JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 22, Number 2, April 2009, Pages 363–385
S 0894-0347(08)00630-9
Article electronically published on November 6, 2008

SMALE'S 17TH PROBLEM: AVERAGE POLYNOMIAL TIME TO COMPUTE AFFINE AND PROJECTIVE SOLUTIONS

CARLOS BELTRÁN AND LUIS MIGUEL PARDO

1. Introduction

In the series of papers [SS93a, SS93b, SS93c, SS94, SS96], Shub and Smale defined and studied in depth a homotopy method for solving systems of polynomial equations. Some articles preceding this new treatment were [Kan49, Sma86, Ren87, Kim89, Shu93]. Other authors have also treated this approach in [Mal94, Yak95, Ded97, BCSS98, Ded01, Ded06, MR02] and more recently in [Shu08, BS08]. In a previous paper [BP08], the authors furthered the program initiated in the series [SS93a] to [SS94], describing an algorithm that computes projective approximate zeros of systems of polynomial equations in polynomial running time, with bounded (small) probability of failure.

In this paper, we give an updated version of some of the concepts introduced in [BP08], and we develop two important extensions of the results therein.

On one hand, we describe a procedure that, instead of assuming a small probability of error, finds an approximate zero of systems of polynomial equations, on the average, in polynomial time (cf. Theorem 1.10). This improvement is made in order to answer explicitly the question posted by Smale in his 17th problem.

On the other hand, we extend the result to the computation of affine solutions of systems of polynomial equations. This new result requires us to understand the probability distribution of the norm of the affine solutions of polynomial systems of equations (Theorem 1.9 below).

These results may be summarized as follows.

Theorem 1.1 (Main)**.** *There exists a uniform probabilistic algorithm that computes an approximate zero - both projective and affine - of systems of polynomial equations (with probability of success* 1*). The average number of arithmetic operations of this algorithm is polynomial in the size of the input.*

More specifically, the kind of algorithm that we obtain belongs to the class **Average ZPP** (for **Z**ero error probability, **P**robabilistic, Average **P**olynomial Time), or equivalently **Average Las Vegas**. At the end of this Introduction we have added an Appendix where we clarify the terminology.

In this paper, c denotes a positive constant.

Received by the editors November 2, 2006.

2000 *Mathematics Subject Classification.* Primary 65H20, 14Q20; Secondary 12Y05, 90C30.

Key words and phrases. Approximate zero, homotopy methods, complexity of methods in algebraic geometry.

The authors' research was partially supported by FECYT, Spanish Ministry of Science, MTM2007-62799, and an NSERC grant.

1.1. **Background.** Let $n \in \mathbb{N}$ be a positive integer. Let $(d) = (d_1, \ldots, d_n) \in \mathbb{N}^n$ be a list of positive degrees, and let $\mathcal{H}_{(d)}$ be the space of all systems $f = [f_1, \ldots, f_n] : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ of n polynomial equations and n unknowns, of respective degrees bounded by $d_1, \ldots, d_n$. Observe that $\mathcal{H}_{(d)}$ is also equivalent to the space of all systems of homogeneous polynomial equations $f : \mathbb{C}^{n+1} \longrightarrow \mathbb{C}^n$ of degrees d_i. Indeed, for each system in the unknowns $X_1, \ldots, X_n$, we can consider its homogeneous counterpart, adding a new unknown X_0 to homogenize all the monomials of each equation to the same degree d_i. The set of projective zeros of $f = [f_1, \ldots, f_n] \in \mathcal{H}_{(d)}$ is

$$V_{\mathbb{P}}(f) = \{x \in \mathbb{P}_n(\mathbb{C}) : f_i(x) = 0, 1 \leq i \leq n\} \subseteq \mathbb{P}_n(\mathbb{C}),$$

where the f_i are seen as polynomial homogeneous mappings $f_i : \mathbb{C}^{n+1} \longrightarrow \mathbb{C}$. Observe that this set is naturally contained in the complex n-dimensional projective space $\mathbb{P}_n(\mathbb{C})$. Namely, a point $x \in \mathbb{C}^{n+1}$ is a zero of f_i, $1 \leq i \leq n$, if and only if $\lambda x \in \mathbb{C}^{n+1}$ is a zero of f_i, where $0 \neq \lambda \in \mathbb{C}$ is any complex number. On the other hand, we may consider the set of affine solutions of an element $f \in \mathcal{H}_{(d)}$,

$$V_{\mathbb{C}}(f) = \{x \in \mathbb{C}^n : f_i(x) = 0, 1 \leq i \leq n\} \subseteq \mathbb{C}^n,$$

where the f_i are seen as polynomial mappings $f_i : \mathbb{C}^n \longrightarrow \mathbb{C}$. Let φ_0 be the standard embedding of $\mathbb{C}^n$ into $\mathbb{P}_n(\mathbb{C})$,

$$\begin{array}{rccc} \varphi_0 : & \mathbb{C}^n & \longrightarrow & \mathbb{P}_n(\mathbb{C}) \setminus \{x_0 = 0\}, \\ & (x_1, \ldots, x_n) & \mapsto & (1 : x_1 : \cdots : x_n). \end{array} \tag{1.1}$$

Then, we may write $V_{\mathbb{C}}(f) = \varphi_0^{-1}(V_{\mathbb{P}}(f))$. Namely,

$$x \in V_{\mathbb{C}}(f) \Longrightarrow \varphi_0(x) \in V_{\mathbb{P}}(f),$$

$$z = (z_0 : \cdots : z_n) \in V_{\mathbb{P}}(f), z_0 \neq 0 \Longrightarrow \varphi_0^{-1}(z) \in V_{\mathbb{C}}(f).$$

We denote by $d = \max\{d_i : 1 \leq i \leq n\}$ the maximum of the degrees, and by $\mathcal{D} = d_1 \cdots d_n$ the Bézout number associated with the list (d). From now on, we assume that $d \geq 2$. We denote by $N+1$ the complex dimension of $\mathcal{H}_{(d)}$ as a vector space. We may consider N as the size of the input of our target systems. Here we summarize the notation.

n	Number of equations
$(d) = (d_1, \ldots, d_n)$	List of degrees
d	$\max\{d_1, \ldots, d_n\}$
$\mathcal{H}_{(d)}$	Space of systems of equations associated with (d)
$N+1$	Complex dimension of $\mathcal{H}_{(d)}$
$\mathcal{D}$	Bezóut number$=d_1 \cdots d_n$

As in [SS93a], we consider $\mathcal{H}_{(d)}$ equipped with the Bombieri-Weyl Hermitian product $\langle \cdot, \cdot \rangle_\Delta$ and the associated Hermitian structure (see Section 2 for a precise definition). We denote by $\mathbb{S}_{(d)}$ the sphere in $\mathcal{H}_{(d)}$ for this Hermitian product, with the inherited Riemannian structure. We will also consider the incidence variety W,

$$W = \{(f, \zeta) \in \mathcal{H}_{(d)} \times \mathbb{P}_n(\mathbb{C}) : f \neq 0,\ \zeta \in V_{\mathbb{P}}(f)\}.$$

As pointed out in [BCSS98], W is a differentiable manifold of complex dimension $N+1$.

The projective Newton operator was initially described in [Shu93], and studied in depth in the series of papers by Shub and Smale. Let $f \in \mathcal{H}_{(d)}$ be seen as a

system of homogeneous polynomial equations and let $z \in \mathbb{C}^{n+1}$. Let $\widehat{df}(z)$ be the tangent mapping of f at z, considering f as a map from $\mathbb{C}^{n+1}$ to $\mathbb{C}^n$, and let $z^\perp$ be the Hermitian complement of z in $\mathbb{C}^{n+1}$. Assume that $\widehat{df}(z) \mid_{z^\perp}$ is a bijective mapping. Then, we define

$$\widehat{N}_f(z) = z - \left(\widehat{df}(z) \mid_{z^\perp}\right)^{-1} f(z) \in \mathbb{P}_n(\mathbb{C}).$$

We denote by $\widehat{N}_f^{(k)}(z) = \widehat{N}_f \circ \overset{k}{\cdots} \circ \widehat{N}_f(z)$ the projective point obtained after k applications of $\widehat{N}_f$, starting at z. Assume that $z \in \mathbb{P}_n(\mathbb{C})$ is such that $\widehat{N}_f^{(k)}(z)$ is defined for every $k \geq 0$, and that there exists a point $\zeta \in V_{\mathbb{P}}(f)$ such that

$$d_T\left(\widehat{N}_f^{(k)}(z), \zeta\right) \leq \frac{1}{2^{2^k-1}} d_T(z, \zeta), \ \ \forall k \geq 0,$$

where d_T is the tangent of the Riemannian distance in $\mathbb{P}_n(\mathbb{C})$. Then, we say that z is a projective approximate zero of f, with associated (true) zero ζ. Note that using z as a starting point for projective Newton's operator, we may quickly obtain an approximation as close as wanted to ζ. Hence, a major objective is the efficient computation of projective approximate zeros.

In [SS93a], Shub and Smale studied the behavior of the projective Newton operator in terms of a normalized condition number of polynomial systems. For $f \in \mathcal{H}_{(d)}$ and $z \in \mathbb{P}_n(\mathbb{C})$, they defined a quantity $\mu_{\text{norm}}(f, z)$ (see equation (3.1) for a precise definition) that satisfies the following property.

Proposition 1.2 (Shub and Smale)**.** *Let $(f, \zeta) \in W$, and let $z \in \mathbb{P}_n(\mathbb{C})$ be such that*

$$d_T(z, \zeta) \leq \frac{3 - \sqrt{7}}{d^{3/2} \mu_{\text{norm}}(f, \zeta)}.$$

Then, z is a projective approximate zero of f with associated zero ζ.

Now we briefly describe the Homotopic Deformation procedure (HD for short), a procedure that attempts to find projective approximate zeros of systems developed by Shub and Smale (see [SS93a, SS96] and mainly [SS94]).

Let $f \in \mathbb{S}_{(d)}$ be a target system, and let $g \in \mathbb{S}_{(d)}$ be another system that has a known solution $\zeta_0 \in \mathbb{P}_n(\mathbb{C})$. Consider the segment $\{f_t = tf + (1-t)g, t \in [0,1]\}$ joining g and f. Under some regularity hypothesis (see Proposition 3.1), the Implicit Function Theorem defines a differentiable curve $\{(f_t, \zeta_t), t \in [0,1]\} \subseteq \mathcal{H}_{(d)} \times \mathbb{P}_n(\mathbb{C})$, such that $f_t(\zeta_t) = 0, \ \forall t \in [0,1]$. This curve will be denoted by $\Gamma(f, g, \zeta_0)$.

