Found Comput Math (2010) 10: 115–125
DOI 10.1007/s10208-009-9054-4

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

A Note on the Finite Variance of the Averaging Function for Polynomial System Solving

Carlos Beltrán · Michael Shub

Received: 14 October 2008 / Revised: 1 May 2009 / Accepted: 20 August 2009 /
Published online: 3 October 2009

Abstract In the forthcoming paper of Beltrán and Pardo, the average complexity of linear homotopy methods to solve polynomial equations with random initial input (in a sense to be described below) was proven to be finite, and even polynomial in the size of the input. In this paper, we prove that some other higher moments are also finite. In particular, we show that the variance is polynomial in the size of the input.

Keywords Homotopy method · Approximate zero · Probabilistic algorithm

Mathematics Subject Classification (2000) 65H10 · 65H20

1 Introduction

The complexity theory of solving systems of polynomial equations still poses many problems. Among these is whether there is a deterministic uniform algorithm which

Communicated by Michael Todd.

C. Beltrán was partially supported by MTM2007-62799 and a Spanish postdoctoral grant (FECYT).
M. Shub was partially supported by an NSERC Discovery Grant.
Thanks to one of the referees for his many comments that helped to improve the first version of this manuscript.

C. Beltrán
Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Santander, Spain
e-mail: carlos.beltran@unican.es

M. Shub (✉)
Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 2E4
e-mail: shub@math.toronto.edu

finds an approximate zero of a system of n homogeneous polynomials in $n+1$ complex variables of degrees $(d_1, \ldots, d_n)$ with average cost polynomial in the input size. This problem is the deterministic version of the 17th problem on Smale's list [18]. To our knowledge, this problem is not solved even for $n = 2$ and arbitrary degree in the context we consider below. For $n = 1$, the history of numerically solving complex polynomial equations is long. It is surveyed in Pan [9], see also Smale [17]. In [9] many algorithms are given for solving univariate complex polynomials, sometimes together with complexity estimates, which may be construed as affirmatively answering Smale's question in the positive for univariate polynomials. See references [6, 10], and [7] in [9] and Chap. 9 of [5] for approaches which bear a close relationship to our own. There has been much progress since, but we do not try to survey it here as it would take us too far afield except to mention [8], which has a probabilistic and condition aspect. In the form of the eigenvalue problem, a vast literature exists on numerical methods which perform wonders in practice, but to our knowledge, no rigorous complexity estimates are given, except perhaps for the reduction to the characteristic polynomial, which most numerical analysts caution against.

Recently, Beltrán and Pardo [1–4] have given a positive probabilistic answer to Smale's problem for all n and for all vectors of degrees. Here we give a brief indication of their and our results. More details and definitions are given below. Beltrán and Pardo show how to produce a random pair (g, ζ) where g is a polynomial system and ζ a zero of g starting from a random n by $n+1$ matrix. Then, given a pair (g, ζ), they find an approximate zero of f by a homotopy or continuation method, which performs a number of (projective) Newton steps. They give an upper bound for the cost (number of steps) of finding this approximate zero as a function $\mathcal{C}_0(g, f, \zeta)$ of f (see (1.4) below). Then the average

$$\mathcal{A}_1(g, \zeta) = \int_{f \in \mathbb{S}} \mathcal{C}_0(g, f, \zeta)\, \mathrm{d}\mathbb{S}$$

is up to a small factor $cd^{3/2}$, where c is a positive constant, and d the maximum of the d_i's, an upper bound for the average number of steps required by a path-following method starting at the pair (g, ζ) to produce an approximate zero of a system f with probability 1.

Let $(d) = (d_1, \ldots, d_n)$ be a list of positive degrees, and $\mathcal{H}_{(d)}$ be the vector space of homogeneous polynomial functions $f = (f_1, \ldots, f_n) : \mathbb{C}^{n+1} \to \mathbb{C}^n$, where f_j is homogeneous of degree d_j, $j = 1 \ldots n$. We denote $\mathcal{D} = d_1 \cdots d_n$ (Bézout's Number), $d = \max\{d_j : 1 \le j \le n\}$, and assume that $d \ge 2$. Note that $\mathcal{D} \le d^n$ is the number of complex projective solutions of any nondegenerate system $f \in \mathcal{H}_{(d)}$. We let $N+1$ denote the dimension of $\mathcal{H}_{(d)}$ as a vector space, i.e., $N+1$ is the number of monomials of a generic $f \in \mathcal{H}_{(d)}$.

A consequence of Theorem 1 in [3] is[1]:

[1]The paper [3] deals with slightly different quantities $\mathcal{C}$ and $\mathcal{A}$ satisfying $\mathcal{C}_0 \le \sqrt{2}\mathcal{C}$ and $\mathcal{A}_1 \le \sqrt{2}\mathcal{A}$, thanks to inequality (2.1) below. Thus, the results in [3] are still valid up to a small constant $\sqrt{2}$. The quantities $\mathcal{A}$ of [3] and $\mathcal{A}_1$ of this paper are versions of the integrals studied in the article [16].

