Some Speed-Ups and Speed Limits for Real Algebraic Geometry

J. Maurice Rojas*

Department of Mathematics
City University of Hong Kong
83 Tat Chee Avenue
Kowloon, HONG KONG
mamrojas@math.cityu.edu.hk
http://www.cityu.edu.hk/ma/staff/rojas

Abstract

We give new positive and negative results (some conditional) on speeding up computational algebraic geometry over the reals:

1. A new and sharper upper bound on the number of connected components of a semialgebraic set. Our new bound is novel in that it is stated in terms of the volumes of certain polytopes and, for a large class of inputs, beats the best previous bounds by a factor exponential in the number of variables.

2. A new algorithm for approximating the real roots of certain sparse polynomial systems. Two features of our algorithm are (a) arithmetic complexity **polylogarithmic** in the degree of the underlying complex variety (as opposed to the super-linear dependence in earlier algorithms) and (b) a simple and efficient generalization to certain univariate exponential sums.

3. Assuming the truth of the **Generalized Riemann Hypothesis (GRH)**, detecting the vanishing of **$\mathcal{A}$-discriminants** (operators which include all known multivariate resultants and discriminants as special cases) can be done within the complexity class **AM**.

4. Detecting whether a real algebraic surface (given as the common zero set of some input straight-line programs) is not smooth can be done in polynomial time within the classical Turing model (resp. BSS model over $\mathbb{C}$) only if **P**$=$**NP** (resp. **NP**$\subseteq$**BPP**).

The last result follows easily from an unpublished observation of Steve Smale. We then conclude by discussing some of the implications of (3) and (4) — especially whether the algorithm from (2) can be further generalized.

*This research was partially funded by a Hong Kong CERG Grant.

1 Introduction and Main Results

We provide new speed-ups for some fundamental computations in real algebraic geometry. Our techniques are motivated by recent results from algebraic geometry but the proofs are almost completely elementary. We then conclude with a discussion of how much farther these techniques can still be pushed.

In particular, roughly speaking, we also show that if singularity detection for curves over $\mathbb{C}$ can be done in polynomial time then we must have **P**$=$ **NP** or **NP** $\subseteq$ **BPP**. This can be thought of as a lower bound on the complexity of elimination theory. On the other hand, we show that the truth of GRH implies that deciding the vanishing of $\mathcal{A}$-discriminants (a central problem from elimination theory) can be done within the complexity class **AM**. The best previous bound was **EXPTIME** — a fact implicit from, say, [Stu93].

This work is a part of an ongoing program by the author [Roj97, Roj98, Roj99a, Roj99b] to dramatically sharpen current complexity bounds from algebraic geometry in terms of more intrinsic geometric invariants. We will give precise statements of these results shortly, so let us begin by considering the number of connected components of a semialgebraic set.[1]

[1]A **semialgebraic set** is simply a subset of $\mathbb{R}^n$ defined

1.1 Sharper Intrinsic Bounds

The topology of semialgebraic sets is intimately related to complexity theory in many ways. For example, the seminal work of Dobkin, Lipton, Steele, and Yao [DL79, SY82] (see also [BCSS98, Ch. 16]) relates upper bounds on the number of connected components to lower bounds on the algebraic circuit complexity of certain problems. More directly, upper bounds on connected components are an important ingredient in complexity upper bounds for the first order theory of the reals [BPR96].

Our first main theorem significantly improves earlier bounds on the number of connected components by Oleinik, Petrovsky, Milnor, Thom and Basu [OP49, Mil64, Tho65, Bas96].[2] The main novelty of our new bound is its greater sensitivity to the monomial term structure of the input polynomials. Letting $\mathbf{O}$ and $\hat{e}_i$ respectively denote the origin and the $i^{\underline{\text{th}}}$ standard basis vector in $\mathbb{R}^N$, $x := (x_1, \ldots, x_n)$, and normalizing k-dimensional volume $\mathrm{Vol}_k(\cdot)$ so that the standard k-simplex $\Delta_k := \{x \in \mathbb{R}^k \mid x_1, \ldots, x_n \geq 0, \sum x_j \leq 1\}$ has volume 1, our result is the following.

Main Theorem 1
Let $f_1, \ldots, f_{p+s} \in \mathbb{R}[x_1, \ldots, x_n]$ and suppose $S \subseteq \mathbb{R}^n$ is the solution set of the following collection of polynomial inequalities:

$$\begin{aligned} f_i(x) &= 0, \quad i \in \{1, \ldots, p\} \\ f_i(x) &> 0, \quad i \in \{p+1, \ldots, p+s\} \end{aligned}$$

Let $Q \subset \mathbb{R}^n$ be the convex hull of the union of $\{\mathbf{O}, \hat{e}_1, \ldots, \hat{e}_n\}$ and the set of all a with $x^a := x_1^{a_1} \cdots x_n^{a_n}$ a monomial term of some f_i. Then S has at most

$$\min\{n+1, \frac{s+1}{s-1}\} 2^n s^n \mathrm{Vol}_n(Q)$$

connected components.

For those familiar with **Newton polytopes** [BKK76], our Q above is simply the convex hull of the union of Δ_n and the Newton polytopes of all the f_i.

Letting d be the maximum of the total degrees of the f_i, the best previous general upper bounds, quoted from [BCSS98, Ch. 16, Prop. 5] and [Bas96] respectively, were $(sd+1)(2sd+1)^n$ and $(p+s)^n \mathcal{O}(d)^n$. Our bound is no worse than $\min\{n+1, \frac{s+1}{s-1}\}(2sd)^n$ (better than both preceding bounds) and is frequently much better. Consider the following examples:

Example 1 (Spikes) *Suppose we pick all the f_i to have the same monomial term structure, and in such a way that Q has small volume but great length in the direction $(1, \ldots, 1) \in \mathbb{R}^n$. In particular, let us assume that the only monomial terms occuring in the f_i are $1, x_1, \ldots, x_{n-1}$ and $(x_1 \cdots x_n), (x_1 \cdots x_n)^2, \ldots, (x_1 \cdots x_n)^D$. Then it is easy to check that Q is a "long and skinny" bypyramid, with one apex at the origin and the other at $(D, \ldots, D) \in \mathbb{R}^n$. We then obtain, via two simple determinants, that $\mathrm{Vol}_n(Q) = D+1$ and thus our bound reduces to $\min\{n+1, \frac{s+1}{s-1}\} 2^n s^n (D+1)$. However, the aforementioned older bounds are easily seen to reduce to $(nsD+1)(2nsD+1)^n$ and $((p+s)\mathcal{O}(nD))^n$.*