The HD is a procedure that constructs a polygonal path that closely follows $\Gamma(f, g, \zeta_0)$. This path has initial vertex (g, ζ_0) and final vertex (f, z_1), for some $z_1 \in \mathbb{P}_n(\mathbb{C})$. The output of the algorithm is $z_1 \in \mathbb{P}_n(\mathbb{C})$. The polygonal path is constructed by "homotopy steps" (path following methods), each of which is an application of the projective Newton operator, with an appropriate step size selection.

In [SS94], Shub and Smale proved the existence of optimal initial pairs that guarantee a good performance of this algorithm, but their existential proof does not give any hint on how to compute these pairs (a full discussion of these aspects with historical comments may be read in detail in [BP06]). The lack of hints on how to find these initial pairs leads both to Shub and Smale's Conjecture (as in [SS94]) and to the problem mentioned in the title of this paper.

Smale's 17th problem. *Can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?*

Here, the term "uniform" emphasizes the fact that the algorithm demanded by Smale must be described explicitly (see the discussion in the Appendix to the Introduction), and "average" is used in the sense of the Bombieri-Weyl product. That is, in our notation, "average on $\mathbb{S}_{(d)}$". Smale also wrote:

Certainly finding zeros of polynomials and polynomial systems is one of the oldest and most central problems of mathematics. Our problem asks if, under some conditions specified in the problem, it can be solved systematically by computers.

1.2. **Statement of the main outcomes.** The formal statement of our main results requires the introduction of some precise terminology. We also recall some concepts and results from our previous paper [BP08].

1.2.1. *Projective solutions.* An important quantity helps to control the number of homotopy steps used in HD algorithms and, hence, the complexity of the algorithm. For a pair $f, g \in \mathbb{S}_{(d)}$ and a solution $\zeta \in V(g)$, we define

$$\mu_{\mathrm{norm}}(f, g, \zeta) = \sup_{(h,z)\in\Gamma(f,g,\zeta)} [\mu_{\mathrm{norm}}(h, z)].$$

Definition 1.3. We say that (g, ζ) is an *efficient initial pair* if

$$\mathrm{E}_{f\in\mathbb{S}_{(d)}}[\mu_{\mathrm{norm}}(f, g, \zeta)^{2-\beta}] \leq 10^5 n^5 N^2 d^{3/2} \log_2 \mathcal{D},$$

where $\beta = \frac{1}{\log_2 \mathcal{D}}$.

Note that the dependence on ε of this concept (appearing in our previous paper [BP08]) has disappeared. The following result follows from the arguments of Shub and Smale in [SS94] (see Section 3.3 for a proof).

Theorem 1.4. *Assume that (g, ζ_0) is an efficient initial pair. Then, the average number of projective Newton steps performed by Shub and Smale's homotopy with initial pair (g, ζ_0) is less than or equal to*

$$cn^5 N^2 d^3 \log_2 \mathcal{D}.$$

Hence, the average number of arithmetic operations is less than or equal to

$$cn^6 N^3 d^3 \log_2 d \log_2 \mathcal{D}.$$

Thus, the knowledge of an efficient initial pair yields an algorithm based on HD that finds a projective approximate zero of systems of polynomial equations in polynomial time, on the average. However, up to this moment such an initial pair is not explicitly known. In [BP08], the authors described a simple probabilistic method to find these pairs. To this end, the authors defined and analyzed the behavior of the quantity

$$A_\varepsilon(g, \zeta) = \mathrm{Prob}_{f\in\mathbb{S}_{(d)}}[\mu_{\mathrm{norm}}(f, g, \zeta) > \varepsilon^{-1}], \tag{1.2}$$

where $(g, \zeta) \in W$, and $\varepsilon > 0$ is some positive real number.

Definition 1.5. Let $\mathcal{G} \subseteq W$ be a subset with a probability measure. We say that $\mathcal{G}$ is a *strong questor set* for initial pairs if for every $\varepsilon > 0$,

$$\mathrm{E}_{\mathcal{G}}[A_\varepsilon] \leq 10^4 n^5 N^2 d^{3/2} \varepsilon^2,$$

where E means expectation.

The importance of Definition 1.5 relies on the following theorem that will be obtained as a consequence of a lemma from Probability Theory (see subsection 3.4.1).

Theorem 1.6. *Let $\mathcal{G} \subseteq \mathcal{H}_{(d)} \times \mathbb{P}_n(\mathbb{C})$ be a strong questor set for initial pairs. Then,*

$$\mathrm{E}_{((g,\zeta),f)\in\mathcal{G}\times\mathbb{S}_{(d)}}[\mu_{\mathrm{norm}}(f,g,\zeta)^{2-\beta}] \leq 2\cdot 10^4 n^5 N^2 d^{3/2}\log_2 \mathcal{D},$$

where $\beta = \frac{1}{\log_2 \mathcal{D}}$.

Fubini's Theorem and Markov's Inequality immediately yield the following corollary.

Corollary 1.7 (Existence of good initial pairs)**.** *Let $\sigma \in (0,1)$. Then, there exists a measurable set $\mathcal{C}_\sigma \subseteq \mathcal{G}$ such that*

$$\mathrm{Prob}_{(g,\zeta)\in\mathcal{G}}[(g,\zeta)\in\mathcal{C}_\sigma] \geq 1-\sigma, \tag{1.3}$$

and such that for every $(g,\zeta)\in\mathcal{C}_\sigma$,

$$\mathrm{E}_{f\in\mathbb{S}_{(d)}}[\mu_{\mathrm{norm}}(f,g,\zeta)^{2-\beta}] \leq \frac{2\cdot 10^4 n^5 N^2 d^{3/2}\log_2 \mathcal{D}}{\sigma}.$$

In particular, with probability of at least $4/5$, a randomly chosen pair $(g,\zeta)\in\mathcal{G}_{(d)}$ is an efficient initial pair.

This result means the following: if we are able to find a computationally tractable strong questor set, then we can probabilistically find an efficient initial pair (g,ζ). Hence, we obtain an algorithm that finds a projective zero of systems, on the average, in polynomial running time. A similar idea (i.e. the usage of a set that contains good initial points for iterative algorithms) has been recently developed in [HSS01], for the case of univariate polynomial solving.

In [BP08], the authors explicitly described a family of systems $\mathcal{G}_{(d)}$ associated with the list of degrees (d), such that one can easily choose at random a point $(g,\zeta)\in\mathcal{G}_{(d)}$. This rather technical family will be described with detail in Subsection 3.5. The main result in [BP08] may be written as follows (see [BP08, Proposition 36]):

Theorem 1.8. *The family $\mathcal{G}_{(d)}$ is a strong questor set for initial pairs. Namely, for every $\varepsilon > 0$:*

$$\mathrm{E}_{\mathcal{G}_{(d)}}[A_\varepsilon] \leq 10^4 n^5 N^2 d^{3/2}\varepsilon^2.$$

Hence, we have an explicit and efficient description of a strong questor set, for each list of degrees (d). This theorem simply means that we can easily compute an efficient initial pair, using a probabilistic method. Hence, from Corollary 1.7, we can compute a projective approximate zero of a system of equations, on the average, in polynomial running time. This is the projective version of our Main Theorem (Theorem 1.1).

1.2.2. *Affine solutions.* Usually, we are interested in the search of affine (not projective) approximate zeros of systems. We recall some properties of the affine Newton operator. Let $f\in\mathcal{H}_{(d)}$, $f:\mathbb{C}^n\longrightarrow\mathbb{C}^n$, and let $z\in\mathbb{C}^n$ be such that the differential matrix $df(z)\in\mathcal{M}_n(\mathbb{C})$ is of maximal rank. The affine Newton operator of f at z is

$$N_f(z) = z - (df(z))^{-1}f(z) \in \mathbb{C}^n.$$

We denote by $N_f^{(k)}(z) = N_f \circ \overset{k}{\cdots} \circ N_f(z)$ the affine point obtained after k applications of N_f to z, if it exists. Assume there exists a true zero $\zeta \in \mathbb{C}^n$ of f such that for every positive integer $k \in \mathbb{N}$, the vector $N_f^{(k)}(z)$ is defined and

$$\|N_f^{(k)}(z) - \zeta\| \leq \frac{1}{2^{2^k-1}}\|z - \zeta\|.$$

Then, we say that z is an affine approximate zero of f, with associated zero ζ.

As we have seen, the affine solutions of f are related to some of its projective solutions *via* the standard embedding φ_0 of equation (1.1). This suggests that, in order to find an affine approximate zero of f, we may first find a projective approximate zero $z \in \mathbb{P}_n(\mathbb{C})$ of f, with associated zero $\zeta \in \mathbb{P}_n(\mathbb{C})$, and then we may consider the affine point $\varphi_0^{-1}(z) \in \mathbb{C}^n$, if it exists. An initial drawback is that $\varphi_0^{-1}(z)$ may not be an affine approximate zero associated with $\varphi_0^{-1}(\zeta)$. In Proposition 4.5 below we will prove that that this process can be done if we are able to state an upper bound for the quantity $\|\varphi_0^{-1}(\zeta)\|$.

Hence, the probability distribution of the norm of the affine solutions of systems in $\mathcal{H}_{(d)}$ turns out to be an essential ingredient for the analysis of the complexity of finding affine approximate zeros. We prove the following result (Corollary 4.9 of Section 4).

Theorem 1.9. *Let $\delta > 0$. Then, the probability that a randomly chosen system $f \in \mathbb{S}_{(d)}$ has an affine solution $\zeta \in \mathbb{C}^n$ with $\|\zeta\| > \delta$ is at most*

$$\frac{\mathcal{D}\sqrt{\pi n}}{\delta}.$$

Now we describe an algorithm that finds affine approximate zeros of systems. For every $(g, \zeta) \in \mathcal{H}_{(d)} \times \mathbb{P}_n(\mathbb{C})$, we define the following procedure (which will be called Affine Homotopic Deformation, AHD for short).

ALGORITHM: AHD

Input: $f \in \mathbb{S}_{(d)}$.

PART ONE

Apply Shub and Smale's HD to the segment $[g, f]$, starting at the initial pair (g, ζ). Let $z \in \mathbb{P}_n(\mathbb{C})$ be the output of this procedure.

PART TWO

For i from 1 to ∞ do

- Let $z^i = \widehat{N}_f^{(i)}(z) \in \mathbb{P}_n(\mathbb{C})$, where $\widehat{N}_f$ is the projective Newton operator associated with f.
- Check (using Smale's α-Theorem [Ded06]) if the affine point $\varphi_0^{-1}(z^i) \in \mathbb{C}^n$ is an affine approximate zero of f. In this case, halt.

Output: an affine approximate zero of f.