Theorem 1 *The expectation*

$$\mathrm{E}\big(\mathcal{A}_1(g,\zeta)\big) \le 16\sqrt{2}\pi nN. \tag{1.1}$$

We describe the probability measures of this theorem below.

Theorem 1 has a simple interpretation via Markov's inequality in terms of the probability that $\mathcal{A}_1(g,\zeta)$ is small, see [3].

Corollary 1 *For* $0<\sigma<1$, *with probability* $1-\sigma$, $\mathcal{A}_1(g,\zeta) \le \sigma^{-1}16\sqrt{2}\pi nN$.

In this paper we extend the polynomial conclusion of Beltrán and Pardo to some higher moments. Hence, in particular, the central limit theorem will apply. Let

$$\mathcal{A}_k(g,\zeta) = \int_{f\in\mathbb{S}} \big(\mathcal{C}_0(g,f,\zeta)\big)^k \,\mathrm{d}\mathbb{S}.$$

Theorem 2 *Let* $2 \le k < 3$. *Then, the expectation of* $\mathcal{A}_k(g,\zeta)$ *satisfies*

$$\mathrm{E}\big(\mathcal{A}_k(g,\zeta)\big) < \infty.$$

Moreover, let $2 \le k < 3 - \frac{1}{2\ln \mathcal{D}}$. *Then, the expectation* $\mathrm{E}(\mathcal{A}_k(g,\zeta))$ *satisfies*

$$\mathrm{E}\big(\mathcal{A}_k(g,\zeta)\big) \le 2^{2k+k/2+4} e\pi^k n^{3k-4} N^2 \mathcal{D}^{4k-8} \ln \mathcal{D}.$$

In particular, $\mathrm{E}(\mathcal{A}_2(g,\zeta)) \le 512 e\pi^2 n^2 N^2 \ln \mathcal{D}$.

We will prove Theorem 2 in the next section of the paper.

Now Theorems 1 and 2 have a simple interpretation. Let $C_1 = 32\sqrt{2}\pi$ and $C_2 = 1024 e\pi^2$.

Corollary 2 *For* $0<\sigma<1$, *with probability* $1-\sigma$, $\mathcal{A}_1(g,\zeta) \le C_1\sigma^{-1}nN$ *and* $\mathcal{A}_2(g,\zeta) \le C_2\sigma^{-1}n^2N^2\ln\mathcal{D}$.

Example 1 Fix $\sigma = 1/2$. Then the corollary asserts that with probability $1/2$ in the choice of (g,ζ), the mean of the number of (projective) Newton steps for finding, with probability one, an approximate zero of a system f of homogeneous polynomial equations of degrees $(d_1,\dots,d_n)$ is bounded above by $O(d^{3/2}nN)$, and the variance is bounded above by $O(n^2N^2d^3\ln\mathcal{D})$.

Now we describe the probability distribution we use. We consider the Bombieri–Weyl inner product in $\mathcal{H}_{(d)}$ (cf. [15]) and the associated norm $\|\cdot\|$ and Riemannian structure in the sphere $\mathbb{S} = \mathbb{S}(\mathcal{H}_{(d)}) = \{f \in \mathcal{H}_{(d)} : \|f\| = 1\}$, normalized in such a way that the total volume of $\mathbb{S}$ is 1.

The set $V = \{(g,\zeta) \in \mathbb{S}\times\mathbb{P}(\mathbb{C}^{n+1}) : g(\zeta) = 0\}$ is called the solution variety. We consider $\Omega = V$ endowed with the probability measure

$$\mathrm{P}(A) = \int_{g\in\mathbb{S}} \frac{1}{\mathcal{D}} \sum_{\zeta: g(\zeta)=0} \chi_A(g,\zeta)\,\mathrm{d}\mathbb{S} \tag{1.2}$$

for any Borel set $A \subseteq \Omega$. Let $\mathrm{d}\Omega$ be the associated volume element. Denote by $\mathrm{E}(\phi)$ the expected value of any measurable function $\phi : \Omega \to \mathbb{R} \cup \{\infty\}$, and by $\mathrm{Var}(\phi)$ its variance.[2]

Theorem 2 can be applied to estimate the higher moments of $\mathcal{A}_1(g, \zeta)$ as well.

Theorem 3 *Let $2 \leq k < 3$. Then, the expectation of $\mathcal{A}_1(g, \zeta)^k$ satisfies*

$$\mathrm{E}\big(\mathcal{A}_1(g, \zeta)^k\big) < \infty.$$

Moreover, let $2 \leq k < 3 - \frac{1}{2 \ln \mathcal{D}}$. Then, the expectation $\mathrm{E}(\mathcal{A}_1(g, \zeta)^k)$ satisfies

$$\mathrm{E}\big(\mathcal{A}_1(g, \zeta)^k\big) \leq 2^{2k+k/2+4} e \pi^k n^{3k-4} N^2 \mathcal{D}^{4k-8} \ln \mathcal{D}.$$

In particular, $\mathrm{E}(\mathcal{A}_1(g, \zeta)^2) \leq 512 e \pi^2 n^2 N^2 \ln \mathcal{D}$.