Example 2 (Bounded Multidegree) *Suppose now that instead of bounding the total degree of the f_i, we only require that the degree of f_i with respect to any x_j be at most d'. It is then easy to check that Q is an axes parallel hypercube with side length d'. So our new bound reduces to $\min\{n+1, \frac{s+1}{s-1}\}(2sd')^n$. However, the old bounds are easily seen to reduce to $(snd'+1)(2snd'+1)^n$ and $((p+s)\mathcal{O}(nd'))^n$.*

We can further compare our new bound to the best previous bounds in very simple polyhedral terms: Let Δ_Q denote the smallest scaled standard n-simplex, $\gamma\Delta_n$, containing Q. Then, since volume is monotonic under containment, our bounds are least favorable when $Q = \Delta_Q$. In particular, our bounds are never worse than the aforementioned earlier results, since $Q \subseteq d\Delta_n$ always.

Remark 1 *It is interesting to note that there are sharper (even optimal) results relating polytope volumes and connected components for* **complex** *varieties, beginning with the remarkable work of Bernshtein, Kushnirenko, and Khovanski [BKK76] a bit over twenty years ago. (See also [DK86].[3]) However, Main Theorem 1 presents the first non-trivial upper bounds on the number of connected*

by the solutions of a finite collection of polynomial inequalities.

[2]These papers actually bound **Betti numbers**, which in turn are an upper bound on the number of connected components. However, Main Theorem 1 can be extended to Betti numbers as well, and this will be covered in a forthcoming paper.

[3]We also point out that the classical Bézout's theorem [Mum95] is optimal only for a small class of polynomial systems. So the results of [BKK76] include Bézout's theorem as a very special case.

components of **semialgebraic** *sets with this combinatorial flavor. Finding an* **optimal** *upper bound for semialgebraic sets, even in the special case of nondegenerate real algebraic varieties, is a much harder problem and is still open.*

Our bound can be further improved in various ways and this is detailed in section 2. In particular, we give sharper versions tailored for certain special cases, and we prove analogues (for **all** our bounds) depending only on n, s, and the number of monomial terms which appear in at least one f_i.

It is also interesting to note that the techniques involved in our proof of Main Theorem 1, when combined with other recent results of the author [Roj99a], also yield similar improvements on the complexity of quantifier elimination over real-closed fields. This will be pursued in a forthcoming paper of the author.

1.2 Superfast Real Solving for Certain Fewnomial Systems

The complexity of solving systems of **fewnomials** (polynomials with few monomial terms[4]) has only been addressed recently. Indeed, the vast majority of work in computational algebra has so far been stated only in terms of degrees of polynomials, thus ignoring the finer monomial term structure. Notable exceptions include [CKS99] (solving a single univariate fewnomial over $\mathbb{Z}$ in polynomial time), [Len98] (solving a single univariate fewnomial over $\mathbb{Q}$ in polynomial time), and [Roj98, MP98, Roj99a] ("solving" polynomial systems over $\mathbb{R}$ or $\mathbb{C}$ within time near polynomial in the degree of the underlying complex variety[5]).

An important open question which still remains is whether the complexity of "solving" fewnomials can be **sub-linear** in the degree of the underlying complex variety. (We will clarify what it means to "solve" momentarily.) As an example, can one ε-approximate the roots of a univariate fewnomial of degree d, in the interval $[0, R]$, using significantly less than $\Theta(d \log\log \frac{R}{\varepsilon})$ arithmetic steps? Doing this even for **bi**nomials (i.e., quickly finding $d^{\underline{\text{th}}}$ roots) is nontrivial [Ye94]. The preceding complexity limit, up to a factor polylogarithmic in d, is the best current bound for solving a general univariate polynomial of degree d [NR96].

Our next main theorem gives an affirmative answer for certain fewnomial systems and univariate exponential sums over $\mathbb{R}$. More precisely, if $f(x) = \sum_{a \in \mathcal{A}} c_a x^a$, where $\mathcal{A} \subset \mathbb{R}$ finite and the coefficients c_a are all real, we call f an **exponential k-sum**. We define the **degree** of such an f to be $m_f \max_{a,a' \in \mathcal{A}}\{a - a'\}$, where $m_f := \max\{1, \frac{1}{\min\{a-a'\}}\}$ and the minimum is over all a and a' with $c_a, c_{a'} > 0$. (If f has only one positive coefficient then we set $m_f := 1$.) We also say that f has **j sign alternations** iff there are j distinct $a \in \mathcal{A}$ such that $c_a c_{a'} < 0$ and $\mathcal{A} \cap (a, a') = \emptyset$ for some $a' \in \mathcal{A}$ with $a' > a$. So, for instance, $47x^{2.53} - 10.3x^{0.9} - \pi - 10x^{-3} - x^{-5.5}$ has one sign alternation but $x^3 - 2x + 2$ has two. Also, when $\mathcal{A} \subset \mathbb{Z}$, we simply call f a **k-nomial**.

Main Theorem 2 *Let f be any exponential k-sum of degree d with at most one sign alternation. Then, given an oracle for evaluating x^r for any $x, r \in \mathbb{R}$, one can ε-approximate all the roots of f in $(0, R)$ using $\mathcal{O}(k(\log d + \log\log \frac{R}{\varepsilon}))$ arithmetic operations over $\mathbb{R}$ (including oracle calls). In particular, restricting to k-nomials and removing the oracle, we can still do the same using $\mathcal{O}(k \log d(\log d + \log\log \frac{R}{\varepsilon}))$ arithmetic operations over $\mathbb{R}$, with d agreeing with the ordinary degree of a univariate Laurent polynomial.*

Remark 2 *We point out that even the* **tri***nomial case is difficult. For example, while one can count the number of real roots of a trinomial of the form $x^d + ax + b$ within $\mathcal{O}(\log d)$ arithmetic operations [Ric93] (regardless of sign alternations), doing the same for general trinomials is still an open problem. From a more numerical point of view, even the use of Newton's method is subtle for trinomials: deciding whether a given initial point converges quadratically to a root of $x^3 - 2x + 2$ is undecidable in the BSS model over $\mathbb{R}$ [BCSS98, Sec. 2.4]. Nevertheless, this need not stop us from finding* **some** *good starting point, as we will soon see.*

Our algorithm, aside from an algebraic trick, closely follows an algorithm of Ye [Ye94] which efficiently blends binary search and Newton's method. By combining these ideas with a few facts on the **Smith normal form** of an integral matrix [Ili89], we can also derive the following complexity result on binomial systems.