The proof of the following result relies on Theorem 1.9.

Theorem 1.10. *Let $(g,\zeta) \in W$ be an efficient initial pair. Then, the average number of arithmetic operations performed by* AHD *is less than or equal to*

$$cn^6N^3d^3 \log_2 d(\log_2 \mathcal{D}).$$

From Theorem 1.10, the knowledge of an efficient initial pair (g,ζ) yields an algorithm that finds an affine approximate zero of systems of polynomial equations, on the average, in polynomial running time. Finally, Corollary 1.7 and Theorem 1.8 above allow us to find (g,ζ), using a simple probabilistic procedure. This is the affine version of our Main Theorem (Theorem 1.1).

This paper is structured as follows. In Section 2 we describe in some detail the notation, concepts and previous results related to our main results. In Section 3 we recall the main results which lead to the resolution of the projective case and we describe in detail the construction of the strong questor set $\mathcal{G}_{(d)}$. We will also prove Theorem 1.6. Section 4 is devoted to proving Theorems 1.9 and 1.10.

1.3. **Appendix to the Introduction.** At this point we must clarify what the terms uniform algorithm and **Average ZPP** mean. For a precise discussion we direct the reader to [BCSS98] and the references therein. Non–uniform algorithms have been known in theoretical computer science literature for a long time. A classical definition of a non–uniform complexity class in the boolean setting is Nick Pippenger's NC classes. In the continuous setting, we have the notions of $\mathbf{NC}_{\mathbb{R}}$ classes in [CMP95, MP98] or [BCSS98, Chpt. 18] and the references therein. Roughly speaking, a non–uniform algorithm is a sequence of circuits $\mathcal{C}_N$ such that for every N, the circuit solves instances of length N. Non–uniform upper complexity bounds are sometimes useful (and give hints) to achieve lower complexity bounds and uniform efficient algorithms. However, they stay at a theoretical level, since non–uniform algorithms are not implementable. Non–uniform lower complexity bounds are also essential to understand the troubles with complexity in many limit problems.

On the other hand, there are several definitions of what a uniform algorithm is. In the boolean setting, the Turing machine model of complexity seems to be the natural one. As S. Cook observed in [Coo85], NC classes also accept different forms of "uniformity". In the continuous setting the work by Blum, Shub and Smale [BSS89] established a context that defines what a uniform algorithm should be. Roughly speaking, a uniform algorithm is an algorithm that has a finite machine that performs the required computations. In terms of circuit (non–uniform) complexity classes, the machine contains a procedure that implicitly generates the sequence of circuits (cf. for example [CMP92]).

Another simplified description is that a uniform algorithm is something that one may implement as a single program in any standard programming language, whereas non–uniform algorithms require an implementation for each input length (and hence an infinite number of "programs").

Uniform algorithms can be deterministic or non-deterministic. This distinction leads to the classical $\mathbf{P} \neq \mathbf{NP}$ question of Cook's Conjecture in the boolean setting and the two problems concerning Hilbert Nullstellensatz (complex) and 4−Feasible (real) in the continuous setting (cf. [Koi96] or [BCSS98] and the references therein).

A particularly useful class of non-deterministic, uniform algorithms are probabilistic algorithms. A probabilistic algorithm is a uniform algorithm that starts its computations by randomly guessing an instance from an appropriate class of data

associated with the given input length. When the probability that the guessed instance leads to an error is bounded (usually by a constant smaller than 1/2), they are called "bounded error probability algorithms". If that probability is equal to 0 they are called "zero error probability algorithms".

Note that the running time (number of arithmetic operations) of a probabilistic algorithm may depend on the initially-guessed instance. Thus, we must fix a way of measuring the time of the algorithm on a given input. There are two widely accepted ways of doing so. The time of the algorithm for a given input is defined as

- the worst-case time among all choices of initial guesses, or
- the average of the time among all choices of initial guesses.

In this paper, we use the second of these two options. Namely, for every polynomial system $f \in \mathbb{S}_{(d)}$, we have

$$t(f) = \mathrm{E}_{(g,\zeta)\in\mathcal{G}_{(d)}}[\sharp \text{ Arithmetic ops. of HD with input } f \text{ and guess } (g,\zeta)].$$

With this terminology, when the running time of a zero error probability algorithm is bounded by a polynomial in the input length, the problem is said to be in the class **ZPP** (**Z**ero error probability, **P**robabilistic, **P**olynomial time). For bounded error probability algorithms (measuring the running time as the worst case instead of average among guesses), the term is **BPP**. Note that, in the Turing setting, **ZPP** $\subseteq$ **BPP** (cf. for example [DK00, Chapter 8]). As an example, the primality tests in [SS77, SS78] or [Mil76, Rab80] are **BPP**, but not **ZPP**.

Usually the complexity classes as **ZPP** or **BPP** are used for decisional problems (problems with YES/NO answers). It is common to use the term **Las Vegas** for the non-decisional version of **ZPP**.

The algorithm described in the Introduction of this paper can be written as

---

Input: a polynomial system f (normalized such that its norm equals 1).

Let (d) be the list of degrees of the polynomials in f. Guess at random $(g,\zeta) \in \mathcal{G}_{(d)}$.

Compute the polygonal path of the Homotopy Deformation in $[g,f]$ (described in [SS94]).

Output: a projective approximate zero of f (if desired, apply AHD above to obtain an affine approximate zero of f).

Note that for some input systems, the algorithm may run forever, and no output will be obtained.

Then, Theorem 1.6 (using Theorem 3.2 below) reads: the average running time of this algorithm is bounded by a polynomial in the size of the input. More specifically, there is a constant c such that for every multiindex (d),

$$\mathrm{E}_{f\in\mathbb{S}_{(d)}}[t(f)] \leq N^c.$$

Moreover, note that Theorem 1.6 also implies that for every input system $f \in \mathbb{S}_{(d)}$ (outside of some zero measure set), the probability that a random choice of $(g, \zeta) \in \mathcal{G}_{(d)}$ leads to an approximate zero of f is exactly equal to 1.

Thus, our algorithm differs from a usual **ZPP** (or Las Vegas) algorithm in that the polynomial running time is not the worst-case but is an average case, and zero-measure sets can be omitted. In the language of Theoretical Computer Science, the running time $t(N)$ is not defined as the maximum of the running times for inputs of length N, but as the average of these running times. Hence, we say that our algorithm is **Average ZPP**, or **Average Las Vegas**. It is not deterministic (as in probabilistic primality tests cited above), although it is certainly uniform. Thus, it gives an affirmative answer to Smale's 17th problem.

The result as stated asks the algorithm to be able to choose random real numbers with the normal distribution (which leads to random choice of points in spheres).

2. Notions and notation

Recall that we have fixed a positive integer number $n \in \mathbb{N}$ that will be the number of equations in our target systems. For every positive integer number $l \in \mathbb{N}$, we fix some ordering in the set

$$\mathcal{E}_l = \{\alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1} : \alpha_0 + \cdots + \alpha_n = l\}.$$

Then, we define $H_l = \prod_{\alpha \in \mathcal{E}_l} \mathbb{C}$. Observe that the vector space H_l may be understood in two different ways:

- as the space of homogeneous polynomial mappings $\mathbb{C}^{n+1} \longrightarrow \mathbb{C}$ of degree l. In fact, for $(a_\alpha)_{\alpha \in \mathcal{E}_l} \in H_l$, we may consider the mapping
$$\begin{array}{ccc} \mathbb{C}^{n+1} & \longrightarrow & \mathbb{C}, \\ (x_0, \ldots, x_n) & \mapsto & \sum_{\alpha=(\alpha_0,\ldots,\alpha_n)\in\mathcal{E}_l} a_\alpha x_0^{\alpha_0} \cdots x_n^{\alpha_n}. \end{array}$$
- as the space of polynomial mappings $\mathbb{C}^n \longrightarrow \mathbb{C}$ of degree at most l. In fact, for $(a_\alpha)_{\alpha \in \mathcal{E}_l} \in H_l$, we may consider the mapping
$$\begin{array}{ccc} \mathbb{C}^{n} & \longrightarrow & \mathbb{C}, \\ (x_1, \ldots, x_n) & \mapsto & \sum_{\alpha=(\alpha_0,\ldots,\alpha_n)\in\mathcal{E}_l} a_\alpha x_1^{\alpha_1} \cdots x_n^{\alpha_n}. \end{array}$$

Now, let $(d) = (d_1, \ldots, d_n) \in \mathbb{N}^n$ be a list of positive integers. We consider the vector space of systems

$$\mathcal{H}_{(d)} = \prod_{i=1}^{n} H_{d_i},$$

as $\mathcal{H}_{(d)}$ may be understood as the set of homogeneous polynomial mappings $\mathbb{C}^{n+1} \longrightarrow \mathbb{C}^n$ of respective degrees given by the list $(d_1, \ldots, d_n)$, or as the set of (not necessarily homogeneous) polynomial mappings $\mathbb{C}^n \longrightarrow \mathbb{C}^n$ of degrees bounded by the same list.

We may denote by $f, g, \ldots$ the elements in $\mathcal{H}_{(d)}$. We denote by $N+1$ the complex dimension of $\mathcal{H}_{(d)}$. Namely,

$$N + 1 = \dim(\mathcal{H}_{(d)}) = \dim(H_{d_1}) + \cdots + \dim(H_{d_n}) = \binom{n + d_1}{n} + \cdots + \binom{n + d_n}{n}.$$

An element $f = [f_1, \ldots, f_n] \in \mathcal{H}_{(d)}$ is usually given as the list of coefficients of $f_1, \ldots, f_n$. Hence, $N + 1$ may be seen as the size of the input. As said in the Introduction, we denote by $d = \max\{d_i : 1 \leq i \leq n\}$ the maximum of the degrees,

and by $\mathcal{D} = d_1 \cdots d_n \leq d^n$ the Bézout number associated with the list (d). From now on, we assume that $d \geq 2$.

