_{\mathcal{A}}$ `has codimension` $\geq$`2!`" *and* `STOP`.

(-2) *(Preprocessing) For any radiant subset* $\mathcal{I}$ *set* $\beta'_{\mathcal{I}} = -\sum_{j \in \mathcal{I}} \beta_j$ *and let* $s_{\mathcal{I}}$ *denote the row vector* $s_i B$ *for any fixed* $i \in \mathcal{I}$.

(-1) *(Preprocessing) Sort the* $\beta'_{\mathcal{I}}$ *in order of increasing counter-clockwise angle with the x-coordinate ray and let* R *denote the resulting ordered collection of* $\beta'_{\mathcal{I}}$.

(0) *(Preprocessing) For each radiant subset* $\mathcal{I}$, *compute* $v_{\mathcal{I}} \in \mathbb{Q}^2$, *the intersection of the lines* $s_{\mathcal{I}} + \mathbb{R}\beta'_{\mathcal{I}}$ *and* $s_{\mathcal{I}'} + \mathbb{R}\beta'_{\mathcal{I}'}$, *where* $\beta'_{\mathcal{I}'}$ *is the counter-clockwise neighbor of* $\beta'_{\mathcal{I}}$.

(1) *Set ConeCount* $= 0$.

(2) *Via binary search, attempt to find a pair of adjacent rays of the form* $(v_{\mathcal{I}} + \mathbb{R}_+\beta'_{\mathcal{I}}, v_{\mathcal{I}} + \mathbb{R}_+\beta'_{\mathcal{I}'})$ *containing* $\mathrm{Log}|c|B$.

  (a) *If (ConeCount* $= 0$ *and there is no such pair of rays)*
  or
  (ConeCount $= 1$ *and there is such a pair of rays)*
  then output "`Your` f `lies in at least 2 distinct chamber cones.`" *and* `STOP`.

  (b) *If ConeCount* $= 0$ *and there is such a pair of rays, delete* $\beta'_{\mathcal{I}}$ *and* $\beta'_{\mathcal{I}'}$ *from* R, *set ConeCount* $=$ *ConeCount* $+ 1$, *and* `GOTO STEP` *(2).*

(3) *Output* "`Your` f `lies in the unique chamber cone determined by` $\mathcal{I}$ `and` $\mathcal{I}'$`.`" *and* `STOP`.

REMARK 3.10. *An important detail for large scale computation is that the preprocessing steps (-5)–(0) need only be done* once *per support* $\mathcal{A}$. *This can significantly increase efficiency in applications where one has just one (or a few)* $\mathcal{A}$ *and one needs to answer chamber cone membership queries for numerous* f *with the same support.* $\diamond$

Proof of Correctness of Algorithm 3.9: First note that the computed matrix B has columns that form a basis for the right null-space of $\mathcal{A}$. This follows since our assumptions on $\mathcal{A}$ ensure that the rank of $\hat{\mathcal{A}}$ is $n+1$; thus, the last 2 rows of $H^\top$ consist solely of zeroes.

By construction, Theorem 3.4 then implies that the $\beta'_{\mathcal{I}}$ are exactly the reduced rays for $\nabla_{\mathcal{A}}$, modulo an invertible linear map. (The invertible map arises because right-multiplication by B induces an injective projection of the right null-space of $\hat{\mathcal{A}}$ onto $\mathbb{R}^2$.)

It is then clear that the preprocessing steps do nothing more than provide us a B suitable for Theorem 3.4 and a sorted set of reduced rays ready for chamber cone membership queries via binary search, should $\nabla_{\mathcal{A}}$ be a hypersurface. (Corollary 3.7 implies that we correctly detect when $\nabla_{\mathcal{A}}$ is not a hypersurface.) In particular, since the reduced chamber cones cover $\mathbb{R}^2$, the correctness of Steps (1)–(3) is clear and we are done. ■

In what follows, we will use the "soft-Oh" notation $O^*(h)$ to abbreviate bounds of the form $O\left(h(\log h)^{O(1)}\right)$.

Complexity Analysis of Algorithm 3.9: We begin our analysis from the more involved point of view of bit complexity. Our arithmetic complexity bound will then follow quickly from this study.

By Theorem 2.16, Step (-5) takes $O\left(n^{3.376}\tau^2\right)$ bit operations. Also, the resulting bit-size for the entries of B is $O(n\tau)$.

The complexity of Step (-4) is negligible, save for the approximation of certain logarithms. The latter won't come into play until we start deciding on which side of a ray a point lies, so let us analyze the remaining preprocessing steps.

Step (-3) can be accomplished easily by a greedy approach: one iterates through the rows $\beta_2,\ldots,\beta_{n+3}$ to find which ones are multiples of β_1. Once this is finished, one checks whether the resulting set of indices is radiant or not, and then one repeats this process with the remaining rows of B. In summary, we need $O(n^2)$ arithmetic operations on numbers of bit-size $O(n\tau)$, giving a total of $O^*(n^3\tau)$ for the number of bit operations.

Step (-2) has negligible complexity.

The comparisons in Step (-1) can be accomplished by computing the cosine and sine of the necessary angles using dot products and cross products. Via the well-known asymptotically optimal sorting algorithms, it is then clear that Step (-1) requires $O(n\log n)$ arithmetic operations on integers of bit-size $O(n\tau)$, contributing a total of $O^*(n^2\tau\log n)$ bit operations.

Step (0) has negligible complexity.