Main Theorem 3 *Let $c_1, \ldots, c_n \in \mathbb{R} \setminus \{0\}$ and let $[d_{ij}]$ be any $n \times n$ matrix with nonnegative integer entries. Finally, let $f_i := x_1^{d_{i1}} \cdots x_n^{d_{in}} + c_i$ for*

[4] Results on fewnomials usually hold on a much broader class of functions: the so-called **Pfaffian** functions [Kho91].

[5] Joos Heintz and his school have a similar result over $\mathbb{C}$ with a larger complexity bound which, however, is applicable to the more general setting of **straight-line programs** [GHMP95].

all i. Then we can ε-approximate all the roots of $f_1 = \cdots = f_n = 0$ in the **orthant wedge** *$\{x \in \mathbb{R}^n \mid x_1, \ldots, x_n \geq 0, \sum x_i^2 \leq R\}$ within*

$$\mathcal{O}((n + \log \max\{|d_{ij}|\})^{6.376}) \text{ bit operations,}$$

followed by

$$\mathcal{O}\left(\log|\det[d_{ij}]| \left[n^3 \log^2(n \max\{|d_{ij}|\}) + \log\log \frac{R}{\varepsilon}\right]\right)$$

rational operations over $\mathbb{R}$.

If the above binomial system has only finitely many **complex** roots, then their number is exactly $|\det[d_{ij}]|$. This follows easily from **Bernshtein's theorem** [BKK76]. It is also interesting to note that for our preceding example, the fastest previous general (sequential) algorithms for polynomial system solving over $\mathbb{R}$ or $\mathbb{C}$ are only known to run in time **polynomial** in $|\det[d_{ij}]|$ [MP98, Roj99a] — that is, super-linear in the degree of the underlying complex variety.

One can of course solve slightly more general systems of fewnomials by threading together the algorithms of Main Theorems 2 and 3. We will say more on the likelihood of farther-reaching extensions of our last two results after first discussing an intriguing result relating complexity classes and singularities.

1.3 Obstructions to Superfast Degeneracy Detection

The preceding two algorithmic results circumvent degeneracy problems in simple but subtle ways. For instance, Main Theorem 2 clearly deals with equations having at most one positive real root, while the binomial systems of Main Theorem 3 are easily seen to have no repeated complex roots (cf. section 3). Thus, the respective hypotheses of these results (restricting sign alternations and/or number of monomial terms) allow us to approximate roots without stopping for a singularity check.

It seems hard to completely solve a system of equations without knowing something about its degeneracies, either a priori or during run-time. So let us present a result which gives solid evidence that detecting degeneracies may be quite difficult. In what follows, unless otherwise mentioned, we use the standard **sparse encoding** for multivariate polynomials [Pla84, Koi96]. Thus the **size** of a polynomial like $x^d + x - 47$ will be $\Theta(\log d)$ and not $\Theta(d)$, whether in the Turing model or the BSS model over $\mathbb{C}$.

Main Theorem 4 *Suppose any of the following problems can be solved in polynomial time via a Turing machine (resp. BSS machine over $\mathbb{C}$). Then* $\mathbf{P} = \mathbf{NP}$ *(resp.* $\mathbf{NP} \subseteq \mathbf{BPP}$*).*

1. *Decide if an input polynomial $f \in \mathbb{Z}[x_1]$ (resp. $f \in \mathbb{C}[x_1]$) vanishes at an $n^{\underline{\text{th}}}$ root of unity.*
2. *Decide if two input polynomials $f, g \in \mathbb{Z}[x_1]$ (resp. $f, g \in \mathbb{C}[x_1]$) have a common root.*
3. *Given a nonzero input polynomial $f \in \mathbb{Z}[x_1, x_2]$ (resp. $f \in \mathbb{C}[x_1, x_2]$) decide if the curve $\{(x_1, x_2) \in (\mathbb{C}^*)^2 \mid f(x_1, x_2) = 0\}$ has a singularity.*
4. *Given input polynomials $f, g \in \mathbb{Z}[x_1, x_2, x_3, x_4]$ (resp. $f, g \in \mathbb{R}[x_1, x_2, x_3, x_4]$), in the straight-line program encoding, defining a surface $S \subset \mathbb{R}^4$, decide if S has a singularity.*
5. *Given any finite subset $\mathcal{A} \subset \mathbb{Z}^2$ and a vector of coefficients $(c_a \mid a \in \mathcal{A}) \in \mathbb{Z}^{\#\mathcal{A}}$ (resp. $\in \mathbb{C}^{\#\mathcal{A}}$), decide if the $\mathcal{A}$-discriminant of the bivariate polynomial $\sum_{a \in \mathcal{A}} c_a x^a$ vanishes.*

Furthermore, problems (1)–(3), and the natural extension of problem (5) to $\mathcal{A} \subseteq \mathbb{Z}^n$, can all be done within **AM**.

Remark 3 *Note that in problem (4) we are already given that S is a surface. Determining whether this is true or not turns out to be* **NP***-hard (resp.* $\mathbf{NP}_{\mathbb{R}}$*-complete) in the Turing model (resp. BSS model over $\mathbb{R}$) [Koi99].*

For any $\mathcal{A} \subset \mathbb{Z}^n$, the **$\mathcal{A}$-discriminant**, $\mathcal{D}_A$, is defined to be the unique (up to sign) irreducible polynomial in $\mathbb{Z}[c_a \mid a \in \mathcal{A}]$ such that $f_{\mathcal{A}}(x) := \sum_{a \in \mathcal{A}} c_a x^a$ has a singularity in its zero set (in $(\mathbb{C}^*)^n$) $\Longrightarrow \mathcal{D}_A = 0$ [GKZ94]. This important operator lies at the heart of **sparse elimination theory**, which is the part of algebraic geometry surrounding this paper.