We consider $\mathcal{H}_{(d)}$ equipped with the Bombieri-Weyl Hermitian product $\langle \cdot, \cdot \rangle_\Delta$. We denote by Δ the diagonal matrix given by

$$\Delta = Diag\left(\binom{d_i}{\alpha}^{-1/2} \right)_{\substack{1 \leq i \leq m \\ \alpha \in \mathcal{E}_{d_i}}},$$

where $\binom{d_i}{\alpha}$ is the multinomial coefficient. Namely,

$$\binom{d_i}{\alpha} = \frac{d_i!}{\alpha_0! \cdots \alpha_n!} \in \mathbb{N}.$$

Then,

$$\langle f, g \rangle_\Delta = \langle \Delta f, \Delta g \rangle_2, \ \ \forall f, g \in \mathcal{H}_{(d)}^m,$$

where $\langle \cdot, \cdot \rangle$ is the usual Hermitian product in $\mathcal{H}_{(d)}$. The main property of the Bombieri-Weyl Hermitian product is its unitary invariance (see for example [BCSS98, Teor. 1, page 218]): for every unitary matrix $U \in \mathcal{U}_{n+1}$ and for every pair of elements $f, g \in \mathcal{H}_{(d)}$, we have

$$\langle f, g \rangle_\Delta = \langle f \circ U, g \circ U \rangle_\Delta. \tag{2.1}$$

For any element $f \in \mathcal{H}_{(d)}$, we may consider the derivative of f as a linear map. This concept changes if we see f as a map with domain in $\mathbb{C}^{n+1}$ or $\mathbb{C}^n$. Hence, if we consider $f : \mathbb{C}^{n+1} \longrightarrow \mathbb{C}^n$ as a homogeneous polynomial mapping, then we denote by $\widehat{d}f(x)$ the derivative of f at x, where $x \in \mathbb{C}^{n+1}$ is a point. On the other hand, if we consider $f : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ as a (possibly not homogeneous) polynomial mapping, then we denote by $df(x)$ the derivative of f at x, where $x \in \mathbb{C}^n$ is a point.

For every $\zeta \in \mathbb{P}_n(\mathbb{C})$, let V_ζ be the set of systems that have a projective zero at ζ. Namely,

$$V_\zeta = \{ f \in \mathcal{H}_{(d)} : \zeta \in V_{\mathbb{P}}(f) \} \subseteq \mathcal{H}_{(d)}.$$

Let $e_0 = (1 : 0 : \cdots : 0) \in \mathbb{P}_n(\mathbb{C})$ be a point that we can fix as a "north pole". We will use the following subspaces of $\mathcal{H}_{(d)}$:

$$L_{e_0} = \{ g = [g_1, \ldots, g_n] \in \mathcal{H}_{(d)} : g_i = X_0^{d_i - 1} \sum_{j=1}^{n} a_{ij} X_j, 1 \leq i \leq n \} \subseteq V_{e_0},$$

$$L_{e_0}^\perp = \{ g \in V_{e_0} : \langle g, f \rangle_\Delta = 0, \ \forall f \in L_{e_0} \} \subseteq V_{e_0}.$$

Namely, L_{e_0} is the set of homogeneous systems that have a zero at e_0 and are linear in the variables $X_1, \ldots, X_n$. The set $L_{e_0}^\perp$ may be seen as the set of homogeneous systems of order 2 at e_0.

For an element $f \in \mathcal{H}_{(d)}$ and a point $x \in \mathbb{C}^{n+1}$, let $T_x f$ be the restriction of the tangent mapping $\widehat{d}f(x)$ to the Hermitian complement of x. Namely,

$$T_x f = (\widehat{d}f(x)) \mid_{x^\perp} .$$

We will also use the mapping

$$\begin{array}{rccc} \psi_{e_0} : & L_{e_0} & \longrightarrow & \mathcal{M}_n(\mathbb{C}), \\ & g & \mapsto & \Delta(d)^{-1/2} T_{e_0} g, \end{array}$$

where $\Delta(d)^{-1/2}$ is given by the formula

$$\Delta(d)^{-1/2} = Diag(d_1^{-1/2}, \ldots, d_n^{-1/2}) \in \mathcal{M}_n(\mathbb{R}).$$

Observe that ψ_{e_0} is a linear isometry (see for example [BCSS98, Lemma 17, page 235]).

We consider $\mathbb{P}_n(\mathbb{C})$ equipped with the canonical Riemannian metric. The Riemannian distance between two points $x, y \in \mathbb{P}_n(\mathbb{C})$ will be denoted by $d_R(x, y)$. As in [BCSS98], we will frequently use the so-called "tangent distance" $d_T(x, y)$ defined as

$$d_T(x, y) = \tan d_R(x, y) = \sqrt{\frac{\|x\|^2\|y\|^2}{|\langle x, y\rangle|^2} - 1},$$

where some affine representatives of x and y have been chosen. The tangent distance d_T is not quite a distance, as it does not satisfy the triangle inequality. But if x and y are near (in terms of d_R), then $d_T(x, y)$ is a good approximation of $d_R(x, y)$.

2.1. **Some geometric integration theory.** The Coarea Formula is a classic integral formula which generalizes Fubini's Theorem. The most general version we know is Federer's Coarea Formula (cf. [Fed69]), but for our purposes a smooth version as used in [BCSS98, SS93b] or [How] suffices.

Definition 2.1. Let X and Y be Riemannian manifolds, and let $F : X \longrightarrow Y$ be a C^1 surjective map. Let $k = \dim(Y)$ be the real dimension of Y. For every point $x \in X$ such that the differential $DF(x)$ is surjective, let $v_1^x, \ldots, v_k^x$ be an orthonormal basis of $Ker(DF(x))^\perp$. Then, we define the Normal Jacobian of F at x, NJ_xF, as the volume in the tangent space $T_{F(x)}Y$ of the parallelepiped spanned by $DF(x)(v_1^x), \ldots, DF(x)(v_k^x)$. In the case that $DF(x)$ is not surjective, we define $NJ_xF = 0$.

Theorem 2.2 (Coarea Formula)**.** *Let X, Y be two Riemannian manifolds of respective dimensions $k_1 \geq k_2$. Let $F : X \longrightarrow Y$ be a C^1 surjective map, such that $dF(x)$ is surjective for almost all $x \in X$. Let $\psi : X \longrightarrow \mathbb{R}$ be an integrable mapping. Then,*

$$\int_X \psi \; dX = \int_{y \in Y} \int_{x \in F^{-1}(y)} \psi(x) \frac{1}{NJ_xF} \; d(F^{-1}(y)) \; dY, \tag{2.2}$$

where NJ_xF is the normal jacobian of F at x.

Observe that the integral on the right-hand side of equation (2.2) may be interpreted as follows: from Sard's Theorem, for every $y \in Y$ except for a zero measure set, y is a regular value of F. Then, $F^{-1}(y)$ is a differentiable manifold of dimension $k_1 - k_2$, and it inherits from X a structure of Riemannian manifold. Thus, it makes sense to integrate functions on $F^{-1}(y)$.

The following proposition is easy to prove (see for example [BCSS98]).

Proposition 2.3. *Let X, Y be two Riemannian manifolds, and let $F : X \longrightarrow Y$ be a C^1 map. Let $x_1, x_2 \in X$ be two points. Assume that there exist isometries $\varphi_X : X \longrightarrow X$ and $\varphi_Y : Y \longrightarrow Y$ such that $\varphi_X(x_1) = x_2$ and*

$$F \circ \varphi_X = \varphi_Y \circ F.$$

Then,

$$NJ_{x_1}F = NJ_{x_2}F.$$

Moreover, if F is bijective and $G = F^{-1} : Y \longrightarrow X$, then

$$NJ_xF = \frac{1}{NJ_{F(x)}G}, \quad x \in X.$$

3. Smale's 17th problem: Projective solutions

3.1. **The projective Newton operator.** The normalized condition number of [SS93a] is defined as follows. For $f \in \mathcal{H}_{(d)}$ and $z \in \mathbb{P}_n(\mathbb{C})$,

$$\mu_{\text{norm}}(f,z) = \|f\|_\Delta \|(T_z f)^{-1} Diag(\|z\|^{d_i-1} d_i^{1_2})\|_2, \tag{3.1}$$

and $\mu_{\text{norm}}(f,z) = +\infty$ if $T_z f$ is not onto. Also, $\mu_{\text{norm}}(f,z) = \mu_{\text{norm}}(\lambda f, \eta z)$ for every $\lambda, \eta \in \mathbb{C} \setminus \{0\}$ (i.e. μ_{norm} depends only on the projective class of f and z). The following set plays a crucial role in the study of the projective Newton operator:

$$\Sigma' = \{(f,\zeta) \in W : \det(T_z f) = 0\},$$

where W is the incidence variety defined in the Introduction. Observe that Σ' (usually called the discriminant variety) consists of the set of pairs $(f,\zeta) \in W$ such that ζ is a singular solution of f. Let $\pi_1 : W \longrightarrow \mathcal{H}_{(d)} \setminus \{0\}$ be the projection onto the first coordinate. Then, the set of critical points of π_1 is exactly Σ' (cf. [BCSS98], [SS93b]).

3.2. **NHD in the space of systems.** Let $f, g \in \mathbb{S}_{(d)}$ be two polynomial systems, $f \neq \pm g$, and let $\zeta \in \mathbb{P}_n(\mathbb{C})$ be a solution of g. Let $[g,f]$ be the segment joining g and f. Namely,

$$[g,f] = \{tf + (1-t)g, t \in [0,1]\} \subseteq \mathcal{H}_{(d)}.$$

We also use the notation (g,h) to represent the corresponding "open interval". Namely, $[g,f] = \{tf + (1-t)g, t \in (0,1)\} \subseteq \mathcal{H}_{(d)}$. The following result is a consequence of the Implicit Function Theorem (cf. for example [BCSS98]).

Proposition 3.1. *Let $(f,g,\zeta_0) \in \mathcal{H}_{(d)} \times W$ be such that $\|f\|_\Delta = \|g\|_\Delta = 1$, $f \neq \pm g$. Let $\Gamma(f,g,\zeta_0)$ be the connected component of $\pi_1^{-1}([g,f]) \subseteq W$ that contains the point (g,ζ_0). If $\Gamma(f,g,\zeta_0) \cap \Sigma' = \emptyset$, then it is a smooth curve. Moreover, for each $h \in [g,f]$, there exists a unique solution $\zeta_0' \in V_{\mathbb{P}}(h)$ of h such that $(h,\zeta_0') \in \Gamma(f,g,\zeta_0)$.*

Proof. Let $p = \pi_1 |_{\Gamma(f,g,\zeta_0)} : \Gamma(f,g,\zeta_0) \longrightarrow [g,f]$ be the restriction of π_1 to $\Gamma(f,g,\zeta_0)$. Note that both $\Gamma(f,g,\zeta_0)$ and $[g,f]$ are Haursdoff, compact and path-wise connected. Thus, the image of p is a closed interval inside $[g,f]$, and the Implicit Function Theorem applied to π_1 implies that p is surjective. Now, p is obviously proper, and we conclude that it is a covering map (see for example [Ho75]) and a global homeomorphism. The proposition immediately follows. □

In the hypothesis of Proposition 3.1, let $\zeta = (\pi_1 |_{\Gamma(f,g,\zeta_0)})^{-1}(f)$ be the (unique) solution of f that belongs to $\Gamma(f,g,\zeta_0)$. The following result is one of the main outcomes of the series of papers due to Shub and Smale (cf. [SS94, Prop. 7.2]).