Thus, the complexity of the Preprocessing Steps (-5)–(0) is $O(n^{3.376}\tau^2)$ bit operations.

Continuing on to Steps (1)–(3), we now see that we are faced with $O(\log n)$ sidedness comparisons between a point and an oriented line. More precisely, we need to evaluate $O(\log n)$ signs of determinants of matrices of the form $\begin{bmatrix}\mathrm{Log}|c|B-s_J\\ \beta'_J\end{bmatrix}$. Each such sign evaluation, thanks to Algorithm 2.18 and Lemma 2.19, takes

$$O\Big(n30^{2n+5}L(\sigma+n\tau)L(\sigma)^{n+3}L(n\tau)^{n+2}\Big)$$

bit operations.

We have thus proved our desired bit complexity bound which, while polynomial in $\tau+\sigma$ for fixed n, is visibly exponential in n. Note, however, that the exponential bottleneck occurs only in the sidedness comparisons of Step (2).

To obtain an improved arithmetic complexity bound, observe that the sidedness comparisons can be replaced by computations of signs of differences of monomials, simply by exponentiating the resulting linear forms in logarithms. Via recursive squaring [**BS96**, Thm. 5.4.1, pg. 103], it is then clear that each such comparison requires only $O(n^2\tau)$ arithmetic operations. Thus, the overall number of arithmetic operations drops to polynomial in $n+\tau$ and we are done. ■

Let us now state some final combinatorial constructions before fully describing how chamber cones apply to real root counting.

3.3. Canonical Viro Diagrams and the Probability of Lying in Outer Chambers. Our use of outer chambers and chamber cones enables us to augment an earlier construction of Viro. Let us first recall that a *triangulation* of a point set $\mathcal{A}$ is a simplicial complex Σ whose vertices lie in $\mathcal{A}$.

DEFINITION 3.11. *We say that a triangulation of $\mathcal{A}$ is* coherent *iff its maximal simplices are exactly the domains of linearity for some function ℓ that is convex, continuous, and piecewise linear on the convex hull of $\mathcal{A}$. In particular, we will sometimes define such an ℓ by fixing the values $\ell(a)$ for just those $a\in\mathcal{A}$ and then employing the faces of*

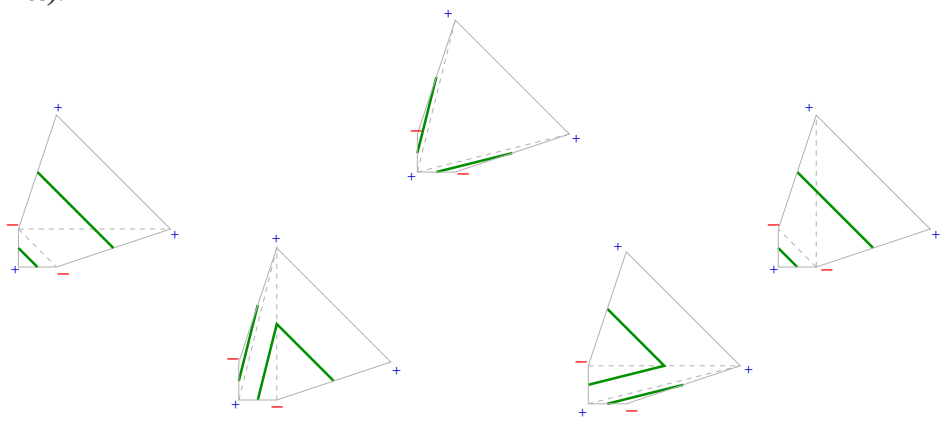
$\mathrm{Conv}(\{(a,\ell(a)) \mid a\in\mathcal{A}\})$ *having inner normal with positive last coordinate. The resulting graph is known as the* lower hull *of the* lifted *point set* $\{(a,\ell(a)) \mid a\in\mathcal{A}\}$. $\diamond$

DEFINITION 3.12. *(See Proposition 5.2 and Theorem 5.6 of* [**GKZ94**, Ch. 5, pp. 378–393]*.) Suppose that* $\mathcal{A}\subset\mathbb{Z}^n$ *is finite and that the convex hull of* $\mathcal{A}$ *has positive volume and boundary* ∂A*. Suppose also that* $\mathcal{A}$ *is equipped with a coherent triangulation* Σ *and a function* $s:\mathcal{A}\to\{\pm\}$ *which we will call a* distribution of signs for $\mathcal{A}$*. Any edge with vertices of opposite sign is called an* alternating edge*, and we define a piece-wise linear manifold — the* Viro diagram $\mathcal{V}_{\mathcal{A}}(\Sigma,s)$ *— in the following local manner: For any n-cell* $C\in\Sigma$*, let* L_C *be the convex hull of the set of all midpoints of alternating edges of* C*, and set*

$$\mathcal{V}_{\mathcal{A}}(\Sigma,s) := \bigcup_{C \text{ an } n\text{-cell}} L_C \setminus \partial A.$$

When $\mathcal{A}=\mathrm{Supp}(f)$ *and* s *is the corresponding sequence of coefficient signs, then we also call* $\mathcal{V}_\Sigma(f):=\mathcal{V}_{\mathcal{A}}(\Sigma,s)$ *the* Viro diagram of f corresponding to Σ. $\diamond$

EXAMPLE 3.13. *Consider* $f(x)=1-x_1-x_2+3x_1^4x_2+3x_1x_2^4$*. Then* $\mathrm{Supp}(f)=\{(0,0),(1,0),(0,1),(1,4),(4,1)\}$ *and its convex hull is a pentagon. There are exactly* 5 *coherent triangulations, giving* 5 *possible Viro diagrams for* f *(drawn in thicker green lines):*