The $\mathcal{A}$-discriminant in fact contains all known multivariate resultants and discriminants as special cases, and also appears in residue theory and hypergeometric functions [GKZ94]. Thus, a corollary of our last main result is that sparse elimination theory, even in low dimensions, might lie beyond the reach of **P**.

Remark 4 *It is interesting to note that nontrivial lower bounds on the complexity of computing $\mathcal{A}$-discriminants in the* **one-dimensional** *case $\mathcal{A} \subset \mathbb{Z}$ are unknown. However, it is easy to show (via [GKZ94, pg. 274]) that one can at least find D_A in* **polynomial time** *when $\mathcal{A} \subset \mathbb{Z}^n$ has less than $n+3$ elements.*

We now prove our main theorems in order of appearance.

2 Binomial Fiberings and Main Theorem 1

Remark 5 *Throughout this section, "nonsingular" (or "smooth") for a real algebraic variety will mean that the underlying* **complex** *variety is nonsingular in the sense of the usual Jacobian criterion (see, e.g., [Mum95]). Also, we will let k denote the number of monomial terms which appear in at least one of $f_1, \ldots, f_{p+s}$.*

We begin with the following important special case of Main Theorem 1. This lemma is also frequently significantly sharper than many earlier results and may be of independent interest.

Lemma 1 *Following the notation of Main Theorem 1, suppose $p = 1$, $s = 0$, and S is smooth compact hypersurface. Then S has at most $\frac{1}{\min\{2,n\}}\mathrm{Vol}_n(Q')$ connected components, where Q' is the convex hull of the union of $\{\mathbf{O}\}$ and the set of all a with x^a a monomial term of f_1.*

Proof: The main idea will be to show that (for $n \geq 2$) the number of connected components is bounded above by half the number of critical points of a projection of a perturbed version of S. This idea is quite old, but we will introduce an unusual projection which permits a much sharper upper bound than before.

In particular, consider the function x^a with $a \in \mathbb{Z}^n \setminus \{\mathbf{O}\}$ to be selected later. The case $n=1$ of our bound is trivial so let us now assume $n \geq 2$. Clearly, any connected component of S (not lying in a hypersurface of the form $x^a = \text{constant}$) must have at least two special points: one locally maximizing, and the other locally minimizing, x^a. Since there are only finitely many connected components (by any earlier bound such as [OP49]), and every component is $(n-1)$-dimensional, there must therefore be an $a \in \mathbb{Z}^n \setminus \{\mathbf{O}\}$ so that every component (not lying entirely within the union of coordinate hyperplanes) contributes at least two critical points of x^a. Pick a in this way, subject to the additional minor restricition that the g.c.d. of the coordinates of a is 1.

Now consider $\tilde{f} := f_1 + \delta$ for some $\delta \in \mathbb{R}$ to be selected later. By Sard's theorem [Hir94], there is a set $W \subseteq \mathbb{R}$ of full measure such that $\delta \in W \Longrightarrow S_\delta = \{x \in \mathbb{R}^n \mid \tilde{f} = 0\}$ is nonsingular. Also, via a simple homotopy argument, S and S_δ are both smooth compact hypersurfaces and have the same number of connected components for $|\delta|$ sufficiently small. (Much stronger versions of this fact can be found in [Bas96].) Furthermore, note that for all but finitely many δ, no connected component of S_δ lies inside the union of the coordinate hyperplanes. We will pick $\delta \neq 0$ so that all these conditions, and one more to be described below, hold.

Note that the critical points of the function x^a on S_δ are just the solutions in $\mathbb{R}^n$ of

$$(\star) \qquad \tilde{f} = \frac{\partial \tilde{f}}{\partial y_2} = \cdots = \frac{\partial \tilde{f}}{\partial y_n} = 0,$$

where the y_i are new variables to be described shortly. Our final condition on δ (which is easily seen to hold for all but finitely many δ) will simply be that all real solutions to the above polynomial system lie in $(\mathbb{R}^*)^n := (\mathbb{R} \setminus \{0\})^n$. Note that a corollary of all our assumptions so far is that all **complex** solutions of $(\star)$ will be nonsingular and, in particular, the number of complex solutions is finite.

We are now essentially done: The number of connected components of S and S_δ are the same, and the latter quantity is bounded above by half the number of critical points (on S_δ) of the function x^a. This number of critical points can be computed in terms of polytope volumes as follows: Via the Smith normal form [Smi61], we can find an invertible change of variables on $(\mathbb{R}^*)^n$ such that $y_1 := x^a$ and $y_2, \ldots, y_n$ are monomials in the x_i. Furthermore, this change of variables induces the action of a unimodular matrix on the exponent vectors of $\tilde{f}$. In particular, $\tilde{f}$ can be considered as a polynomial in $\mathbb{R}[y_1^{\pm 1}, \ldots, y_n^{\pm 1}]$ and the number of monomial terms (and Newton polytope volume) of $\tilde{f}$ is preserved under this change of variables. Thus, up to a monomial change of variables, the critical points of the function x^a on S_δ are exactly the solutions in $(\mathbb{R}^*)^n$ of $(\star)$.

The key to our new bound is to finish things off by picking a bound **other** than Bézout's theorem here. In particular, by Bernshtein's theorem [BKK76], the number of solutions in $(\mathbb{C}^*)^n$ is at most the **mixed volume** of Q' and $n-1$ other polytopes with translates contained in Q'. By the monotonicity of the mixed volume [BZ88], the latter quantity is at most the mixed volume of n copies of Q' and, by the definition of mixed volume, this is just $\mathrm{Vol}_n(Q')$. ■

We point out that a key new ingredient in our proof is that the monomial change of variables we use (as opposed to the linear changes of variables used in most earlier proofs) preserves sparsity. This allows us to take full advantage of more powerful and refined techniques to bound the number of real roots, and thus get new bounds on the number of real connected components. For example, substituting Bernshtein's theorem for Bézout's theorem in the older proofs would not have yielded any significant improvement.

However, we need not have been so heavy-handed and only used tools over $\mathbb{C}$. We could have also used the following alternative bound on the number of real roots.