Theorem 3.2 (Shub and Smale)**.** *Let $0 \leq \nu \leq 1$ and $(g,\zeta_0) \in W$. Let $f \in \mathbb{S}_{(d)} \setminus \{\pm g\}$. Then,*

$$k \leq c(\mu_{\text{norm}}(f,g,\zeta_0))^{2-\nu} \mathcal{D}^{2\nu} d^{3/2}$$

steps of projective Newton operator, starting from (g, ζ_0), are sufficient to produce an approximate zero z of f. Moreover, the (true) projective zero associated with z is ζ and

$$d_T(z, \zeta)) \leq \frac{3-\sqrt{7}}{2d^{3/2}\mu_{\mathrm{norm}}(f,\zeta)}. \tag{3.2}$$

3.3. **Proof of Theorem 1.4.** Let (g, ζ_0) be an efficient initial pair, and let $\beta = \frac{1}{\log_2 \mathcal{D}}$. From Definition 1.3,

$$\mathrm{E}_{f\in\mathbb{S}_{(d)}}[\mu_{\mathrm{norm}}(f,g,\zeta)^{2-\beta}] \leq 10^5 n^5 N^2 d^{3/2} \log_2 \mathcal{D}.$$

From Theorem 3.2, taking $\nu = \beta$ so that $\mathcal{D}^{2\nu}$ is a universal constant, we conclude that the average number of steps of the HD with initial pair (g, ζ_0) is at most

$$cd^{3/2}\mathrm{E}_{f\in\mathbb{S}_{(d)}}[\mu_{\mathrm{norm}}(f,g,\zeta_0)^{2-\beta}] \leq cn^5 N^2 d^3 \log_2 \mathcal{D},$$

as wanted.

3.4. **Further properties of strong questor sets.** In this subsection we will prove Theorem 1.6. We will use a well-known fact from probability theory (for a proof, see for example [BP07b, Lemma 43]).

Lemma 3.3. *Let ξ be a positive real valued random variable such that for every positive real number $t > 0$*

$$\mathrm{Prob}[\xi > t] < ct^{-\alpha},$$

where Prob$[\cdot]$ holds for Probability, and $c > 0, \alpha > 1$ are some positive constants. Then,

$$\mathrm{E}[\xi] \leq c^{\frac{1}{\alpha}} \ \frac{\alpha}{\alpha - 1}.$$

Lemma 3.4. *Let X, Y be two probability spaces and let $X \times Y$ be endowed with the product probability measure. Let $\xi : X \times Y \longrightarrow [0, \infty)$ be a positive real random variable and let*

$$\begin{array}{rccc} T: & X \times (0,\infty) & \longrightarrow & [0,1], \\ & (x,\varepsilon) & \mapsto & \mathrm{Prob}_{y\in Y}[\xi(x,y) > \varepsilon^{-1}]. \end{array}$$

Assume that for every $\varepsilon > 0$ we have

$$\mathrm{E}_{x\in X}[T(x,\varepsilon)] \leq K\varepsilon^2,$$

for some $K \geq 1$ independent of ε. Then, for every $0 < \beta < 2$,

$$\mathrm{E}_{X\times Y}[\xi^{2-\beta}] \leq \frac{2K}{\beta}.$$

Proof. Let $t > 0$. Fubini's Theorem yields

$$\begin{aligned} \mathrm{Prob}_{(x,y)\in X\times Y}[\xi(x,y)^{2-\beta} > t] &= \int_{x\in X} \mathrm{Prob}_{y\in Y}[\xi(x,y)^{2-\beta} > t] dX \\ &= \int_{x\in X} T(x, t^{\frac{-1}{2-\beta}}) dX \leq K t^{\frac{-2}{2-\beta}}. \end{aligned}$$

From Lemma 3.3, we conclude that

$$\mathrm{E}_{(x,y)\in X\times Y}[\xi(x,y)^{2-\beta}] \leq K^{\frac{2-\beta}{2}} \frac{2}{\beta} \leq \frac{2K}{\beta}. \qquad \square$$

3.4.1. *Proof of Theorem* 1.6. Apply Lemma 3.4 with $X = \mathcal{G}$, $Y = \mathbb{S}_{(d)}$ and $\xi((g,\zeta),f) = \mu_{\text{norm}}(f,g,\zeta)$. □

3.5. **The family of good initial pairs.** Now we recall the description of the strong questor set for initial pairs found in [BP08]. Let

$$Y = [0,1] \times B^1(L_{e_0}^{\perp}) \times B^1(\mathcal{M}_{n\times(n+1)}(\mathbb{C})) \subseteq \mathbb{R} \times \mathbb{C}^{N+1},$$

where $B^1(L_{e_0}^{\perp})$ is the closed ball of radius one in $L_{e_0}^{\perp}$ for the canonical Hermitian metric and $B^1(\mathcal{M}_{n\times(n+1)}(\mathbb{C}))$ is the closed ball of radius one in the space of $n \times (n+1)$ complex matrices for the standard Frobenius norm. We assume that Y is endowed with the product of the respective Riemannian structures and the corresponding measures and probabilities.

Let $\tau \in \mathbb{R}$ be the real number given by

$$\tau = \sqrt{\frac{n^2+n}{N}},$$

and let us fix any (a.e. continuous) mapping $\phi : \mathcal{M}_{n\times(n+1)}(\mathbb{C}) \longrightarrow \mathcal{U}_{n+1}$ such that for every matrix $M \in \mathcal{M}_{n\times(n+1)}(\mathbb{C})$ of maximal rank, ϕ associates a unitary matrix $\phi(M) \in \mathcal{U}_{n+1}$ satisfying $M\phi(M)e_0 = 0$. In other words, $\phi(M)$ transforms e_0 into a vector in the kernel $Ker(M)$ of M. Our statements below are independent of the chosen mapping ϕ that satisfies this property. There are several different ways to define this mapping ϕ, using simple procedures from Linear Algebra. Observe that the first column of the matrix $M\phi(M)$ is zero. Namely,

$$M\phi(M) = (0\ C),$$

where $C \in \mathcal{M}_n(\mathbb{C})$ is a square matrix. As no confusion is possible, we also denote by $M\phi(M)$ the matrix C.

Then, we consider the associated system

$$\psi_{e_0}^{-1}(M\phi(M)) \in L_{e_0} \subseteq \mathcal{H}_{(d)},$$

where ψ_{e_0} is as defined in Section 2.

Then, we define a mapping $G_{(d)} : Y \longrightarrow V_{e_0}$ as follows. For every $(t,h,M) \in Y$, let

$$G_{(d)}(t,h,M) = \left(1 - \tau^2 t^{\frac{1}{n^2+n}}\right)^{1/2} \frac{\Delta^{-1}h}{||h||_2} + \tau t^{\frac{1}{2n^2+2n}} \psi_{e_0}^{-1}\left(\frac{M\phi(M)}{||M||_F}\right) \in V_{e_0}.$$

Finally, we define the set $\mathcal{G}_{(d)} \subseteq W$ as

$$\mathcal{G}_{(d)} = \{(g,e_0) \in W : \exists y \in Y, G_{(d)}(y) = g\}.$$

Namely, $\mathcal{G}_{(d)}$ is the set of pairs $(g,\zeta) \in \mathcal{H}_{(d)} \times \mathbb{P}_n(\mathbb{C})$ such that $\zeta = e_0$ and g is in the image of Y under $G_{(d)}$. The probability distribution in $\mathcal{G}_{(d)}$ is the one inherited from Y by $G_{(d)}$. Namely, in order to choose a random point in $\mathcal{G}_{(d)}$ we choose a random point $y \in Y$ and we consider the pair $(G_{(d)}(y), e_0) \in \mathcal{G}_{(d)}$.

As said in the Introduction, the main result in [BP08] is Theorem 1.8, and in that paper the authors use it to describe an algorithm that computes a projective approximate zero of systems $f \in \mathcal{H}_{(d)}$ in polynomial time, assuming a small probability of failure.

4. Smale's 17th problem: Affine solutions

Now we turn to the case of affine solutions of systems of polynomial equations. Our space of inputs is again $\mathcal{H}_{(d)}$ (or the sphere $\mathbb{S}_{(d)}$), and we look for affine approximate zeros (points in $\mathbb{C}^n$) of our systems.

4.1. **From projective to affine approximate zeros.** In this section, we will show how to obtain an affine approximate zero of f from the information contained in a projective approximate zero of f. We start with the following result, which is a consequence of the γ and μ theories of Shub and Smale (cf. for example [Sma86, SS93a, Ded06]).

Proposition 4.1. *With the notation above, let $\zeta = (\zeta_0 : \cdots : \zeta_n) \in \mathbb{P}_n(\mathbb{C})$ be a projective zero of f, such that $\zeta_0 \neq 0$, and let $z = (z_0 : \cdots : z_n) \in \mathbb{P}_n(\mathbb{C})$ be a projective point such that $z_0 \neq 0$. Assume that*

$$\|\varphi_0^{-1}(z) - \varphi_0^{-1}(\zeta)\| \leq \frac{3-\sqrt{7}}{d^{3/2}\mu_{\mathrm{norm}}(f,\zeta)}.$$

Then, $\varphi_0^{-1}(z)$ is an affine approximate zero of f, with associated zero $\varphi_0^{-1}(\zeta)$.