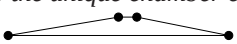
Note that all these diagrams have exactly 2 *connected components, with each component isotopic to an open interval. Note also that* f *is a* 2*-variate* $(2+3)$*-nomial.* $\diamond$

DEFINITION 3.14. *Suppose that* $\mathcal{A}\subset\mathbb{Z}^n$ *is a near-circuit and* $\nabla_{\mathcal{A}}$ *is a hypersurface. Also let* B *be any real* $(n+3)\times 2$ *matrix whose columns are a basis for the right null space of* $\hat{\mathcal{A}}$*. For any* $f\in\mathcal{F}_{\mathcal{A}}$*, define*

$$v(f)=(v_1(f),\ldots,v_{n+3}(f)) := \sum_{i\in\mathcal{I}} e_i + \sum_{j\in\mathcal{J}} e_j,$$

where $\mathcal{I}$ *and* $\mathcal{J}$ *are the unique radiant subsets corresponding to the unique chamber cone containing* $\mathrm{Log}|c|$*. (We set* $v(f):=\mathbf{O}$ *should there not be a unique such chamber cone.) Let* $\widetilde{\mathrm{ArchNewt}}(f)$ *be the convex hull of* $\{(a_i,v_i) \mid i\in\{1,\ldots,n+3\}\}$*, and let* Σ_f *denote the triangulation of* $\mathcal{A}$ *induced by the lower hull of* $\widetilde{\mathrm{ArchNewt}}(f)$*. We call* $\widetilde{\mathrm{ArchNewt}}(f)$ *the* renormalized Archimedean Newton polygon of f*. Also, call any polynomial of the form* $\sum_{a_i\in Q} c_i x^{a_i}$ *— with* Q *a cell of* Σ_f *— a* canonical lower polynomial *for* f*. Finally, we write* $\mathcal{V}(f):=\mathcal{V}_\Sigma(f)$ *for the* canonical Viro diagram *of* f. $\diamond$

Note that for an $f \in \mathbb{C}\left[x_1^{\pm 1}, \ldots, x_n^{\pm 1}\right]$ its standard Newton polytope lies in $\mathbb{R}^n$, while $\widetilde{\text{ArchNewt}}(f)$ lies in $\mathbb{R}^{n+1}$.

EXAMPLE 3.15. *Let* $f(x_1) = 1 - \frac{1}{2}x^{404} + x^{405} - 2x^{808}$, $c = \left(1, -\frac{1}{2}, 1, -2\right)$, *and* $\mathcal{A} = \{0, 404, 405, 808\}$. *A routine calculation reveals that* $\{\{2\}, \{3\}\}$ *is the pair of radiant subsets corresponding to the unique chamber cone containing* $\text{Log}|c|$. *Thus,* $v(f) = (0,1,1,0)$ *and* $\widetilde{\text{ArchNewt}}(f)$ *is*  *(modulo some artistic stretching). In particular,* Σ_f *has the single cell* $[0, 808]$, *which is an alternating cell, and so* $\mathcal{V}(f)$ *consists of a single point. More than coincidentally,* f *has exactly* 1 *positive root.* $\diamond$

EXAMPLE 3.16. *Returning to Example 3.13, let* $c = (1, -1, -1, 3, 3)$. *A routine calculation reveals that the unique chamber cone containing* $\text{Log}|c|$ *is defined by the pair of radiant subsets* $\{\{2\}, \{3\}\}$. *Thus,* $v(f) = (0,1,1,0,0)$ *and* Σ_f *is the upper middle triangulation from the illustration of Example 3.13.* $\mathcal{V}(f)$ *then consists of* 2 *disjoint open intervals and, more than coincidentally, the positive zero set of* f *has exactly* 2 *connected components, each homeomorphic to an open interval.* $\diamond$

THEOREM 3.17. *Following the notation above, set* $\hat{f}_t(x) = \sum_{i=1}^{n+3} c_i t^{v_i(f)} x^{a_i}$ *and assume in addition that* $\mathcal{A} \cap Q$ *has cardinality* n *for all facets* Q *of* $\text{Conv}\mathcal{A}$. *Then* $c = (c_1, \ldots, c_{n+3})$ *lies in an outer chamber* $\Longrightarrow$ *the positive zero sets of* $\hat{f}_t$, *as* t *ranges over* $(0,1]$, *are each diffeotopic to the positive zero set of* $\hat{f}_1$. *In particular,* $\hat{f}_1 = f$ *and thus, when* c *lies in an outer chamber, the positive zero set of* f *is isotopic to* $\mathcal{V}(f)$.

REMARK 3.18. *For* $n = 1$ *we obtain that the number of positive roots of the tetranomial* f *is exactly the cardinality of its canonical Viro diagram.* $\diamond$

Proof: By construction, the image of $\text{Log}\left|\left(c_1 t^{v_1(f)}, \ldots, c_{n+3} t^{v_{n+3}(f)}\right)\right|$ as t ranges over $(0,1]$ is a ray entirely contained in a unique chamber cone. Moreover, by assumption (and since outer chambers are log convex), the ray is also contained entirely in $\text{Log}|\cdot|$ of an outer chamber. The first part of our theorem now follows from Lemma 2.13.