Proof The Hölder inequality (see, for example, [11, p. 63]) implies

$$\begin{aligned}\mathrm{E}\big(\mathcal{A}_1(g, \zeta)^k\big) &= \int_{(g,\zeta)\in\Omega} \left(\int_{f \in \mathbb{S}} \mathcal{C}_0(g, f, \zeta)\, \mathrm{d}\mathbb{S} \right)^k \mathrm{d}\Omega \\ &\leq \int_{(g,\zeta)\in\Omega} \int_{f \in \mathbb{S}} \mathcal{C}_0(g, f, \zeta)^k\, \mathrm{d}\mathbb{S}\, \mathrm{d}\Omega = \mathrm{E}\big(\mathcal{A}_k(g, \zeta)\big).\end{aligned}$$

So Theorem 2 finishes the proof. □

Now we explain a little how homotopy algorithms work and describe the function $\mathcal{C}_0(g, f, \zeta)$.

Let Σ' be the set of critical points of the projection $\pi : V \to \mathbb{S}$, $\pi(g, \zeta) = g$. Note that $(g, \zeta) \in \Sigma'$ if $g(\zeta) = 0$ and the Jacobian matrix $Dg(\zeta)$ is not of maximal rank.

A key ingredient in our analysis below is the condition number: For $(g, \zeta) \in V$, we define

$$\mu(g, \zeta) = \big\| \big(Dg(\zeta)\,|_{\zeta^\perp} \big)^{-1} \mathrm{Diag}\big(\|\zeta\|^{d_i - 1} d_i^{1/2} \big) \big\|, \quad (g, \zeta) \in V \setminus \Sigma',$$

or $\mu(g, \zeta) = \infty$ if $(g, \zeta) \in \Sigma'$. Note that $\mu(g, \zeta)$ is essentially equal to the operator norm of the inverse of the differential matrix $Dg(\zeta)$, restricted to the orthogonal complement of ζ. The rest of the factors in this definition are normalizing factors which make results look prettier and allow projective computations, see [15] for more details. Sometimes μ is denoted μ_{norm} or μ_{proj}, but we keep the most simple notation here. An important property of $\mu(g, \zeta)$ is that it bounds the operator norm of the derivative of the (locally defined) inverse of π. Namely, if $g = g_t \in \mathbb{S}$ is a smooth curve parameterized by a real $t \in (-\varepsilon, \varepsilon)$ and if $\zeta = \zeta_0$ (with $\|\zeta\| = 1$) is a solution of g_0 such that $\mu(g_0, \zeta_0) < \infty$, then ζ can locally be smoothly deformed in such a way

[2] The set Ω was denoted $\mathcal{G}_{(d)}$ in [3]. In that paper, an alternative description of $\mathcal{G}_{(d)}$ is given which allows one to randomly choose pairs $(g, \zeta) \in \Omega$. Then, it is proven that the obtained probability distribution is the one described above for Ω.

that ζ_t is a solution of g_t, and the tangent vector $\dot{\zeta}_0 = \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\zeta_t \in T_\zeta \mathbb{P}(\mathbb{C}^{n+1}) \equiv \zeta^\perp$ satisfies

$$\|\dot{\zeta}_0\| \leq \mu(g,\zeta)\left\| \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} g_t \right\|. \tag{1.3}$$

Let $(g,\zeta) \in V$ and $f \in \mathbb{S}$, $f \neq -g$. Let $L_{g,f} \subseteq \mathbb{S}$ be the (shorter) arc of the great circle joining g and f, and let $\Gamma(g,f,\zeta)$ be the connected component of $\pi^{-1}(L_{g,f})$ that contains (g,ζ). Thus, $\Gamma(g,f,\zeta)$ is a set of pairs $(\mathit{system}, \mathit{solution}) \in V$. The implicit function theorem guarantees that $\Gamma(g,f,\zeta)$ is a smooth arc if it does not intersect Σ' and that in that case $\pi|_{\Gamma(g,f,\zeta)}$ is a bijection, so that for $h \in L_{g,f}$, there is a unique zero ζ_h with $(h,\zeta_h) \in \Gamma(g,f,\zeta)$. Homotopy algorithms attempt to approximate this path starting at a known pair (g,ζ) to produce an approximate zero of the input problem f. The reader may find more background in [1, 2, 5, 15].