Khovanski's Theorem on Real Fewnomials (Special Case) *[Kho91, Sec. 3.12, Cor. 6] Suppose that for all $i \in \{1,\ldots,n\}$, $f_i \in \mathbb{R}[x_1,\ldots,x_n,m_1,\ldots,m_k]$ has total degree q_i, where the m_j are monomials in x. Assume further that the variety S defined by $f_1,\ldots,f_n$ is zero-dimensional and nonsingular. Then S has at most $(1+\sum q_i)^k 2^{k(k-1)/2}\prod q_i$ connected components in the positive orthant.* ■

We call any set of the form $\{x \in \mathbb{R}^n \mid \pm x_1,\ldots,\pm x_n \geq 0\}$ a **closed orthant**. When all signs are positive we call the corresponding closed orthant the **nonnegative orthant**. The analogous constructions where all inequalities are strict are, respectively, an **open** orthant and the **positive** orthant.

As an immediate corollary, our proof above yields the following alternative upper bound on the number of components of a smooth compact real algebraic hypersurface.

Corollary 1 *Following the notation and assumptions of lemma 1, the number of connected components of S is also at most $2^{n-1}(n+1)^{k+1}2^{k(k+1)/2}$. In particular, S has at most $\frac{1}{2}(n+1)^k 2^{k(k-1)/2}$ connected components contained entirely within the positive orthant.*

Proof: Following the notation of our last proof, note that multiplying any equation of $(\star)$ by a monomial in $y_1,\ldots,y_n$ does not affect the roots in $(\mathbb{R}^*)^n$. Thus, we can assume $(\star)$ has only $k+1$ distinct monomial terms. Also note that the monomial change of variables $x \mapsto y$ maps orthants onto orthants, and that the case $n=1$ is trivial. The first portion of our corollary then follows immediately from our last proof (using Khovanski's Theorem on Fewnomials with $q_1=\cdots=q_n=1$ instead of Bernshtein's Theorem), upon counting roots in all open orthants. The second portion follows even more easily, upon observing that we do not need δ if we only want to count critical points in an open orthant. ■

Returning to the proof of Main Theorem 1, the next step is to prove a slightly more general upper bound. Again, the following result is frequently much sharper than many earlier bounds and may also be of independent interest.

Lemma 2 *Following the notation of Main Theorem 1, suppose now that $s=0$, so that S is a real algebraic variety (not necessarily smooth or compact). Then S has at most $2^{n-1}\mathrm{Vol}_n(Q)$ connected components.*

Proof: The main trick is to reduce to the case considered by our preceding lemma. In particular, define $F_{\delta,\varepsilon} := f_1^2+\cdots+f_p^2+\varepsilon^2(\sum x_i^2)-\delta^2 \in \mathbb{R}[x_1,\ldots,x_n]$ and let $S_{\delta,\varepsilon}$ be the set of real zeroes of $F_{\delta,\varepsilon}$. It then follows that for sufficiently small (and suitably restricted) $\delta,\varepsilon>0$, $S_{\delta,\varepsilon}$ is a smooth compact hypersurface and the number of connected components of $S_{\delta,\varepsilon}$ is no smaller than the number of connected components of S. The proof of this fact is standard and a very clear account can be found in [BCSS98, Sec. 16.1].

In any event, the number of connected components of $S_{\delta,\varepsilon}$ is clearly at most $\frac{1}{2}\mathrm{Vol}_n(\mathrm{Conv}(2Q'\cup\{2\hat{e}_1,\ldots,2\hat{e}_n\}))$, thanks to our preceding lemma. Since the last quantity is just $\frac{1}{2}2^n\mathrm{Vol}_n(Q)$ we are done. ■

We can combine the proof of lemma 2 with Khovanski's Theorem on Fewnomials to obtain the following generalization of corollary 1. This result, while giving a slightly looser bound than an earlier result of Khovanski [Kho91, Sec. 3.14, Cor. 5], removes all nondegeneracy assumptions from his result.

Corollary 2 *Following the notation and assumptions of lemma 2, the number of connected components of S is also bounded above by* $4^{n-\frac{1}{2}}(2n+1)^{k+1}2^{k(k+1)/2}$.

Proof: Combining the proofs of lemmata 2 and 1, and since we are only counting roots in $(\mathbb{R}^*)^n$, we see that the number of connected components is at most half the number of solutions in $(\mathbb{R}^*)^n$ of the following polynomial system:

$$(\star\star) \quad \bar{F}_{\delta,\varepsilon}=y_2\frac{\partial\bar{F}_{\delta,\varepsilon}}{\partial y_2}=\cdots=y_n\frac{\partial\bar{F}_{\delta,\varepsilon}}{\partial y_n}=0,$$

where $\bar{F}_{\delta,\varepsilon}$ is the variant of $F_{\delta,\varepsilon}$ where we substitute $\sum y_i^2$ for $\sum x_i^2$. (It is a simple exercise to verify that the proof of lemma 2 still goes through with this variation.) Now simply note, via the chain rule of calculus, that every polynomial in $(\star\star)$ is of degree at most 2 in $y_1, \ldots, y_n$ and the set of monomials appearing in $f_1, \ldots, f_p$. Also note that the polynomials in $(\star\star)$ are polynomials in a total of $k+1$ monomial terms. So by Khovanski's Theorem on Real Fewnomials, and counting roots in all open orthants, we are done. ■

We are now ready to prove Main Theorem 1.

Proof of Main Theorem 1: We reduce again, this time to lemma 2. The trick here is to note that every connected component of S is in turn a connected component of S' where $S' := \{x \in \mathbb{R}^n \mid f_1(x) = \cdots = f_p(x) = 0, f_{p+1}(x) \neq 0, \ldots, f_{p+s}(x) \neq 0\}$. Every connected component of S' is in turn a projection (onto the first n coordinates) of a connected component of S'', where $S'' \subset \mathbb{R}^{n+1}$ is the real zero set of the polynomial system $(f_1, \ldots, f_p, -1 + z\prod_{i=p+1}^{p+s} f_i)$. This reduction is not new and appears, for example, in [BCSS98, Sec. 16.3].