The final part of our theorem is then just a reformulation of Viro's Theorem on the isotopy type of toric deformations of real algebraic sets (see, e.g., [**GKZ94**, Thm. 5.6]). ■

The main contribution of our paper is thus an efficient method to associate a *canonical* Viro diagram to the positive zero set of a given f, so that both C^1 manifolds have the same topology. Such a method appears to be new, although the necessary ingredients have existed in the literature since at least the 1990s. In particular, to the best of our knowledge, all earlier applications of Viro's method designed clever f having the same topology as some specially tailored Viro diagram, thus going in the opposite direction of our construction.

We state up front that our method for finding isotopy type does *not* work for all f. However, our development yields a sufficient condition — outer chamber membership — that holds with high probability under the stable log-uniform content.

THEOREM 3.19. *Suppose that* $\mathcal{A} \subset \mathbb{Z}^n$ *is a near-circuit and* $\nabla_{\mathcal{A}}$ *is a hypersurface. Suppose also that the coefficients of* $f \in \mathcal{F}_{\mathcal{A}} \cap \mathbb{R}[x_1, \ldots, x_n]$ *are independently chosen via the stable log-uniform content over* $\mathbb{R}$. *Then with probability* 1, f *lies in some outer chamber. In particular, if we assume in addition that* $\mathcal{A} \cap Q$ *has cardinality* n *for all facets* Q *of* $\text{Conv}\mathcal{A}$, *the positive zero set of* f *is isotopic to* $\mathcal{V}(f)$ *with probability* 1.

Proof: By Theorem 3.4, $\text{Amoeba}(\Delta_{\mathcal{A}})$ is an n-plane bundle over $\text{Amoeba}\left(\overline{\Delta}_{\mathcal{A}}\right)$, where $\overline{\Delta}_{\mathcal{A}} \in \mathbb{Z}[a,b]$ and $\Delta_{\mathcal{A}}(c_1, \ldots, c_{n+3}) = \gamma(c)\overline{\Delta}_{\mathcal{A}}(\alpha(c), \beta(c))$ for suitable monomials α, β, γ

in the variables c_i. Furthermore, thanks to Corollary 8 of [**PST05**], Amoeba$(\overline{\Delta}_{\mathcal{A}})$ is *solid*; that is, the complement of Amoeba$(\overline{\Delta}_{\mathcal{A}})$ has no bounded convex connected components.

Let c denote the coefficient vector of f. It follows that f lies in an outer chamber if and only if $\mathrm{Log}|c| \notin \mathrm{Amoeba}(\Delta_{\mathcal{A}})$. In particular, by the Passare-Rullgård Theorem the volume of Amoeba$(\Delta_{\mathcal{A}}) \cap C$ in any large centered cube C occupies a vanishingly small fraction of C. This proves the first assertion. The final assertion is an immediate consequence of the first and Theorem 3.17. ■

Theorem 1.4 follows easily from Theorems 3.17 and 3.19. The applications of Theorems 3.17 and 3.19 to computational real topology will be pursued in another paper.

3.4. Proving Theorem 1.4. Consider the following algorithm for counting the positive roots of "most" real univariate tetranomials.

ALGORITHM 3.20.
Input: *A tetranomial $f \in \mathbb{R}[x_1]$ with support $\mathcal{A}$.*
Output: *A number in $\{0,1,2,3\}$ that is exactly the number of positive roots of f whenever f is in an outer chamber of $\nabla_{\mathcal{A}}$.*
Description:

(1) *Via Algorithm 3.9, and any sub-quadratic planar convex hull algorithm (see, e.g.,* [**OSvK00**]*), compute the canonical Viro diagram $\mathcal{V}(f)$.*
(2) *If f did not lie in a unique chamber cone then output* `"Your f does not lie in an outer chamber, please use an alternative method."` *and* `STOP`.
(3) *Output the cardinality of $\mathcal{V}(f)$ and* `STOP`.

Assuming Algorithm 3.20 is correct, we can count the real roots of f by applying Algorithm 3.20 to $f(x_1)$ and $f(-x_1)$. (Whether f vanishes at 0 can trivially be checked in constant time.) Theorem 1.4 thus follows upon proving the correctness of our last algorithm and providing a suitable complexity bound.

Proof of Correctness of Algorithm 3.20: By Theorem 3.17, the number of positive roots of f is exactly the cardinality of $\mathcal{V}(f)$ whenever f is in an outer chamber. ■

Complexity Analysis of Algorithm 3.20: First observe that Algorithm 3.20 gives a correct answer with probability 1 (relative to the stable log-uniform content) by Theorem 3.19. We finish by proving the complexity bound in the statement of the theorem.

Consider first the more refined setting of bit complexity. From our complexity analysis of Algorithm 3.9, it is clear that Step (1) requires at most

$$O\big(\log^2 D\big) + O\big(L(\sigma + \log D)L(\sigma)^4 L(\log D)^3\big)$$

bit operations, modulo the computation of $\mathcal{V}(f)$. The complexity of computing $\mathcal{V}(f)$ is essentially dominated by that of computing the convex hull of 4 points with coordinates of bit-size $O(\log D)$, which is clearly negligible in comparison. The complexity of Steps (2) and (3) is also negligible. Thus, we obtain a final bit complexity bound of $O^*\big((\sigma + \log D)\sigma^4 \log^3 D\big)$.