Let the arc $L_{g,f}$ be parameterized by some real parameter $t \in [a,b]$, and let $(\dot{h},\dot{\zeta}_h) \in T_{(h,\zeta_h)}V$ be the tangent vector to $\Gamma(g,f,\zeta)$ at (h,ζ_h). Define

$$\begin{aligned}\mathcal{C}_0(g,f,\zeta) &= \int_{h\in L_{g,f}} \mu(h,\zeta_h)\left\|(\dot{h},\dot{\zeta}_h)\right\| \mathrm{d}L_{g,f} \\ &= \int_a^b \mu\big(h(t),\zeta_{h(t)}\big)\left\|\big(h'(t),(\zeta_h)'(t)\big)\right\| \mathrm{d}t,\end{aligned} \tag{1.4}$$

or ∞ if $\Gamma(g,f,\zeta)$ intersects Σ'. By the change of variables formula, this number is independent of the chosen parameterization. Note that $\mathcal{C}_0(g,f,\zeta)$ is the length of the arc $\Gamma(g,f,\zeta)$ in the so-called "condition metric", namely the metric in $V \subseteq \mathbb{S}\times\mathbb{P}(\mathbb{C}^{n+1})$ obtained by pointwise multiplying the product metric by the condition number μ, see [14].

From [14] we know that $cd^{3/2}\mathcal{C}_0(g,f,\zeta)$ (c a universal constant) is an upper bound for the number of steps of a particular path-following method to approximate a zero of f starting from an approximation to ζ. More exactly, the homotopy method of [14] starts at the pair (g,z) where z is an approximate zero of g near ζ, and then it chooses a small step t and considers the polynomial system h which lies on the arc $L_{g,f}$ at distance t from g. An approximate zero of h is obtained by one application of Newton's method $N(h)$ with initial pair z (more exactly, one has to use projective Newton's method $N_{\mathbb{P}}(h)$ of [13]). This scheme is repeated until we reach f. The main result of [14] is that, if the homotopy step t is chosen properly, the total number of iterations is at most $cd^{3/2}\mathcal{C}_0(g,f,\zeta)$ with a universal constant c.

However, in [14] the method for choosing the homotopy step t is not described. In a future paper we will compute explicitly a value for these steps and all the constants involved in this process.

We finally want to study the variance of the running time of the following algorithm: Input f, choose $(g,\zeta) \in V$ at random, then perform the homotopy method starting at (g,ζ) to solve f. That is, we choose a random pair for each input system f, instead of fixing some random (g,ζ) and using it for solving every system $f \in \mathbb{S}$ as suggested by Example 1. Then for fixed $f \in \mathbb{S}$, let

$$\mathcal{B}(f) = \mathrm{E}\big(\mathcal{C}_0(g,f,\zeta)\big) = \int_{(g,\zeta)\in\Omega} \mathcal{C}_0(g,f,\zeta)\,\mathrm{d}\Omega,$$

which is thus (up to a small factor $cd^{3/2}$) an upper bound for the average number of steps performed by the homotopy algorithm to solve f with random pair (g, ζ). Note that by Fubini's Theorem, inequality (1.1) above also reads:

Corollary 3

$$\int_{f\in\mathbb{S}} \mathcal{B}(f)\,d\mathbb{S} \leq 16\sqrt{2}\pi nN.$$

Namely, the expected value of the number of steps for polynomial system solving (with random initial pair (g, ζ)) is almost linear in the size of the input. In the following result we prove that the variance of this quantity is again finite.

Theorem 4 *Let $2 \leq k < 3$. Then, $\mathcal{B}$ belongs to $L^k(\mathbb{S})$, that is,*

$$\int_{f\in\mathbb{S}} \mathcal{B}(f)^k\,d\mathbb{S} < \infty.$$

Moreover, let $2 \leq k < 3 - \frac{1}{2\ln \mathcal{D}}$. Then,

$$\int_{f\in\mathbb{S}} \mathcal{B}(f)^k\,d\mathbb{S} \leq 2^{2k+k/2+4} e\pi^k n^{3k-4} N^2 \mathcal{D}^{4k-8} \ln \mathcal{D}.$$

In particular, the variance of $\mathcal{B} : \mathbb{S} \to [0, \infty]$ is at most $512e\pi^2 n^2 N^2 \ln \mathcal{D}$.

That is:

Corollary 4 *The average number of homotopy steps performed by the homotopy algorithm with random initial pair (g, ζ) (see [3, Corollary 2] for a more detailed description of the random choice of (g, ζ)) has variance at most $O(d^3 n^2 N^2 \ln \mathcal{D})$, thus polynomial in the size of the input.*

Theorem 4 is proven in the third section of the paper.

Remark 1 In Theorems 2, 3, and 4, if $3 - \frac{1}{2\ln \mathcal{D}} \leq k < 3$, we have the upper bound

$$2^{2k+k/2+3}\pi^k n^{3k-4} N^2 \frac{\mathcal{D}^{2k-2}}{3-k},$$

valid for

$$\mathrm{E}\big(\mathcal{A}_k(g, \zeta)\big), \qquad \mathrm{E}\big(\mathcal{A}_1(g, \zeta)^k\big), \quad \text{and} \quad \int_{f\in\mathbb{S}} \mathcal{B}(f)^k\,d\mathbb{S}.$$

Finally, in the case $n = 1$ all the bounds in the theorems above can be improved by a factor of 2^{2k+1}.