Proof. From [Sma86, Th. C], it suffices to prove that

$$\|\varphi_0^{-1}(z) - \varphi_0^{-1}(\zeta)\| \leq \frac{3-\sqrt{7}}{2\gamma(f,\varphi_0^{-1}(\zeta))},$$

where $\gamma(f,\varphi_0^{-1}(\zeta))$ is the quantity defined in [Sma86]. Now, from [SS93a, Prop. 3 of Section I.3 and Th. 2 of Section I.4] we know that

$$\gamma(f,\varphi_0^{-1}(\zeta)) \leq \frac{d^{3/2}}{2}\mu_{\mathrm{norm}}(f,\zeta), \tag{4.1}$$

and the proposition follows. Note that the notation in [SS93b] is slightly different, as μ_{norm} is denoted μ_{proj}. See also [BP07a, Prop. 3.4] for a proof of equation (4.1). □

Lemma 4.2. *Let $v, w \in \mathbb{C}^n$ be two complex vectors. Then,*

$$\|v-w\|^2 \leq (1+\|v\|^2)(1+\|w\|^2) - |1+\langle v,w\rangle|^2.$$

Proof. Let R and I be the respective real and imaginary parts of $\langle v,w\rangle$. On one hand, we have

$$\|v-w\|^2 = \langle v-w, v-w\rangle = \|v\|^2 + \|w\|^2 - 2R.$$

On the other hand,

$$\begin{aligned}(1+\|v\|^2)(1+\|w\|^2) - |1+\langle v,w\rangle|^2 &= (1+\|v\|^2)(1+\|w\|^2) - |1+R+\sqrt{-1}I|^2\\ &= (1+\|v\|^2)(1+\|w\|^2) - (1+R)^2 - I^2\\ &= 1+\|v\|^2+\|w\|^2+\|v\|^2\ \|w\|^2 - 1 - 2R - (R^2+I^2)\\ &= \|v\|^2+\|w\|^2-2R+\|v\|^2\ \|w\|^2 - |\langle v,w\rangle|^2.\end{aligned}$$

Hence,

$$(1+\|v\|^2)(1+\|w\|^2)-|1+\langle v,w\rangle|^2=\|v-w\|^2+\|v\|^2\ \|w\|^2-|\langle v,w\rangle|^2\geq\|v-w\|^2,$$

as wanted. □

Lemma 4.3. *Let $\zeta=(\zeta_0:\cdots:\zeta_n), z=(z_0:\cdots:z_n)\in\mathbb{P}_n(\mathbb{C})$ be two projective points, such that $\zeta_0, z_0\neq 0$. Then,*

$$\|\varphi_0^{-1}(z)-\varphi_0^{-1}(\zeta)\|\leq(1+\|\varphi_0^{-1}(\zeta)\|\ \|\varphi_0^{-1}(z)\|)\ d_T(z,\zeta).$$

Proof. We denote $v=\varphi_0^{-1}(\zeta)$, $w=\varphi_0^{-1}(z)$. Then, the point $(1,v)\in\mathbb{C}^{n+1}$ is a representative of ζ and the point $(1,w)\in\mathbb{C}^{n+1}$ is a representative of z. Hence,

$$\begin{aligned}\frac{\|v-w\|^2}{d_T(z,\zeta)^2}&=\frac{\|v-w\|^2}{\frac{\|(1,v)\|^2\|(1,w)\|^2}{|\langle(1,v),(1,w)\rangle|^2}-1}\\&=\frac{\|v-w\|^2}{\|(1,v)\|^2\|(1,w)\|^2-|\langle(1,v),(1,w)\rangle|^2}|\langle(1,v),(1,w)\rangle|^2\\&=\frac{\|v-w\|^2}{(1+\|v\|^2)(1+\|w\|^2)-|1+\langle v,w\rangle|^2}|1+\langle v,w\rangle|^2.\end{aligned}$$

From Lemma 4.2, we conclude that

$$\frac{\|v-w\|^2}{d_T(z,\zeta)^2}\leq|1+\langle v,w\rangle|^2.$$

Thus,

$$\frac{\|v-w\|}{d_T(z,\zeta)}\leq|1+\langle v,w\rangle|\leq 1+|\langle v,w\rangle|\leq 1+\|v\|\|w\|,$$

as wanted. □

Lemma 4.4. *Let $\zeta=(\zeta_0:\cdots:\zeta_n)$ and $z=(z_0:\cdots:z_n)$ be two projective points such that $\zeta_0\neq 0$, and let $\varepsilon\geq 0$ be such that $d_T(z,\zeta)\leq\varepsilon$. Moreover, assume that*

$$d_R(\zeta,e_0)+\varepsilon<\frac{\pi}{2}. \tag{4.2}$$

Then, $z_0\neq 0$ and

$$\|\varphi_0^{-1}(z)-\varphi_0^{-1}(\zeta)\|\leq\frac{1+\|\varphi_0^{-1}(\zeta)\|^2}{1-\|\varphi_0^{-1}(\zeta)\|\varepsilon}\ \varepsilon.$$

Proof. Let $\beta_z=d_R(z,e_0)$ and $\beta_\zeta=d_R(\zeta,e_0)$ be the respective Riemannian distances from z,ζ to e_0. Observe that

$$\beta_z\leq\beta_\zeta+d_R(z,\zeta)\leq\beta_z+d_T(z,\zeta)\leq\beta_z+\varepsilon.$$

Moreover,

$$\|\varphi_0^{-1}(z)\|=\tan\beta_z,\quad\|\varphi_0^{-1}(\zeta)\|=\tan\beta_\zeta.$$

Now, tan is an increasing function in $[0,\frac{\pi}{2})$. From inequality (4.2), we conclude

$$\|\varphi_0^{-1}(z)\|\leq\tan(\beta_\zeta+d_R(z,\zeta))=\frac{\tan\beta_\zeta+d_T(z,\zeta)}{1-\tan\beta_\zeta d_T(z,\zeta)}\leq\frac{\|\varphi_0^{-1}(\zeta)\|+\varepsilon}{1-\|\varphi_0^{-1}(\zeta)\|\varepsilon}.$$

From Lemma 4.3, we conclude that

$$\|\varphi_0^{-1}(z)-\varphi_0^{-1}(\zeta)\|\leq\left(1+\|\varphi_0^{-1}(\zeta)\|\ \frac{\|\varphi_0^{-1}(\zeta)\|+\varepsilon}{1-\|\varphi_0^{-1}(\zeta)\|\varepsilon}\right)\ \varepsilon,$$

and the lemma follows. □

The following result yields a method to obtain an affine approximate zero of f from a projective approximate zero of f, if some properties are satisfied.

Proposition 4.5. *Let $f \in \mathcal{H}_{(d)}$. Let $\zeta \in V_{\mathbb{P}}(f)$ be a projective zero of f, $\zeta_0 \neq 0$, and let $z = (z_0, \ldots, z_n) \in \mathbb{P}_n(\mathbb{C})$ be a projective approximate zero of f with associated zero ζ, such that*

$$d_T(z, \zeta) \leq \frac{3 - \sqrt{7}}{d^{3/2}\mu_{\text{norm}}(f, \zeta)}.$$

Let $z^k = \widehat{N}_f^{(k)}(z)$, where $k \in \mathbb{N}$ is such that

$$k \geq \log_2 \log_2(4(1 + \|\varphi_0^{-1}(\zeta)\|^2)).$$

Then,

$$\|\varphi_0^{-1}(z^k) - \varphi_0^{-1}(\zeta)\| \leq \frac{3 - \sqrt{7}}{d^{3/2}\mu_{\text{norm}}(f, \zeta)}.$$

In particular (from Proposition 4.1), $\varphi_0^{-1}(z^k)$ is an affine approximate zero of f with associated zero $\varphi_0^{-1}(\zeta)$.

Proof. From the definition of projective approximate zero,

$$d_T(z^k, \zeta) \leq \frac{1}{2^{2^k - 1}} d_T(z, \zeta) \leq \frac{1}{2(1 + \|\varphi_0^{-1}(\zeta)\|^2)} \frac{3 - \sqrt{7}}{d^{3/2}\mu_{\text{norm}}(f, \zeta)}.$$

The reader may check that we are under the hypotheses of Lemma 4.4. We conclude that

$$\|\varphi_0^{-1}(z^k) - \varphi_0^{-1}(\zeta)\| \leq \frac{1 + \|\varphi_0^{-1}(\zeta)\|^2}{1 - \|\varphi_0^{-1}(\zeta)\|\varepsilon}\, \varepsilon,$$

where $\varepsilon = \frac{1}{2(1+\|\varphi_0^{-1}(\zeta)\|^2)} \frac{3-\sqrt{7}}{d^{3/2}\mu_{\text{norm}}(f,\zeta)}$. The proposition easily follows. □

4.2. **The average size of affine solutions.** Proposition 4.5 yields a way of obtaining an affine approximate zero $\varphi_0^{-1}(y)$ of f from a projective approximate zero of f with associated zero ζ. However, we need to control $\|\varphi_0^{-1}(\zeta)\|$ in order to estimate the number of steps of the projective Newton operator necessary to guarantee that $\varphi_0^{-1}(y)$ is really an affine approximate zero of f. Now we dedicate some paragraphs to study the probability distribution of $\|\varphi_0^{-1}(\zeta)\|$ under our hypotheses.

Lemma 4.6. *Let $1 - 2n \leq \alpha < 2$. Then*

$$\frac{1}{\nu[\mathbb{P}_n(\mathbb{C})]} \int_{x \in \mathbb{P}_n(\mathbb{C})} \|\varphi_0^{-1}(x)\|^\alpha \, d\mathbb{P}_n(\mathbb{C}) = \frac{\Gamma\left(1 - \frac{\alpha}{2}\right)\Gamma\left(n + \frac{\alpha}{2}\right)}{\Gamma(n)},$$

where $\nu[\mathbb{P}_n(\mathbb{C})] = \pi^n/\Gamma(n+1)$ is the volume of the complex projective space.

Proof. In [BP07b, Lemma 21], the authors proved that

$$NJ_x\varphi_0 = \frac{1}{(1 + \|x\|^2)^{n+1}}.$$

Then, from Theorem 2.2 we have

$$\int_{x \in \mathbb{C}^n} \frac{\|x\|^\alpha}{(1 + \|x\|^2)^{n+1}} \, d\mathbb{C}^n = \int_{z \in \mathbb{P}_n(\mathbb{C})} \|\varphi_0^{-1}(z)\|^\alpha d\mathbb{P}_n(\mathbb{C}).$$

On the other hand, using polar coordinates,

$$\int_{x\in\mathbb{C}^n} \frac{\|x\|^\alpha}{(1+\|x\|^2)^{n+1}}\, d\mathbb{C}^n = \frac{2\pi^n}{\Gamma(n)} \int_0^\infty \frac{t^{2n-1+\alpha}}{(1+t^2)^{n+1}}\, dt = \frac{\pi^n}{\Gamma(n)} \frac{\Gamma\left(1-\frac{\alpha}{2}\right)\Gamma\left(n+\frac{\alpha}{2}\right)}{\Gamma(n+1)}.$$

The lemma follows, as

$$\nu[\mathbb{P}_n(\mathbb{C})] = \frac{\pi^n}{\Gamma(n+1)}. \qquad \square$$

Theorem 4.7. *Let $1-2n \leq \alpha < 2$. Let $\mathcal{B}^\alpha_{(d)}$ be the expected value of the norm of the affine solutions of a random system of polynomial equations, powered to α. Namely,*

$$\mathcal{B}^\alpha_{(d)} = \frac{1}{\nu_\Delta[\mathbb{S}_{(d)}]} \int_{f\in\mathbb{S}_{(d)}} \frac{1}{\mathcal{D}} \sum_{\zeta\in V_{\mathbb{P}}(f)} \|\varphi_0^{-1}(\zeta)\|^\alpha\, d\mathbb{S}_{(d)}.$$