As for arithmetic complexity, our earlier analysis of Algorithm 3.9 specializes easily to give an upper bound of $O(\log^2 D)$. (The speed-up arises from the ease of checking inequalities involving integral powers of real numbers in the BSS model over $\mathbb{R}$.) ■

REMARK 3.21. *It is important to note that when f lies in a chamber cone but* not *in any outer chamber, Algorithm 3.20 can give a wrong answer. However, thanks to Theorem 3.19, such an occurence has probability* 0 *under the stable log-uniform content.* ⋄

4. Proving Theorem 1.6

Here, we prove the negative result of our paper: sparse positive univariate polynomials cannot always be expressed as sparse sums of sparse squares. This result is an obstruction to using sum of squares techniques for fast root counting. To prepare for the proof, we first set up some notation. Let $\mathbb{N}_0$ denote the set of nonnegative integers, and fix positive integers ℓ and m. Let $P = [p_{i,j}] \in \mathbb{N}_0^{\ell \times m}$ be a matrix of nonnegative integers, ordered as

$$p_{1,1} \geq p_{2,1} \geq \cdots \geq p_{\ell,1} \quad \text{and} \quad p_{i,j} > p_{i,(j+1)},$$

for all $i \in \{1,\ldots,\ell\}$ and $j \in \{1,\ldots,m\}$. Also, let $a_{i,j}$ be indeterminates over the same index set. Consider now the following polynomial:

$$\begin{aligned} S_P(x_1) &:= \sum_{i=1}^{\ell} \left(\sum_{j=1}^{m} a_{i,j} x_1^{p_{i,j}} \right)^2 \in \mathbb{N}_0[x_1][a_{i,j} \mid 1 \leq i \leq \ell \ , \ 1 \leq j \leq m] \\ &= g_{2p_{1,1}}(P) x_1^{2p_{1,1}} + \cdots + g_1(P) x_1 + g_0(P), \end{aligned} \tag{4.1}$$

in which each nonzero $g_i(P)$ is a homogeneous (quadratic) polynomial in $\mathbb{N}_0[a_{i,j} \mid 1 \leq i \leq \ell \ , \ 1 \leq j \leq m]$. Note that there are at most ℓm^2 distinct powers of x_1 occuring in the monomial term expansion of $S_P(x_1)$ and thus at most ℓm^2 of the g_i are nonzero. We will refer to the integer $p_{i,j}$ as the *exponent* corresponding to the *coefficient* $a_{i,j}$.

LEMMA 4.1. *For any fixed $\ell, m \geq 1$, the following set of polynomials is finite:*

$$G_{m,\ell} := \{g_i(P) \mid P \in \mathbb{N}_0^{\ell \times m} \text{ and } i \in \{1,\ldots,2p_{1,1}\}\}.$$

Proof: Note that the coefficient of any g_i is clearly a nonnegative integer bounded above by $2m\ell$ (independent of P). Note also that each g_i involves at most ℓm variables $a_{i,j}$. Since each g_i is quadratic, it has no more than $\ell m(\ell m - 1)/2$ monomial terms. So there are at most $(2m\ell)^{\ell m(\ell m-1)/2}$ distinct polynomials in $G_{m,\ell}$. ■

Suppose now that $f = \sum_{i=0}^{d} f_i x_1^i$ is a sum of ℓ squares, each involving at most m terms. Then, there is a set of exponents P and an assignment $\overline{a}_{i,j} \in \mathbb{R}$ for the coefficients $a_{i,j}$ such that $f = S_P$ identically. Conversely, fixing a set of exponents P, any real point in the variety determined by the equations $g_i = f_i$ gives a representation of f as a sum of ℓ squares, each involving at most m terms.

We will prove Theorem 1.6 using contradiction by showing that a certain infinite family of trinomials cannot all have sparse representations of the form (4.1). For this approach to work, however, we will need to find a single "universal" set of coefficients $\overline{a}_{i,j}$ that represents an infinite number of sums of squares.

LEMMA 4.2. *Let $F \subset \mathbb{R}[x_1]$ be an infinite collection of polynomials which are sums of ℓ squares, each involving at most m terms. Moreover, suppose that the nonzero coefficients of polynomials $f \in F$ come from a finite set C. Then, there is an infinite subset $\{f_1, f_2, \ldots\} \subseteq F$, with corresponding exponent matrices $P_1, P_2, \ldots \in \mathbb{N}_0^{\ell \times m}$, and a single set of real coefficients $\{\overline{a}_{i,j}\}$, such that for all k, the polynomial f_k is obtained from $S_{P_k}(x_1)$ by specializing $a_{i,j} = \overline{a}_{i,j}$ for all i, j.*

Proof: Given $f \in F$, let P_f be an exponent matrix corresponding to the hypothesized sum of squares representation for f. Also, let T be the set of all possible coefficient polynomials g_i occurring in the expansion of S_{P_f} as a polynomial in x_1 for some $f \in F$. The set T is finite, thanks to Lemma 4.1. By assumption, a putative sum of squares expression for an $f \in F$ gives rise to a set of equations of the form $g_i(P_f) = c_{i,f}$, where the g_i are in T and

the $c_{i,f}$ are in C. The set of all such equations is thus finite, and has a non-empty real zero set since every f has a representation as a sum of ℓ squares of univariate m-nomials. Therefore, by the infinite pigeon-hole principle, there is a subset $\{f_k\}_{k\in\mathbb{N}}$ which has the same set of equations governing the coefficients $a_{i,j}$ for all k. Picking any real solution to such a set of equations finishes the proof. ■

To complete the preparation for our proof of Theorem 1.6, let us also recall "little-oh" notation: given any function $h:\mathbb{N}\to\mathbb{R}$, we say that $h(n)=o(n)$ if

$$\lim_{n\to\infty}\frac{h(n)}{n}=0.$$

It is easy to see that the sum of any finite number of such functions is also $o(n)$. Moreover, if $\lim_{n\to\infty}\frac{p(n)}{n}=p$ for some constant p, then $p(n)=np+o(n)$.