In the next section we prove Theorem 2. Some background is provided in the last section for those not familiar with it.

2 Proof of Theorem 2

From now on, assume that $h = h(t) \in \mathbb{S}$ is arc-length parameterized, so $\|\dot{h}\| = 1$. Let $0 < \beta < 1 - 1/k$, and let $(\dot{h}, \dot{\zeta})$ be tangent to V. Then, inequality (1.3) and the fact that μ is bounded below by 1 imply

$$\begin{aligned}\left\|(\dot{h}, \dot{\zeta}_h)\right\| &= \sqrt{\|\dot{h}\|^2 + \|\dot{\zeta}_h\|^2} \leq \sqrt{\|\dot{h}\|^2 + \mu(h, \zeta_h)^2 \|\dot{h}\|^2} \\ &= \sqrt{1 + \mu(h, \zeta_h)^2} \leq \sqrt{2}\mu(h, \zeta_h), \end{aligned} \tag{2.1}$$

which implies

$$\mu(h, \zeta_h)\left\|(\dot{h}, \dot{\zeta}_h)\right\| \leq 2^{\frac{1-\beta}{2}} \mu(h, \zeta_h)^{2-\beta} \left\|(\dot{h}, \dot{\zeta}_h)\right\|^{\beta}.$$

For $\frac{1}{p} + \frac{1}{q} = 1$, the Hölder Inequality thus yields

$$\mathcal{C}_0(g, f, \zeta)^k = \left(\int_{h \in L_{g,f}} \mu(h, \zeta_h)\left\|(\dot{h}, \dot{\zeta}_h)\right\| \mathrm{d}L_{g,f}\right)^k \leq 2^{\frac{k(1-\beta)}{2}} I_1 I_2,$$

where

$$I_1 = \left(\int_{h \in L_{g,f}} \left\|(\dot{h}, \dot{\zeta}_h)\right\|^{q\beta} \mathrm{d}L_{g,f}\right)^{k/q},$$

$$I_2 = \left(\int_{h \in L_{g,f}} \mu(h, \zeta_h)^{p(2-\beta)} \mathrm{d}L_{g,f}\right)^{k/p}.$$

Note that, although not explicitly seen in the notation, ζ_h depends on h and on the chosen root ζ of g, for ζ_h is the unique root of h in the arc $\Gamma(g, f, \zeta)$. We assume from now on that $L_{g,f} \cap \pi(\Sigma') = \emptyset$, namely that every system $h \in L_{g,f}$ has all of its solutions nonsingular. We can assume this without loss of generality, as for almost every choice of g and f, this hypotheses holds, see Sect. 4 below.

Now, let $q = \frac{1}{\beta}$, so that $p = \frac{1}{1-\beta} < k$. Then, from [16, Lemma 7.3.a] (see also Lemma 1 below) we have

$$\begin{aligned} I_1 &= \left(\int_{h \in L_{g,f}} \left\|(\dot{h}, \dot{\zeta}_h)\right\| \mathrm{d}L_{g,f}\right)^{k\beta} \leq \left(\int_{h \in L_{g,f}} \left(1 + \|\dot{\zeta}_h\|\right) \mathrm{d}L_{g,f}\right)^{k\beta} \\ &\leq \left(\pi + 2\mathcal{D}^2\right)^{k\beta} \leq \left(3\mathcal{D}^2\right)^{k\beta} \quad \text{(using that } \mathcal{D} \geq d \geq 2\text{)}. \end{aligned}$$

On the other hand, again the Hölder inequality implies

$$I_2 \leq \pi^{k(1-\beta)-1} \int_{h \in L_{g,f}} \mu(h, \zeta_h)^{k(2-\beta)} \mathrm{d}L_{g,f}.$$

We have proved that

$$\int_{(g,\zeta) \in \Omega} \int_{f \in \mathbb{S}} \mathcal{C}_0(g, f, \zeta)^k \, \mathrm{d}\mathbb{S} \, \mathrm{d}\Omega \leq 2^{\frac{k(1-\beta)}{2}} \pi^{k(1-\beta)-1} \left(3\mathcal{D}^2\right)^{k\beta} J,$$

where

$$J = \int_{(g,\zeta)\in\Omega} \int_{f\in\mathbb{S}} \int_{h\in L_{g,f}} \mu(h,\zeta_h)^{k(2-\beta)}\, \mathrm{d}L_{g,f}\, \mathrm{d}\mathbb{S}\, \mathrm{d}\Omega.$$