Then,

$$\mathcal{B}^\alpha_{(d)} = \frac{\Gamma\left(1-\frac{\alpha}{2}\right)\Gamma\left(n+\frac{\alpha}{2}\right)}{\Gamma(n)}.$$

Proof. Let

$$\mathcal{I}^\alpha_{(d)} = \int_{f\in\mathbb{S}_{(d)}} \sum_{\zeta\in V_{\mathbb{P}}(f)} \|\varphi_0^{-1}(\zeta)\|^\alpha\, d\mathbb{S}_{(d)}.$$

Observe that

$$\begin{aligned} \mathcal{I}^\alpha_{(d)} &= \nu_\Delta[\mathbb{S}_{(d)}]\mathcal{D}\mathcal{B}^\alpha_{(d)}, \\ \mathcal{I}^0_{(d)} &= \nu_\Delta[\mathbb{S}_{(d)}]\mathcal{D}. \end{aligned} \tag{4.3}$$

In fact, equality (4.3) is due to the fact that a generic system $f \in \mathcal{H}_{(d)}$ has exactly $\mathcal{D}$ projective solutions (cf. for example [BCSS98]). Let $\widehat{W} = \{(f,\zeta) \in \mathbb{S}_{(d)} \times \mathbb{P}_n(\mathbb{C}) : \zeta \in V_{\mathbb{P}}(f)\}$ be the incidence variety intersected with $\mathbb{S}_{(d)} \times \mathbb{P}_n(\mathbb{C})$. $\widehat{W}$ is a (real) manifold of real dimension $2N+1$ (see for example [BP08]). Let $p_1 : \widehat{W} \longrightarrow \mathbb{S}_{(d)}$ and $p_2 : \widehat{W} \longrightarrow \mathbb{P}_n(\mathbb{C})$ be the canonical projections. Then, from Theorem 2.2 applied to p_1, we have that

$$\mathcal{I}^\alpha_{(d)} = \int_{(f,x)\in\widehat{W}} \|\varphi_0^{-1}(x)\|^\alpha N J_{(f,x)} p_1\, d\widehat{W}.$$

Moreover, from Theorem 2.2 applied to p_2,

$$\begin{aligned} &\int_{(f,x)\in\widehat{W}} \|\varphi_0^{-1}(x)\|^\alpha N J_{(f,x)} p_1\, d\widehat{W} \\ &= \int_{x\in\mathbb{P}_n(\mathbb{C})} \int_{f\in p_2^{-1}(x)} \|\varphi_0^{-1}(x)\|^\alpha \frac{N J_{(f,x)} p_1}{N J_{(f,x)} p_2}\, dp_2^{-1}(x)\, d\mathbb{P}_n(\mathbb{C}), \end{aligned}$$

where for $x \in \mathbb{P}_n(\mathbb{C})$, $p_2^{-1}(x) \subseteq \mathbb{S}_{(d)} \times \{x\}$ may be identified with the set of systems that have x as a projective zero. We conclude that

$$\mathcal{I}^\alpha_{(d)} = \int_{x\in\mathbb{P}_n(\mathbb{C})} \|\varphi_0^{-1}(x)\|^\alpha \int_{f\in p_2^{-1}(x)} \frac{N J_{(f,x)} p_1}{N J_{(f,x)} p_2}\, dp_2^{-1}(x)\, d\mathbb{P}_n(\mathbb{C}). \tag{4.4}$$

Let $x \in \mathbb{P}_n(\mathbb{C})$ be a projective point and let $U \in \mathcal{U}_{n+1}$ be a unitary matrix such that $Ue_0 = x$. From equality (2.1), the mapping from $p_2^{-1}(x)$ onto $p_2^{-1}(e_0)$, sending each (f, x) to the pair $(f \circ U, e_0)$, is an isometry. Moreover, from Proposition 2.3, for every $f \in p_2^{-1}(e_0)$ we have that $NJ_{(f,e_0)}p_i = NJ_{(f\circ U^{-1},x)}p_i$, $i = 1, 2$. Hence, from Theorem 2.2 we have that

$$\begin{aligned}\int_{f\in p_2^{-1}(x)} \frac{NJ_{(f,x)}p_1}{NJ_{(f,x)}p_2}\, dp_2^{-1}(x) &= \int_{f\in p_2^{-1}(e_0)} \frac{NJ_{(f\circ U^{-1},x)}p_1}{NJ_{(f\circ U^{-1},x)}p_2}\, dp_2^{-1}(e_0)\\ &= \int_{f\in p_2^{-1}(e_0)} \frac{NJ_{(f,e_0)}p_1}{NJ_{(f,e_0)}p_2}\, dp_2^{-1}(e_0),\end{aligned}$$

and we conclude that the inner integral in equation (4.4) does not depend on the choice of $x \in \mathbb{P}_n(\mathbb{C})$. Hence,

$$I^{\alpha}_{(d)} = \left(\int_{f\in p_2^{-1}(e_0)} \frac{NJ_{(f,e_0)}p_1}{NJ_{(f,e_0)}p_2}\, dp_2^{-1}(e_0)\right) \int_{x\in\mathbb{P}_n(\mathbb{C})} \|\varphi_0^{-1}(x)\|^{\alpha} d\mathbb{P}_n(\mathbb{C}).$$

First, assume that $\alpha = 0$. Then, from equation (4.3),

$$\nu_{\Delta}[\mathbb{S}_{(d)}]\mathcal{D} = I^0_{(d)} = \nu[\mathbb{P}_n(\mathbb{C})] \int_{f\in p_2^{-1}(e_0)} \frac{NJ_{(f,e_0)}p_1}{NJ_{(f,e_0)}p_2}\, dp_2^{-1}(e_0),$$

and we obtain that

$$\int_{f\in p_2^{-1}(e_0)} \frac{NJ_{(f,e_0)}p_1}{NJ_{(f,e_0)}p_2}\, dp_2^{-1}(e_0) = \frac{\nu_{\Delta}[\mathbb{S}_{(d)}]\mathcal{D}}{\nu[\mathbb{P}_n(\mathbb{C})]}.$$

We conclude that

$$I^{\alpha}_{(d)} = \frac{\nu_{\Delta}[\mathbb{S}_{(d)}]\mathcal{D}}{\nu[\mathbb{P}_n(\mathbb{C})]} \int_{x\in\mathbb{P}_n(\mathbb{C})} \|\varphi_0^{-1}(x)\|^{\alpha} d\mathbb{P}_n(\mathbb{C}).$$

Hence,

$$\mathcal{B}^{\alpha}_{(d)} = \frac{1}{\nu[\mathbb{P}_n(\mathbb{C})]} \int_{x\in\mathbb{P}_n(\mathbb{C})} \|\varphi_0^{-1}(x)\|^{\alpha}\, d\mathbb{P}_n(\mathbb{C}).$$

The theorem follows from Lemma 4.6. □

An immediate consequence of Theorem 4.7 is the following result.

Corollary 4.8. *Let $\delta > 0$. Let P_δ be the probability that a randomly chosen system $f \in \mathbb{S}_{(d)}$ has an affine solution $\zeta \in V_{\mathbb{C}}(f)$ with $\|\zeta\| > \delta$. Then,*

$$P_\delta \leq \mathcal{D}\frac{\Gamma\left(1-\frac{\alpha}{2}\right)\Gamma\left(n+\frac{\alpha}{2}\right)}{\Gamma(n)\delta^{\alpha}}.$$

Here, $\alpha \in [0, 2)$ is any positive real number.

Proof. Observe that

$$\mathrm{E}_{f\in\mathbb{S}_{(d)}}\left[\max_{\zeta\in V_{\mathbb{C}}(f)} \|\zeta\|^{\alpha}\right] \leq \frac{1}{\nu_{\Delta}[\mathbb{S}_{(d)}]} \int_{f\in\mathbb{S}_{(d)}} \sum_{\zeta\in V_{\mathbb{P}}(f)} \|\varphi_0^{-1}(\zeta)\|^{\alpha}\, d\mathbb{S}_{(d)}.$$

From Theorem 4.7, this last term equals

$$\mathcal{D}\frac{\Gamma\left(1-\frac{\alpha}{2}\right)\Gamma\left(n+\frac{\alpha}{2}\right)}{\Gamma(n)}.$$

From Markov's Inequality, for every positive real number $\delta > 0$, we conclude that

$$\begin{aligned}&\mathrm{Prob}_{f\in\mathbb{S}_{(d)}}\left[\exists \zeta \in V_{\mathbb{C}^n}(f) : \|\zeta\| \geq \delta\right]\\ &= \mathrm{Prob}_{f\in\mathbb{S}_{(d)}}\left[\exists \zeta \in V_{\mathbb{C}^n}(f) : \|\zeta\|^\alpha \geq \delta^\alpha\right] \leq \mathcal{D}\frac{\Gamma\left(1-\frac{\alpha}{2}\right)\Gamma\left(n+\frac{\alpha}{2}\right)}{\Gamma(n)\delta^\alpha}.\end{aligned}$$

□

Corollary 4.9. *Let $\delta > 0$. Let P_δ be the probability that a randomly chosen system $f \in \mathbb{S}_{(d)}$ has an affine solution $\zeta \in V_{\mathbb{C}}(f)$ with $\|\zeta\| > \delta$. Then,*

$$P_\delta \leq \frac{\mathcal{D}\sqrt{\pi n}}{\delta}, \quad \delta > 0,$$
$$P_\delta \leq e\mathcal{D}\frac{n}{\delta^2}\ln\left(\frac{\delta^2}{n}\right), \quad \delta \geq e\sqrt{n}.$$

Proof. This comes directly from Corollary 4.8, using the fact that

$$\frac{\Gamma\left(n+\frac{\alpha}{2}\right)}{\Gamma(n)} \leq n^{\alpha/2}, \quad \alpha \in [1,2)$$

(cf. for example [EGP00]). We also use that

$$\Gamma\left(1-\frac{\alpha}{2}\right) \leq \frac{2}{2-\alpha}, \quad \alpha \in [1,2).$$

For the first inequality of the corollary, let $\alpha = 1$. For the second one, let

$$\alpha = 2\frac{\ln\left(\frac{\delta^2}{n}\right)-1}{\ln\left(\frac{\delta^2}{n}\right)}.$$

□

Theorem 1.9 in the Introduction is a weak form of this Corollary 4.9. For the purposes of numerical solving of systems of equations, the result as stated in the Introduction suffices.