Proof of Theorem 1.6: Suppose, to derive a contradiction, that every positive definite trinomial can be written as a sum of ℓ squares, each involving at most m terms. Consider the following infinite sequence of positive definite trinomials:

$$f_k=x_1^{2k}+x_1^{2k-1}+1,\quad k=1,2,\ldots. \tag{4.2}$$

Using Lemma 4.2, we can find a subsequence f_{k_s} with corresponding exponent matrices $P_{k_1},P_{k_2},\ldots\in\mathbb{N}_0^{\ell\times m}$ and a single set of real numbers $\overline{a}=(\overline{a}_{i,j})$ such that $f_{k_s}(x_1)=S_{P_{k_s}}(x_1,\overline{a})$ as polynomials in x_1, for all positive integers s. Let us also pick $\overline{a}$ so that the number of nonzero coordinates is maximal among all such vectors of coefficients. For clarity of exposition, we will not keep updating the subscripting of indices when taking subsequences.

Given an exponent matrix $P_{k_s}\in\mathbb{N}_0^{\ell\times m}$, define a new matrix

$$\tilde{P}_{k_s}=\frac{1}{k_s}P_{k_s}.$$

This corresponds naturally to the transformation $x_1\mapsto x_1^{1/k_s}$ applied to both sides of the equation $f_{k_s}(x_1)=S_{P_{k_s}}(x_1,\overline{a})$. Since $\deg(f_{k_s})=2k_s$, each matrix $\tilde{P}_{k_s}$ has entries in the interval $[0,1]$. By compactness, we may choose a subsequence P_{k_s} such that $\tilde{P}_{k_s}$ converges in the (entry-wise) Euclidean norm to a matrix $\tilde{P}=[\tilde{p}_{i,j}]\in[0,1]^{\ell\times m}$. Henceforth, we restrict to this subsequence. Clearly, we have $\tilde{p}_{11}=1$, and also that some entry of $\tilde{P}$ is 0. It turns out that 0 and 1 are the only possible values for entries of $\tilde{P}$ which need play a role in (4.1).

CLAIM. *We can choose the subsequence $\{f_{k_s}\}_s$ so that if $0<\tilde{p}_{i,j}<1$, the corresponding coefficient $\overline{a}_{i,j}$ is* 0.

To prove the claim, let us suppose temporarily that $\tilde{P}$ contains $r\geq 3$ entries, $\tilde{p}_1,\ldots,\tilde{p}_r$, with $1=\tilde{p}_1>\cdots>\tilde{p}_r=0$. (Otherwise, the claim is vacuously true.) Each power of x_1 occurring after expanding the squared summands in $S_{P_{k_s}}(x_1)$ is of the form

$$k_s\tilde{p}_u+k_s\tilde{p}_v+o(k_s). \tag{4.3}$$

Thus, for all sufficiently large s, the powers of x_1 occurring in expression (4.1) can be partitioned into classes determined by the distinct values of

$$\tilde{p}_u+\tilde{p}_v,\ \ u,v\in\{1,\ldots,r\}.$$

Note that the numbers (4.3) all become strictly smaller (resp. larger) than $2k_s-1$ (resp. 0) as $s\to\infty$ unless $u=v=1$ (resp. $u=v=r$). (This is because $\tilde{p}_2<1$ and $\tilde{p}_{r-1}>0$.) In particular, for

$$w\in\{2k_s+o(k_s),0+o(k_s)\} \tag{4.4}$$

and s large, the polynomials $g_w(P_{k_s}) \in \mathbb{N}_0[a_{i,j}]$ do not involve the indeterminates $a_{i,j}$ coming from exponents of the form $k_s\tilde{p}_u + o(k_s)$ with $u \notin \{1,r\}$. Moreover, each monomial in $g_w(P_{k_s})$ with w not in one of the classes from (4.4) is divisible by at least one $a_{i,j}$ coming from an exponent of the form $k_s\tilde{p}_u + o(k_s)$ with $u \notin \{1,r\}$ since exponents in f_{k_s} cannot have order $k_s(\tilde{p}_1 + \tilde{p}_r) = k_s + o(k_s)$.

Since the only nonzero coefficients of the sequence (4.2) come from the classes of (4.4), it follows that we may replace with 0 all coefficients $\bar{a}_{i,j}$ corresponding to exponents $k_s p_u + o(k_s)$ with $u \notin \{1,r\}$ and still have the equality of polynomials

$$f_{k_s}(x_1) = S_{P_{k_s}}(x_1, \bar{a}).$$

The claim therefore follows from the maximality property of the chosen set of coefficients $\bar{a}_{i,j}$.