From the definition of the probability measure in Ω, i.e., (1.2), we conclude that

$$J = \frac{1}{\mathcal{D}} \int_{g\in\mathbb{S}} \int_{f\in\mathbb{S}} \sum_{\zeta:g(\zeta)=0} \int_{h\in L_{g,f}} \mu(h,\zeta_h)^{k(2-\beta)}\, \mathrm{d}L_{g,f}\, \mathrm{d}\mathbb{S}\, \mathrm{d}\mathbb{S}.$$

By [3, Theorem 3] (see also Theorem 5 below), this last is at most

$$\frac{2\pi}{\mathcal{D}} \int_{f\in\mathbb{S}} \sum_{\eta:h(\eta)=0} \mu(f,\eta)^{k(2-\beta)}\, \mathrm{d}\mathbb{S}.$$

Finally, from [3, Corollary 5] (see also Proposition 1 below) we conclude that

$$J \le \frac{2\pi\,\Gamma(N+1)\Gamma(n^2+n-k(2-\beta)/2)}{\Gamma(N+1-k(2-\beta)/2)\Gamma(n^2+n)} \frac{2^{k(2-\beta)+2}}{4-k(2-\beta)} n^{3k(2-\beta)/2} \tag{2.2}$$

is valid while $k(2-\beta) < 4$, that is, $\beta > 2-4/k$. We may rewrite this last expression as

$$\frac{2^{2k+3}\pi n^{3k(2-\beta)/2}}{2^{k\beta}(4-k(2-\beta))} \cdot \frac{\Gamma(N+1)}{\Gamma(n^2+n)} \cdot \frac{\Gamma(n^2+n-k(2-\beta)/2)}{\Gamma(N+1-k(2-\beta)/2)}.$$

Assume that $n \ge 2$ and note that $N+1 \ge n^2+n$ and hence the function

$$\alpha \mapsto \frac{\Gamma(n^2+n-\alpha)}{\Gamma(N+1-\alpha)} = \frac{1}{(N-\alpha)\cdots(n^2+n-\alpha)}, \quad 0 \le \alpha \le 2,$$

is an increasing function of α. Thus, using that $k(2-\beta)/2 < 2$, we have

$$\frac{\Gamma(N+1)}{\Gamma(n^2+n)} \cdot \frac{\Gamma(n^2+n-k(2-\beta)/2)}{\Gamma(N+1-k(2-\beta)/2)} \le \frac{\Gamma(N+1)}{\Gamma(n^2+n)} \cdot \frac{\Gamma(n^2+n-2)}{\Gamma(N+1-2)} \le \frac{N^2}{n^4}.$$

We conclude that if $n \ge 2$,

$$J \le \frac{2^{2k+3}\pi n^{3k(2-\beta)/2-4} N^2}{2^{k\beta}(4-k(2-\beta))}.$$

Hence,

$$\begin{aligned} \mathrm{E}\big(\mathcal{A}_k(g,\zeta)\big) &\le 2^{\frac{k(1-\beta)}{2}} \pi^{k(1-\beta)-1} \big(3\mathcal{D}^2\big)^{k\beta} \frac{2^{2k+3}\pi n^{3k(2-\beta)/2-4} N^2}{2^{k\beta}(4-k(2-\beta))} \\ &\le 2^{2k+k/2+3} \pi^k n^{3k-4} N^2 \frac{\mathcal{D}^{2k\beta}}{4-k(2-\beta)} \end{aligned} \tag{2.3}$$

for $\beta \in (2-4/k, 1-1/k)$. The minimum of this function is easily obtained by computing the unique zero of the derivative:

$$\beta = \frac{1+2(2k-4)\ln \mathcal{D}}{2k\ln \mathcal{D}} \in (2-4/k, 1-1/k) \quad \text{if } 2 \le k < 3 - \frac{1}{2\ln \mathcal{D}}, \tag{2.4}$$

which yields the bound

$$\mathrm{E}\big(\mathcal{A}_k(g,\zeta)\big) \le 2^{2k+k/2+3}\pi^k n^{3k-4} N^2 e \mathcal{D}^{4k-8} \ln \mathcal{D}^2,$$

as claimed by the theorem. If $3 - \frac{1}{2\ln \mathcal{D}} \le k < 3$, then the valid interval for β, $\beta \in (2-4/k, 1-1/k)$, is nonempty, and hence $\mathrm{E}(\mathcal{A}_k(g,\zeta))$ is finite. The upper bound of Remark 1 in this last case is easily obtained by letting β approach $1-1/k$.

Finally, from the comments after Proposition 1 below inequality (2.2) can be more precisely stated in the special case $n=1$; namely, in that case, we have

$$J \le \frac{2\pi\Gamma(N+1)\Gamma(2-k(2-\beta)/2)}{\Gamma(N+1-k(2-\beta)/2)\Gamma(2)} \le \frac{4\pi N^2}{4-k(2-\beta)},$$

and then we have

$$\begin{aligned}\mathrm{E}\big(\mathcal{A}_k(g,\zeta)\big) &\le 2^{\frac{k(1-\beta)}{2}}\pi^{k(1-\beta)-1}\big(3\mathcal{D}^2\big)^{k\beta}\frac{4\pi N^2}{4-k(2-\beta)} \\ &\le 2^{2+k/2}\pi^k N^2 \frac{\mathcal{D}^{2k\beta}}{4-k(2-\beta)},\end{aligned}$$

valid for $\beta \in (2-4/k, 1-1/k)$. Hence, we get an improved version of inequality (2.3), and the bound in the remark follows, with an improvement of 2^{2k+1}.