Now lemma 2 tells us that the number of connected components of S'' is at most 2^n times the $(n+1)$-dimensional volume of $\mathrm{Conv}(P_1 \cup (P_2 \times \hat{e}_{n+1}))$, where P_1 (resp. P_2) is the union of $\{\mathbf{O}, \hat{e}_1, \ldots, \hat{e}_n\}$ and the Newton polytopes of $f_1, \ldots, f_p$ (resp. the **Minkowski sum** of the Newton polytopes of $f_{p+1}, \ldots, f_{p+s}$). However, it is a simple exercise to show that $P_2 \subseteq P_3$ where P_3 is the union of $\{\mathbf{O}, \hat{e}_1, \ldots, \hat{e}_n\}$ and the Newton polytopes of $f_{p+1}, \ldots, f_{p+s}$, scaled by a factor of s. Now note that $P_2 \subseteq Q$, $P_3 \subseteq sQ$ and $\mathrm{Conv}(P_1 \cup (P_2 \times \hat{e}_{n+1})) \subseteq \mathrm{Conv}(Q \cup (sQ \times \hat{e}_{n+1}))$.

If $s > 1$ then the last polytope is in turn contained in a pyramid P with apex at $(0, \ldots, 0, \frac{-1}{s-1})$ and base $Q \times \hat{e}_{n+1}$. So we obtain that the number of connected components of S is at most $2^n \mathrm{Vol}_{n+1}(P) = 2^n \frac{s+1}{s-1} \mathrm{Vol}_n(sQ) = \frac{s+1}{s-1} 2^n s^n \mathrm{Vol}_n(Q)$.

If $s = 1$ then $\mathrm{Conv}(Q \cup (sQ \times \hat{e}_{n+1})) = [\mathbf{O}, \hat{e}_{n+1}] \times Q$. So, similar to the previous case, the number of connected components of S is at most $2^n \mathrm{Vol}_{n+1}(P) = 2^n n \mathrm{Vol}_n(Q)$.

Now note that the number of connected components of S will always be at most $\min\{n+1, \frac{s+1}{s-1}\} 2^n s^n \mathrm{Vol}_n(Q)$, with the possible exception of the case $(n, s) = (1, 2)$. So we need only check this final case. However, this is almost trivial, separating the cases $p > 0$ and $p = 0$. ■

We can give an alternative version of Main Theorem 1, solely in terms of n, s, and k, as follows.

Theorem 5 *Following the notation and assumptions of Main Theorem 1, the number of connected components of S is also bounded above by* $4^{n-\frac{1}{2}}(s+1)^n(2(n+1)(s+1)+1)^{k+1} 2^{k(k+1)/2}$. ■

The proof is very similar to that of corollary 2, save only that we substitute the polynomial system from the proof of Main Theorem 1 into the construction of $\bar{F}_{\delta,\varepsilon}$. In particular, we eventually obtain a system of $n+1$ polynomials of degree $2(s+1)$ in a total of $k+1$ monomials, thus allowing yet another application of Khovanksi's beautiful theorem on fewnomials.

3 Alpha Theory and Proving Main Theorems 2 and 3

The proof of Main Theorem 2 hinges on **alpha theory** [BCSS98], which gives useful criteria for when Newton's method converges quadratically. In particular, we will need the following elementary analytic lemma.

Lemma 3 *For any monotonic function $\phi : \mathbb{R} \longrightarrow \mathbb{R}$, let α_ϕ satisfy $\sup_{k>1} |\frac{\phi^{(k)}(x)}{k!\phi'(x)}|^{\frac{1}{k-1}} \leq \frac{\alpha_\phi}{x}$. Then, for $\phi(x) = x^r$, we may take α_ϕ equal to $\lceil |r| \rceil$, 2 or 1, according as $r \in (-\infty, -1) \cup (1, \infty)$, $r \in (0, 1)$, or $r \in (-1, 0)$. More generally, if $\phi = \phi_1 + \phi_2$ with ϕ_1 and ϕ_2 both convex and either both increasing or both decreasing, then we can take $\alpha_\phi = \max\{\alpha_{\phi_1}, \alpha_{\phi_2}\}$.* ■

The first part is a simple exercise while the second part is a proposition from [Ye94].

We are now ready to sketch the proof of Main Theorem 2.

Proof of Main Theorem 2: We begin by changing our function f slightly. First let M be largest exponent occuring in the k-sum f and let m be the smallest real number so that x^m is a monomial term of f with **positive** coefficient. (We assume, by multiplying by -1 if necessary, that the leading coefficient of f is positive.) By dividing out by x^m we may assume that $m = 0$. Via the change of variables $x = y^{1/M}$, we may further assume that $M = 1$. In particular, we now obtain that f is a sum of two increasing convex functions: one a positive linear combination of powers of x (with exponents in $(0, 1]$), the other a negative linear combination of powers of x (with exponents in $(-\infty, 0)$).

By our preceding lemma, we may take $\alpha_f = d$ (the degree of f) since d is no smaller than the degree of our original f. We now invoke the hybrid algorithm from [Ye94, Theorem 3]: This algorithm allows us to ε-approximate the real roots of f in $(0, R)$ using $\mathcal{O}(\log \alpha_f + \log\log \frac{R}{\varepsilon}) = \mathcal{O}(\log d + \log\log \frac{R}{\varepsilon})$ function evaluations and arithmetic operations. To conclude the first part of this main theorem, inverting the change of variables we made requires another $\mathcal{O}(\log d + \log\log \frac{R}{\varepsilon})$ operations via the same algorithm (since taking $n^{\underline{\text{th}}}$ roots is the same as solving an exponential 2-sum). However, we may have decreased the accuracy of our ε-approximation. So we just begin by solving to accuracy $\min\{\varepsilon^{M-m}, \varepsilon\}$ instead to obtain the first part of our main theorem. (Note also that evaluating f requires k uses of our oracle.)

To obtain the second part of our theorem, we simply use the same algorithm without the oracle. This simply introduces another factor of $\log d$ since monomials can now be evaluated by the usual repeated squaring trick. ■

Main Theorem 3 only needs a special case of Main Theorem 2. In fact, [Ye94] contains a slightly modified algorithm for the binomial case with an even better complexity bound of $\mathcal{O}(\log d \log\log \frac{R}{\varepsilon})$, which we will use below. However, we will also require some refined quantitative facts about the Smith normal form of a matrix.