To conclude, we now examine the limiting behavior of the expressions from Equality (4.1). From the claim, it follows that when s is large, we need only consider those exponents from the matrices P_{k_s} that are on the order

$$k_s + o(k_s) \text{ and } 0 + o(k_s).$$

So fix s large enough so that all exponents of P_{k_s} that occur with a nonzero coefficient in (4.1) after substituting $(\bar{a}_{i,j})$ for $(a_{i,j})$ are either strictly greater than $\frac{2}{3}k_s$ or strictly less than $\frac{1}{3}k_s$. Let p be the smallest such exponent greater than $\frac{2}{3}k_s$. When the sum from (4.1) is expanded, the term x_1^{2p} will then appear with positive coefficient; i.e., $g_{2p}(P_{k_s})(\bar{a}) > 0$. (This is because $p > \frac{2}{3}k_s$ and thus, by construction, $2p$ can not be the sum of two exponents other than p and p.) Since the only term of f_{k_s} of positive even degree is $x_1^{2k_s}$, we must then have that $p = k_s$. In particular, it is not possible to obtain a nonzero coefficient for $x_1^{2k_s-1}$ in f_{k_s}. This contradiction completes the proof. ■

Acknowledgements

We thank Alicia Dickenstein and Sandra Di Rocco for useful discussions on when $\mathcal{A}$-discriminant varieties have codimension ≥ 2.

References

[AI11] Avendaño, Martín and Ibrahim, Ashraf, *"Multivariate ultrametric root counting,"* this volume, to appear.

[AAR11] Ascher, Kenneth; Avendaño, Martín; and Rojas, J. Maurice, *"Solving sparse polynomials in time logarithmic in the degree,"* in progress.

[BS96] Bach, Eric and Shallit, Jeff, *Algorithmic Number Theory, Vol. I: Efficient Algorithms*, MIT Press, Cambridge, MA, 1996.

[Bak77] Baker, Alan, *"The theory of linear forms in logarithms,"* in Transcendence Theory: Advances and Applications: proceedings of a conference held at the University of Cambridge, Cambridge, Jan.–Feb., 1976, Academic Press, London, 1977.

[BPR06] Basu, Saugata; Pollack, Ricky; and Roy, Marie-Francoise, *Algorithms in Real Algebraic Geometry*, Algorithms and Computation in Mathematics, vol. 10, Springer-Verlag, 2006.

[Ber03] Bernstein, Daniel J., *"Computing logarithm intervals with the arithmetic-geometric mean iterations,"* available from `http://cr.yp.to/papers.html` , 2003.

[BRS09] Bihan, Frederic; Rojas, J. Maurice; and Stella, Casey E., *"Faster real feasibility via circuit discriminants,"* proceedings of ISSAC 2009 (July 28-31, Seoul, Korea), pp. 39–46, ACM Press, 2009.

[BSZ00] Bleher, Pavel; Shiffman, Bernard; and Zelditch, Steve, *"Universality and scaling of correlations between zeros on complex manifolds,"* Invent. Math. 142 (2000), no. 2, pp. 351–395.

[BCSS98] Blum, Lenore; Cucker, Felipe; Shub, Mike; and Smale, Steve, *Complexity and Real Computation*, Springer-Verlag, 1998.

[Bre76] Brent, Richard P., *"Fast Multiple-Precision Evaluation of Elementary Functions,"* Journal of the Association for Computing Machinery, vol. 23, No. 2, April 1976, pp. 242–251.

[CC07] Curran, Raymond and Cattani, Eduardo, *"Restriction of A-discriminants and dual defect toric varieties,"* Journal of Symbolic Computation 42 (2007), pp. 115–135.

[DFS07] Dickenstein, Alicia; Feichtner, Eva Maria; and Sturmfels, Bernd, *"Tropical discriminants,"* J. Amer. Math. Soc., **20** (2007), pp. 1111–1133.

[DRRS07] Dickenstein, Alicia; Rojas, J. Maurice; Rusek, Korben; Shih, Justin, *"Extremal real algebraic geometry and $\mathcal{A}$-discriminants,"* Moscow Mathematical Journal, vol. 7, no. 3, (July–September, 2007), pp. 425–452.

[DS02] Dickenstein, Alicia and Sturmfels, Bernd, *"Elimination Theory in Codimension 2,"* Journal of Symbolic Computation, vol. 34, no. 2, pp. 119–135, 2002.

[DR06] Di Rocco, Sandra, *"Projective duality of toric manifolds and defect polytopes,"* Proc. London Math. Soc. (3), 93 (2006), pp. 85–104.

[GKZ94] Gel'fand, Israel Moseyevitch; Kapranov, Misha M.; and Zelevinsky, Andrei V.; *Discriminants, Resultants and Multidimensional Determinants*, Birkhäuser, Boston, 1994.

[Hab48] Habicht, Walter, *"Eine Verallgemeinerung des Sturmschen WurzelzÃd'hlverfahrens,"* Comment. Math. Helv. **21** (1948), pp. 99–116.

[HTZEKM09] Hemmer, Michael; Tsigaridas, Elias P.; Zafeirakopolous, Zafeirakis; Emiris, Ioannis Z.; Karavelas, Menelaos I.; and Mourrain, Bernard, *"Experimental evaluation and cross-benchmarking of univariate real solvers,"* proceedings of SNC 2009 (Symbolic-Numeric Computation, Kyoto, Japan, August 2–5), pp. 45–54, ACM Press, 2009.

[Kap91] Kapranov, Misha, *"A characterization of A-discriminantal hypersurfaces in terms of the logarithmic Gauss map,"* Mathematische Annalen, 290, 1991, pp. 277–285.

[KM09] Kojima, M. and Muramatsu, M., *"A Note on sparse SOS and SDP relaxations for polynomial optimization problems over symmetric cones,"* Computational Optimization and Applications Vol. 42 (1), pp. 31–41 (2009).