3 Proof of Theorem 4

Hölder's Inequality implies

$$\begin{aligned}\int_{f\in\mathbb{S}} \mathcal{B}(f)^k \,\mathrm{d}\mathbb{S} &= \int_{f\in\mathbb{S}} \left(\int_{(g,\zeta)\in\Omega} \mathcal{C}_0(g,f,\zeta)\,\mathrm{d}\Omega\right)^k \mathrm{d}\mathbb{S} \\ &\le \int_{f\in\mathbb{S}}\int_{(g,\zeta)\in\Omega} \mathcal{C}_0(g,f,\zeta)^k \,\mathrm{d}\Omega\,\mathrm{d}\mathbb{S} \\ &= \int_{(g,\zeta)\in\Omega}\int_{f\in\mathbb{S}} \mathcal{C}_0(g,f,\zeta)^k \,\mathrm{d}\mathbb{S}\,\mathrm{d}\Omega,\end{aligned}$$

by Fubini's Theorem. Theorem 2 finishes the proof.

4 Some Previous Results Used in the Proof

For the sake of readability and completeness, we recall now some results that have been used in the proof and that were stated in previous papers. We start with a result

that bounds the length of the path of solutions ζ_t when the system f_t moves in a great circle of $\mathbb{S}$. We write this result (and the rest of the results in this section) with the language and notation of this paper.

Lemma 1 [16, Lemma 7.3.a] *Let $f_0 \neq \pm f_1 \in \mathbb{S}$, and let L_{f_0,f_1} be the small piece of the great circle joining f_0 and f_1. Let ζ_0 be a zero of f_0 and assume that ζ_0 can be smoothly deformed to ζ_t, a solution of f_t, for f_t parameterizing L_{f_0,f_1} as f_t moves from f_0 to f_1. Then, the total length of the curve $\zeta_t : 0 \leq t \leq 1$ in $\mathbb{P}(\mathbb{C}^{n+1})$ is at most $2\mathcal{D}^2$. Namely,*

$$\int_{h \in L_{f_0,f_1}} \|\dot{\zeta}_h\| \, \mathrm{d}L_{g,f} = \int_0^1 \left\| \frac{\mathrm{d}\zeta_t}{\mathrm{d}t} \right\| \mathrm{d}t \leq 2\mathcal{D}^2.$$

We have also used in the proof of the theorem that for almost every choice of $g, f \in \mathbb{S}$, the arc $L_{g,f}$ does not contain systems with singular solutions. We explain this idea with more detail now: Recall that $\Sigma' \subseteq V$ is the set of pairs $(g, \zeta) \in V$ such that $Dg(\zeta)$ is not of maximal rank or, equivalently, $\mu(g, \zeta) = \infty$ or ζ is a singular solution of g. Σ' is an algebraic subvariety of V. The set $\{g \in \mathcal{H}_{(d)} : Dg(\zeta)$ is not of maximal rank for some $\zeta, g(\zeta) = 0\}$ is a complex algebraic subvariety of $\mathcal{H}_{(d)}$ and has real codimension 2. Moreover, its intersection with the sphere $\mathbb{S}$, that is, $\Sigma = \pi(\Sigma')$ has also real codimension 2 in $\mathbb{S}$, and hence for every choice of g, f (except for a zero measure subset of $\mathbb{S} \times \mathbb{S}$), the great circle containing g and f does not intersect Σ. Every zero of every system $h \in L_{g,f}$ in that arc is nonsingular, and thus the roots of h are in one-to-one correspondence with the roots of g and f. The reader may see these facts and many other details on the geometric perspective of this problem in [5].

Another result used in the proof is a version of [3, Theorem 3], which we explain now.

Theorem 5 [3, Theorem 3] *For $f \neq \pm g \in \mathbb{S}$, let $\tilde{L}_{g,f}$ be the whole great circle containing g and f. For a measurable function $s : V \to [0, \infty]$, we have*

$$\int_{g \in \mathbb{S}} \int_{f \in \mathbb{S}} \sum_{\zeta : g(\zeta) = 0} \int_{h \in \tilde{L}_{g,f}} s(h, \zeta_h) \, \mathrm{d}\tilde{L}_{g,f} \, \mathrm{d}\mathbb{S} \, \mathrm{d}\mathbb{S} = 2\pi \int_{f \in \mathbb{S}} \sum_{\eta : f(\eta) = 0} s(f, \eta) \, \mathrm{d}\mathbb{S}.$$

(*Recall that ζ_h is the unique root of h that lies in the lifted path $\Gamma(g, h, \zeta)$, and hence it depends on h and on the root ζ of g.*)

Theorem 3 of [3] is this same result with the particular function $s = \mu^2$, but the proof is exactly the same for a generic measurable function s. Note that this is a result from Integral Geometry, similar to many others that can be found, for example, in [12]. The interested reader may find helpful a heuristic explanation of this last result (before diving into the proof of [3, Theorem 3]): Choosing at random two points in $\mathbb{S}$ and then choosing at random a point in the associated great circle must be the same as simply choosing at random a point in the sphere, because no point is preferred to another one with either of the two methods. One can make this argument more rigorous using the uniqueness of Haar's measure.