Lemma 4 *[Ili89] Let $A = [a_{ij}]$ be any $n \times n$ matrix with entries only in $\mathbb{Z}$ and define h_A to be $\log(2n + \max\{|a_{ij}|\})$. Then, within $\mathcal{O}^*((n + h_A)^{6.375})$ bit operations, one can find matrices U, D, V with the following properties:*

1. *U and V both have determinant ± 1 and entries only in $\mathbb{Z}$.*
2. *D is diagonal and has entries only in $\mathbb{Z}$.*
3. *$UAV = D$*
4. *$\det A$ is the product of the diagonal elements of D and $h_U, h_V = \mathcal{O}(n^3(h_A + \log n)^2)$.*

Proof of Main Theorem 3: We begin by immediately applying the Smith normal form to our matrix $[d_{ij}]$. (This accounts for the bit operation count.) Clearly then, we have reduced to the case of n binomials of the form $x_1^{d_1} - \gamma_1, \ldots, x_n^{d_n} - \gamma_n$. The real roots of this polynomial system can then be ε-approximated by n applications of Main Theorem 2. Since $\sum \log d_i = \log \prod d_i = |\det[d_{ij}]|$, this accounts for almost all of the second bound.

To conclude, note that we must still invert our change of variables. By lemma 4, computing this monomial map is almost the final contribution to our second complexity bound. The only missing part is the fact that we may have needed more accuracy at the beginning of our algorithm. Lemma 4 also tells us how much more accuracy we need, thus finally accounting for all of our second complexity bound. ■

4 Smale's Theorem and Main Theorem 4

We begin with the following result of Plaisted.

Plaisted's Theorem *[Pla84] Deciding if an input polynomial $f \in \mathbb{C}[x_1]$ vanishes at an $n^{\underline{\text{th}}}$ root of unity is* **NP***-hard.* ■

In the above (and in what follows) f is given in the sparse encoding and n is also part of the input.

The following unpublished result of Steve Smale gives an intriguing extension of Plaisted's result via computations over new rings.

Smale's Theorem *Suppose we can decide if an input polynomial $f \in \mathbb{C}[x_1]$ vanishes at an $n^{\underline{\text{th}}}$ root of unity within polynomial time, in the BSS model over $\mathbb{C}$. Then* $\mathbf{NP} \subseteq \mathbf{BPP}$.

Proof: Given any complexity class $\mathcal{C}$ over the Turing model, consider its extension $\mathcal{C}_{\mathbb{C}}$ to the BSS model over $\mathbb{C}$. It is then a simple fact that $\mathcal{C}$ is contained in the **Boolean part** of $\mathcal{C}_{\mathbb{C}}$, $\mathbf{BP}(\mathcal{C}_{\mathbb{C}})$ [CKKLW95]. However, we will make use of an inclusion going the opposite way: $\mathbf{BP}(\mathcal{C}_{\mathbb{C}}) \subseteq \mathcal{C}^{\mathbf{BPP}}$ [CKKLW95]. Applying this to the problem at hand, we thus see that the hypothesis of our theorem, thanks to Plaisted's Theorem, implies that $\mathbf{NP} \subseteq \mathbf{BP}(\mathbf{P}_{\mathbb{C}}) = \mathbf{P}^{\mathbf{BPP}} = \mathbf{BPP}$. So we are done. ■

The first part of our final main theorem then follows from some simple reductions to problem (1) from the statement. The second will follow from a result of Koiran [Koi96] and a technical result on the vanishing of $\mathcal{A}$-discriminants [Roj99a].

Proof of Main Theorem 4: We will first prove the lower bound portion of our main theorem.

First note that the assertion concerning problem (1) follows immediately from Smale's Theorem and Plaisted's Theorem. It thus suffices to successively reduce (1) to special cases of all the other problems.

The assertion for (2) is then clear, since via the special case $g(x) = x^n - 1$, any polynomial time algorithm for (2) would give a polynomial time algorithm for (1).

On the other hand, a polynomial time algorithm for problem (5) would imply a polynomial time algorithm for problem (2). This is because problem (2) is essentially the decision problem of whether the **sparse resultant** of f and g [GKZ94] is zero. Via the **Cayley trick** [GKZ94], the $\mathcal{A}$-discriminant for $\mathcal{A} = P \cup (Q \times \hat{e}_2)$ (where P and Q are respectively the supports of f and g) is exactly the sparse resultant of f and g. So this portion is done.

Note also that (3) is just a reformulation of (5).

As for (4), via the Jacobian criterion for singularities [Mum95] applied to the real and imaginary parts of the input to (3), a polynomial time algorithm for (4) (using the straight-line program encoding for the input) would immediately imply a polynomial time algorithm for (3) (using the straight-line program encoding for the input). Such an algorithm would then immediately be a polynomial time algorithm for (3) with inputs given in the sparse encoding.

To conclude, note that Koiran's result that Hilbert's Nullstellensatz is in **AM** [Koi96] **almost** implies that our extension of (5) lies in **AM.** The key difference is that the vanishing of resultants measures degeneracies in a particular **toric variety**, not in $\mathbb{C}^n$ [GKZ94]. However, via the results of [Roj99a], we can reduce checking the vanishing of an $\mathcal{A}$-discriminant to a polynomial number of instances of Hilbert's Nullstellensatz. So our extension of (5) lies in **AM.**

That problems (1)–(3) now lie in **AM** follows easily from our preceding reductions. ■

5 Acknowledgements

The author thanks Felipe Cucker, Askold Khovanski, and Steve Smale for some very useful discussions. In particular, he thanks Felipe Cucker for pointing out an elegant proof of Smale's Theorem.

References

[Bas96] Basu, Saugata, *"On Bounding the Betti Numbers and Computing the Euler Characteristic of Semi-Algebraic Sets,"* Proceedings of the Twenty-eighth Annual ACM STOC (Philadelphia, PA, 1996), pp. 408–417, ACM, New York.

[BPR96] Basu, S., Pollack, R., Roy, M.-F., *"On the Combinatorial and Algebraic Complexity of Quantifier Elimination,"* Journal of the ACM, Vol. 43, No. 6, November 1996, pp. 1002–1045.

[BKK76] Bernshtein, D. N., Kushnirenko, A. G., and Khovanski, A. G., *"Newton Polyhedra,"* Uspehi Mat. Nauk 31 (1976), no. 3(189), pp. 201–202.

[BCSS98] Blum, L., Cucker, F., Shub, M., Smale, S., *Complexity and Real Computation,* Springer-Verlag, 1998.

[BZ88] Burago, Yu. D. and Zalgaller, V. A., *Geometric Inequalities,* Grundlehren der mathematischen Wissenschaften 285, Springer-Verlag (1988).