4.3. **Proof of Theorem 1.10.** Let $\mathcal{A}_1, \mathcal{A}_2$ be the average number of arithmetic operations of the first and the second parts of the algorithm, respectively. Observe that

$$\text{total average number of arith. ops} = \mathcal{A}_1 + \mathcal{A}_2.$$

Moreover, from Theorem 1.4, we know that $\mathcal{A}_1 \leq cn^6N^3d^3\log_2 d\log_2\mathcal{D}$. For $\mathcal{A}_2$, let

$$Q_i \leq cnNi\log_2 d \leq cN^2 i$$

be the number of arithmetic operations required to perform i loops of the algorithm. Note that, depending on the method we use to perform the α-test, it may require to compute some extra Newton steps. This may increase the total complexity by at most a small constant factor. Let R_i be the probability that we reach the i-th loop of the second part of the algorithm. Then,

$$\mathcal{A}_2 \leq \sum_{i=1}^{\infty} R_i Q_i \leq cN^2 \sum_{i=1}^{\infty} iR_i.$$

Now, from Proposition 4.5, the set of systems that are solved in the i-th step of the algorithm contains the set of systems $f \in \mathbb{S}_{(d)}$ such that $\|\varphi_0^{-1}(\zeta)\|^2 \leq 2^{2^i-2}-1$, for every $\zeta \in V_{\mathbb{P}}(f)$. Thus, from Theorem 1.9,

$$R_i \leq \min\left\{1, \frac{\mathcal{D}\sqrt{\pi n}}{2^{2^{i-1}-2}-1}\right\} \leq \min\left\{1, \frac{6\mathcal{D}\sqrt{\pi n}}{2^{2^{i-1}}}\right\}, \quad \forall i \geq 2.$$

We conclude that

$$\mathcal{A}_2 \leq cN^2 \left(\sum_{i=1}^{\lceil \log_2 \log_2 \mathcal{D} \rceil} i + \mathcal{D}\sqrt{\pi n} \sum_{i=1+\lceil \log_2 \log_2 \mathcal{D} \rceil}^{\infty} \frac{i}{2^{2^{i-1}}} \right) \leq cN^2 \sqrt{n} (2 + \log_2 \log_2 \mathcal{D})^2.$$

Note that

$$N^2 \sqrt{n} (2 + \log_2 \log_2 \mathcal{D})^2 \leq cn^6 N^3 d^3 \log_2 d \log_2 \mathcal{D},$$

for any choice of n and the list of degrees (d). Thus, $\mathcal{A}_2 \leq c\mathcal{A}_1$, and the theorem follows. □

References

[BCSS98] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998. MR1479636 (99a:68070)

[BP06] C. Beltrán and L. M. Pardo, *On the complexity of non–universal polynomial equation solving: old and new results*, Foundations of Computational Mathematics: Santander 2005. L. Pardo, A. Pinkus, E. Süli, M. Todd, editors, Cambridge University Press, 2006, pp. 1–35. MR2277103 (2008a:65097)

[BP07a] C. Beltrán and L. M. Pardo, *On the probability distribution of condition numbers of complete intersection varieties and the average radius of convergence of Newton's method in the underdetermined case*, Math. Comp. **76** (2007), no. 259, 1393–1424 (electronic). MR2299780 (2008d:65053)

[BP07b] C. Beltrán and L. M. Pardo, *Estimates on the distribution of the condition number of singular matrices*, Found. Comput. Math. **7** (2007), no. 1, 87–134. MR2283343 (2008b:65059)

[BP08] C. Beltrán and L. M. Pardo, *On Smale's 17th problem: a probabilistic positive solution*, Found. Comput. Math. **8** (2008), no. 1, 1–43. MR2403529

[BS08] C. Beltrán and M. Shub, *Complexity of Bezout's Theorem VII: Distance Estimates in the Condition Metric*, J. FoCM **Online First 10.1007/s10208-007-9018-5** (2008).

[BSS89] L. Blum, M. Shub, and S. Smale, *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*, Bull. Amer. Math. Soc. (N.S.) **21** (1989), no. 1, 1–46. MR974426 (90a:68022)

[CMP92] F. Cucker, J. L. Montaña, and L. M. Pardo, *Time bounded computations over the reals*, Internat. J. Algebra Comput. **2** (1992), no. 4, 395–408. MR1189670 (93m:03067)

[CMP95] ———, *Models for parallel computation with real numbers*, Number-theoretic and algebraic methods in computer science (Moscow, 1993), World Sci. Publ., River Edge, NJ, 1995, pp. 53–63. MR1377740 (97b:68067)

[Coo85] Stephen A. Cook, *A taxonomy of problems with fast parallel algorithms*, Inform. and Control **64** (1985), no. 1-3, 2–22. MR837088 (87k:68043)

[Ded97] J. P. Dedieu, *Estimations for the separation number of a polynomial system*, J. Symbolic Comput. **24** (1997), no. 6, 683–693. MR1487794 (99b:65065)

[Ded01] ———, *Newton's method and some complexity aspects of the zero-finding problem*, Foundations of computational mathematics (Oxford, 1999), London Math. Soc. Lecture Note Ser., vol. 284, Cambridge Univ. Press, Cambridge, 2001, pp. 45–67. MR1836614 (2002d:65050)

[Ded06] ———, *Points fixes, zéros et la méthode de newton*, Collection Mathématiques et Applications, Springer, to appear 2006.

[DK00] Ding-Zhu Du and Ker-I Ko, *Theory of computational complexity*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. MR1789501 (2001k:68036)

[EGP00] N. Elezović, C. Giordano, and J. Pečarić, *The best bounds in Gautschi's inequality*, Math. Inequal. Appl. **3** (2000), no. 2, 239–252. MR1749300 (2001g:33001)

[Fed69] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag, New York, Inc., New York, 1969. MR0257325 (41:1976)

[Ho75] Chung Wu Ho, *A note on proper maps*, Proc. Amer. Math. Soc. **51** (1975), 237–241. MR0370471 (51:6698)

[How] R. Howard, *Analysis on homogeneous spaces*, Class notes, Spring 1994. Royal Institute of Technology, Stocholm.

[HSS01] J. Hubbard, D. Schleicher, and S. Sutherland, *How to find all roots of complex polynomials by Newton's method*, Invent. Math. **146** (2001), no. 1, 1–33. MR1859017 (2002i:37059)

[Kan49] L. Kantorovich, *Sur la méthode de Newton*, Travaux de l'Institut des Mathématiques Steklov **XXVIII** (1949), 104–144.

[Kim89] M.H. Kim, *Topological complexity of a root finding algorithm*, J. Complexity **5** (1989), no. 3, 331–344. MR1018023 (90m:65058)

[Koi96] P. Koiran, *Hilbert's Nullstellensatz is in the polynomial hierarchy*, J. Complexity **12** (1996), no. 4, 273–286, Special issue for the Foundations of Computational Mathematics Conference (Rio de Janeiro, 1997). MR1422712 (98e:68109)

[Mal94] G. Malajovich, *On generalized Newton algorithms: quadratic convergence, path-following and error analysis*, Theoret. Comput. Sci. **133** (1994), no. 1, 65–84, Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993). MR1294426 (95g:65073)

[Mil76] Gary L. Miller, *Riemann's hypothesis and tests for primality*, J. Comput. System Sci. **13** (1976), no. 3, 300–317, Working papers presented at the ACM-SIGACT Symposium on the Theory of Computing (Albuquerque, N.M., 1975). MR0480295 (58:470a)

[MP98] J. L. Montaña and Luis M. Pardo, *On Kolmogorov complexity in the real Turing machine setting*, Inform. Process. Lett. **67** (1998), no. 2, 81–86. MR1638154 (99c:68134)

[MR02] G. Malajovich and J. M. Rojas, *Polynomial systems and the momentum map*, Foundations of computational mathematics (Hong Kong, 2000), World Sci. Publishing, River Edge, NJ, 2002, pp. 251–266. MR2021984 (2004k:65090)

[Rab80] M. O. Rabin, *Probabilistic algorithm for testing primality*, J. Number Theory **12** (1980), no. 1, 128–138. MR566880 (81f:10003)

[Ren87] J. Renegar, *On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials*, Math. Oper. Res. **12** (1987), no. 1, 121–148. MR882846 (88j:65112)

[Shu93] M. Shub, *Some remarks on Bezout's theorem and complexity theory*, From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990) (New York), Springer, 1993, pp. 443–455. MR1246139 (95a:14002)

[Shu08] ———, *Complexity of Bezout's Theorem VI: Geodesics in the Condition (number) Metric*, J. FoCM **Online First 10.1007/s10208-007-9017-6** (2008).

[Sma86] S. Smale, *Newton's method estimates from data at one point*, The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985), Springer, New York, 1986, pp. 185–196. MR870648 (88e:65076)

[SS77] R. Solovay and V. Strassen, *A fast Monte-Carlo test for primality*, SIAM J. Comput. **6** (1977), no. 1, 84–85. MR0429721 (55:2732)

[SS78] ———, *Erratum: "A fast Monte-Carlo test for primality" (SIAM J. Comput.* **6** *(1977), no. 1, 84–85)*, SIAM J. Comput. **7** (1978), no. 1, 118. MR0466001 (57:5885)

[SS93a] M. Shub and S. Smale, *Complexity of Bézout's theorem. I. Geometric aspects*, J. Amer. Math. Soc. **6** (1993), no. 2, 459–501. MR1175980 (93k:65045)

[SS93b] ———, *Complexity of Bezout's theorem. II. Volumes and probabilities*, Computational algebraic geometry (Nice, 1992), Progr. Math., vol. 109, Birkhäuser Boston, Boston, MA, 1993, pp. 267–285. MR1230872 (94m:68086)

[SS93c] ———, *Complexity of Bezout's theorem. III. Condition number and packing*, J. Complexity **9** (1993), no. 1, 4–14, Festschrift for Joseph F. Traub, Part I. MR1213484 (94g:65152)

[SS94] ———, *Complexity of Bezout's theorem. V. Polynomial time*, Theoret. Comput. Sci. **133** (1994), no. 1, 141–164, Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993). MR1294430 (96d:65091)

[SS96] ———, *Complexity of Bezout's theorem. IV. Probability of success; extensions*, SIAM J. Numer. Anal. **33** (1996), no. 1, 128–148. MR1377247 (97k:65310)

[Yak95] J. C. Yakoubsohn, *A universal constant for the convergence of Newton's method and an application to the classical homotopy method*, Numer. Algorithms **9** (1995), no. 3-4, 223–244. MR1339720 (96d:65092)

Departmento de Matemáticas, Estadística y Computación. Fac. de Ciencias. Avda. Los Castros s/n. 39005 Santander, Spain

E-mail address: beltranc@unican.es

Departmento de Matemáticas, Estadística y Computación. Fac. de Ciencias. Avda. Los Castros s/n. 39005 Santander, Spain

E-mail address: luis.pardo@unican.es