We recall now a result that bounds the expected value of the condition number μ.

Proposition 1 [3, Corollary 5] *Let* $0 < \alpha < 4$ *be a real number. Then*,

$$\int_{f\in\mathbb{S}} \sum_{\eta: f(\eta)=0} \mu(f,\eta)^\alpha \, \mathrm{d}\mathbb{S} \leq \frac{\mathcal{D}\Gamma(N+1)\Gamma(n^2+n-\alpha/2)}{\Gamma(N+1-\alpha/2)\Gamma(n^2+n)} \frac{2^{\alpha+2}}{4-\alpha} n^{3\alpha/2}.$$

In the case that $n = 1$, this last result can be improved using [3, Proposition 1] (which we do not include here), proving that

$$\int_{f\in\mathbb{S}} \sum_{\eta: f(\eta)=0} \mu(f,\eta)^\alpha \, \mathrm{d}\mathbb{S} = \frac{\mathcal{D}\Gamma(N+1)\Gamma(n^2+n-\alpha/2)}{\Gamma(N+1-\alpha/2)\Gamma(n^2+n)},$$

as used in the last part of the proof of Theorem 2, to obtain the bound for the case $n = 1$.

References

1. C. Beltrán, L.M. Pardo, On Smale's 17th problem: a probabilistic positive solution, *Found. Comput. Math.* **8**(1), 1–43 (2008).
2. C. Beltrán, L.M. Pardo, Smale's 17th problem: average polynomial time to compute affine and projective solutions, *J. Am. Math. Soc.* **22**, 363–385 (2009).
3. C. Beltrán, L.M. Pardo, Fast linear homotopy to find approximate zeros of polynomial systems (to appear).
4. C. Beltrán, L.M. Pardo, Computing several zeros of polynomial systems: a complexity analysis and Shannon's Entropy (to appear).
5. L. Blum, F. Cucker, M. Shub, S. Smale, *Complexity and Real Computation* (Springer, New York, 1998).
6. D. Coppersmith, C.A. Neff, Roots of a polynomial and its derivatives, in *Proceedings of the Fifth Annual ACM–SIAM Symposium on Discrete Algorithms* (Arlington, VA, 1994) (ACM, New York, 1994), pp. 271–279.
7. M.H. Kim, S. Sutherland, Polynomial root-finding algorithms and branched covers, *SIAM J. Comput.* **23**(2), 415–436 (1994).
8. M.H. Kim, M. Martens, S. Sutherland, *A universal bound for the average cost of rootfinding*. Preprint.
9. V.Y. Pan, Solving a polynomial equation: some history and recent progress, *SIAM Rev.* **39**(2), 187–220 (1997).
10. J. Renegar, On the worst-case arithmetic complexity of approximating zeros of polynomials, *J. Complex.* **3**(2), 90–113 (1987).
11. W. Rudin, *Real and Complex Analysis*, 3rd edn. (McGraw-Hill Book, New York, 1987).
12. L.A. Santaló, *Integral Geometry and Geometric Probability* (Addison-Wesley, Reading, 1976).
13. M. Shub, Some remarks on Bezout's theorem and complexity theory, in *From Topology to Computation: Proceedings of the Smalefest* (Berkeley, CA, 1990), ed. by M.W. Hirsch, J.E. Marsden, M. Shub (Springer, New York, 1993), pp. 443–455.
14. M. Shub, Complexity of Bézout's theorem. VI: Geodesics in the condition (number) metric, *Found. Comput. Math.* **9**(2), 171–178 (2009).
15. M. Shub, S. Smale, Complexity of Bézout's theorem. I. Geometric aspects, *J. Am. Math. Soc.* **6**(2), 459–501 (1993).
16. M. Shub, S. Smale, Complexity of Bezout's theorem. V. Polynomial time, *Theor. Comput. Sci.* **133**(1), 141–164 (1994). Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993).
17. S. Smale, Complexity theory and numerical analysis, in *Acta Numerica, 1997*. Acta Numer., vol. 6 (Cambridge University Press, Cambridge, 1997), pp. 523–551.
18. S. Smale, Mathematical problems for the next century, in *Mathematics: Frontiers and Perspectives* (American Mathematical Society, Providence, 2000), pp. 271–294.