[CKKLW95] Cucker, F., Karpinski, M., Koiran, P., and Lickteig, T., and Werther, K., *"On Real Turing Machines that Toss Coins,"* Proceedings of the 27th STOC, pp. 335–342, ACM Press, 1995.

[CKS99] Cucker, F., Koiran, P., and Smale, S., *"A Polynomial Time Algorithm for Diophantine Equations in One Variable,"* Journal of Symbolic Computation (1999) **27**, pp. 21–29.

[DK86] Danilov, V. I. and Khovanski, A. G., *"Newton Polyhedra and an Algorithm for Calculating Hodge-Deligne Numbers,"* Math. USSR-Izv. 29 (1987), no. 2, pp. 279–298.

[DL79] Dobkin, David and Lipton, Richard, *"On the Complexity of Computations Under Varying Sets of Primitives,"* J. of Computer and System Sciences 18, pp. 86–91, 1979.

[GKZ94] Gel'fand, I. M., Kapranov, M. M., and Zelevinsky, A. V., *Discriminants, Resultants and Multidimensional Determinants,* Birkhäuser, Boston, 1994.

[GHMP95] Giusti, M., Heintz, J., Morais, J. E., Pardo, L. M., *"When Polynomial Equation Systems can be 'Solved' Fast?,"* Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (Paris, 1995), 205–231, Lecture Notes in Comput. Sci. 948, Springer, Berlin, 1995.

[Hir94] Hirsch, Morris, *Differential Topology,* corrected reprint of the 1976 original, Graduate Texts in Mathematics, 33, Springer-Verlag, New York, 1994.

[Ili89] Iliopoulos, Costas S., *"Worst Case Complexity Bounds on Algorithms for Computing the Canonical Structure of Finite Abelian Groups and the Hermite and Smith Normal Forms of an Integer Matrix,"* SIAM Journal on Computing, 18 (1989), no. 4, pp. 658–669.

[Kho91] Khovanski, Askold, *Fewnomials,* AMS Press, Providence, Rhode Island, 1991.

[Koi96] Koiran, Pascal, *"Hilbert's Nullstellensatz is in the Polynomial Hierarchy,"* DIMACS Technical Report 96-27, July 1996. (**Note:** This preprint considerably improves the published version which appeared in Journal of Complexity in 1996.)

[Koi99] __________, *"The Real Dimension Problem is* $\mathbf{NP}_{\mathbb{R}}$*-Complete,"* LIP Research Report 97-36 (ENS Lyon), to appear in Journal of Complexity.

[Len98] Lenstra, Hendrik W., *"Finding Small Degree Factors of Lacunary Polynomials,"* Number Theory in Progress, proceedings of a meeting in honor of the 70th birthday of Andrej Schnizel, W. de Gruyter, to appear.

[Mil64] Milnor, John *"On the Betti Numbers of Real Varieties,"* Proceedings of the Amer. Math. Soc. 15, pp. 275–280, 1964.

[MP98] Mourrain, Bernard and Pan, Victor Y. *"Asymptotic Acceleration of Solving Multivariate Polynomial Systems of Equations,"* Proc. ACM STOC 1998.

[Mum95] Mumford, David, *Algebraic Geometry I: Complex Projective Varieties,* Reprint of the 1976 edition, Classics in Mathematics, Springer-Verlag, Berlin, 1995.

[NR96] Neff, C. Andrew and Reif, John, *"An Efficient Algorithm for the Complex Roots Problem,"* Journal of Complexity **12** (1996), no. 2, pp. 81–115.

[OP49] Oleinik, O. and Petrovski, I., *"On the Topology of Real Algebraic Hypersurfaces,"* Izv. Akad. Akad. Nauk SSSR 13, pp. 389–402, 1949.

[Pla84] Plaisted, David A., *"New NP-Hard and NP-Complete Polynomial and Integer Divisibility Problems,"* Theoret. Comput. Sci. 31 (1984), no. 1–2, 125–138.

[Ric93] Richardson, D., *"Finding the Number of Distinct Real Roots of Sparse Polynomials of the Form $p(x, x^n)$,"* Computational Algebraic Geometry (Nice, 1992), 225–233, Progr. Math., 109, Birkhäuser, Boston, MA, 1993.

[Roj97] Rojas, J. Maurice, *"Toric Laminations, Sparse Generalized Characteristic Polynomials, and a Refinement of Hilbert's Tenth Problem,"* Foundations of Computational Mathematics (Rio de Janeiro, January 1997), pp. 369–381, Felipe Cucker and Mike Shub (eds.), Springer-Verlag (1997).

[Roj98] ________________, *"Intrinsic Near Quadratic Complexity Bounds for Real Multivariate Root Counting,"* Proceedings of the Sixth Annual European Symposium on Algorithms, Lecture Notes on Computer Science, pp. 127–138, vol. 1461, Springer-Verlag (1998).

[Roj99b] ________________, *"On the Complexity of Diophantine Geometry in Low Dimensions,"* Proceedings of the 31st Annual ACM STOC (May 1-4, 1999, Atlanta, Georgia), pp. 527–536, ACM Press, 1999.

[Roj99a] ________________, *"Solving Degenerate Sparse Polynomial Systems Faster,"* Journal of Symbolic Computation, vol. 27 (special issue on elimination theory), 1999.

[Smi61] Smith, H. J. S., *"On Systems of Integer Equations and Congruences,"* Philos. Trans. 151, pp. 293–326 (1861).

[SY82] Steele, J. and Yao, A., *"Lower Bounds for Algebraic Decision Trees,"* J. of Algorithms 3, pp. 1–8, 1982.

[Stu93] Sturmfels, Bernd, *"Sparse Elimination Theory,"* In D. Eisenbud and L. Robbiano, editors, Proc. Computat. Algebraic Geom. and Commut. Algebra 1991, pages 377–396, Cortona, Italy, 1993, Cambridge Univ. Press.

[Tho65] Thom, Reneé, *"Sur l'homologie des variétés algébriques réelles,"* In S. Cairns (Ed.), Differential and Combinatorial Topology, Princeton University Press, 1965.

[Ye94] Ye, Yinyu, *"Combining Binary Search and Newton's Method to Compute Real Roots for a Class of Real Functions,"* J. Complexity 10 (1994), no. 3, 271–280.