[Kos88] Kostlan, Eric J., *"Complexity theory of numerical linear algebra,"* Journal of Computational and Applied Mathematics Volume 22, Issues 2-3, June 1988, pp. 219–230

[Las06] Lasserre, Jean Bernard, *"Convergent SDP-relaxations in polynomial optimization with sparsity,"* SIAM J. Optim., Vol. 17, No. 3, pp. 822–843, Sept. 2006.

[Las09] Lasserre, Jean-Michel, *"Moments and sums of squares for polynomial optimization and related problems,"* Journal of Global Optimization, vol. 45, no. 1, pp. 39–61, Sept. 2009.

[LLL82] Lenstra, Arjen K.; Lenstra (Jr.), Hendrik W.; and Lovász, László, *"Factoring polynomials with rational coefficients,"* Math. Ann. 261 (1982), no. 4, pp. 515–534.

[LM01] Lickteig, Thomas and Roy, Marie-Francoise, *"Sylvester-Habicht sequences and fast Cauchy index computation,"* J. Symbolic Computation (2001) **31**, pp. 315–341.

[MR04] Malajovich, Gregorio and Rojas, J. Maurice, *"High Probability analysis of the condition number of sparse polynomial systems,"* Theoretical Computer Science, special issue on algebraic and numerical algorithms, Vol. 315, no. 2–3, (May 2004), pp. 525–555.

[Nes03] Nesterenko, Yuri, *"Linear forms in logarithms of rational numbers,"* Diophantine approximation (Cetraro, 2000), pp. 53–106, Lecture Notes in Math., 1819, Springer, Berlin, 2003.

[OSvK00] Overmars, Mark; Schwarzkopf, Otfried; and van Kreveld, Marc, *Computational Geometry: Algorithms and Applications*, Springer Verlag, 2000.

[Par03] Parrilo, Pablo A., *"Semidefinite programming relaxations for semialgebraic problems,"* Algebraic and geometric methods in discrete optimization, Math. Program. 96 (2003), no. 2, Ser. B, pp. 293–320.

[PR04] Passare, Mikael and Rullgård, Hans, *"Amoebas, Monge-Ampàre measures, and triangulations of the Newton polytope,"* Duke Math. J. Vol. 121, No. 3 (2004), pp. 481–507.

[PST05] Passare, Mikael; Sadykov, Timur; and Tsikh, August, *"Singularities of hypergeometric functions in several variables,"* Compositio Math. 141 (2005), pp. 787–810.

[PT05] Passare, Mikael and Tsikh, August, *"Amoebas: their spines and their contours,"* Idempotent mathematics and mathematical physics, Contemp. Math., v. 377, Amer. Math. Soc., Providence, RI, 2005, pp. 275–288.

[PRT09] Pébay, Philippe; Rojas, J. Maurice; Thompson, David C., *"Optimization and $\mathbf{NP}_{\mathbb{R}}$-completeness of certain fewnomials,"* proceedings of SNC 2009 (August 3–5, 2009, Kyoto, Japan), pp. 133–142, ACM Press, 2009.

[PRT11] Pébay, Philippe; Rojas, J. Maurice; Thompson, David C., *"Optimizing n-variate $(n+k)$-nomials for small k,"* Theoretical Computer Science, Symbolic-Numeric Computation 2009 special issue, Vol. 412, No. 16 (1 April 2011), pp. 1457–1469.

[PRRT11] Pébay, Philippe; Rojas, J. Maurice; Rusek, Korben; and Thompson, David C., *"Simple homotopies for just the real roots of polynomial systems,"* preprint, Sandia National Laboratories, 2011.

[Pou71] Pourchet, Y., *"Sur la représentation en somme de carrés des polynômes á une indéterminée sur un corps de nombres algébriques,"* Acta Arithm. 19 (1971), pp. 89–104.

[RS02] Rahman, Q. I. and Schmeisser, G., *Analytic Theory of Polynomials*, London Mathematical Society Monographs 26, Oxford Science Publications, 2002.

[RY05] Rojas, J. Maurice and Ye, Yinyu, *"On solving sparse polynomials in logarithmic time,"* Journal of Complexity, special issue for the 2002 Foundations of Computation Mathematics (FOCM) meeting, February 2005, pp. 87–110.

[Sal76] Salamin, E., *"Computation of π using arithmetic-geometric mean,"* Math. Comput., 30 (1976), pp. 565–570

[SS96] Shub, Mike and Smale, Steve, *"The complexity of Bezout's theorem IV: probability of success; extensions,"* SIAM J. Numer. Anal., **33** (1996), no. 1, pp. 128–148.

[Sto00] Storjohann, Arne, *"Algorithms for matrix canonical forms,"* doctoral dissertation, Swiss Federal Institute of Technology, Zurich, 2000.

[Stu35] Sturm, Jacques Charles-François, *"Mémoire sur la résolution des équations numériques,"* Inst. France Sc. Math. Phys., **6** (1835).

HARVARD UNIVERSITY, MASSACHUSETTS HALL, CAMBRIDGE, MA 02138
E-mail address: hypo3400@gmail.com

MATHEMATICAL SCIENCES RESEARCH INSTITUTE, BERKELEY, CA 94720.
E-mail address: chillar@msri.org

MIT, 77 MASS. AVE., CAMBRIDGE, MA 02139
E-mail address: dpopov@mit.edu

TAMU 3368, TEXAS A&M UNIVERSITY, COLLEGE STATION, TX 77843-3368
E-mail address: rojas@math.tamu.edu