Solving Polynomial Equations in Smoothed Polynomial Time and a Near Solution to Smale's 17th Problem

[Extended Abstract] *

Peter Bürgisser†
Institute of Mathematics
University of Paderborn
D-33098 Paderborn, Germany
pbuerg@upb.de

Felipe Cucker‡
Dept. of Mathematics
City University of Hong Kong
Kowloon Tong, Hong Kong
macucker@cityu.edu.hk

ABSTRACT

The 17th of the problems proposed by Steve Smale for the 21st century asks for the existence of a deterministic algorithm computing an approximate solution of a system of n complex polynomials in n unknowns in time polynomial, on the average, in the size N of the input system. A partial solution to this problem was given by Carlos Beltrán and Luis Miguel Pardo who exhibited a randomized algorithm, call it LV, doing so. In this paper we further extend this result in several directions. Firstly, we perform a smoothed analysis (in the sense of Spielman and Teng) of algorithm LV and prove that its smoothed complexity is polynomial in the input size and σ^{-1}, where σ controls the size of the random perturbation of the input systems. Secondly, we perform a condition-based analysis of LV. That is, we give a bound, for each system f, of the expected running time of LV with input f. In addition to its dependence on N this bound also depends on the condition of f. Thirdly, and to conclude, we return to Smale's 17th problem as originally formulated for deterministic algorithms. We exhibit such an algorithm and show that its average complexity is $N^{\mathcal{O}(\log\log N)}$. This is nearly a solution to Smale's 17th problem.

Categories and Subject Descriptors

F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems—*computations on polynomials*; G.1.5 [**Numerical Analysis**]: Roots of Nonlinear Equations—*Continuation (homotopy) methods*

General Terms

Algorithms

Keywords

polynomial equation solving, homotopy methods, approximate zero, complexity, polynomial time, smoothed analysis

1. INTRODUCTION

In 2000, Steve Smale published a list of mathematical problems for the 21st century [20]. The 17th problem in the list reads as follows:

> *Can a zero of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?*

Smale pointed out that "it is reasonable" to homogenize the polynomial equations by adding a new variable and to work in projective space after which he made precise the different notions intervening in the question above. We provide these definitions in full detail in Section 2. Before doing so, in the remaining of this section, we briefly describe the recent history of Smale's 17th problem and the particular contribution of the present paper. The following summary of notations should suffice for this purpose.

We denote by $\mathcal{H}_{\mathbf{d}}$ the linear space of complex homogeneous polynomial systems in $n+1$ variables, with a fixed degree pattern $\mathbf{d} = (d_1, \ldots, d_n)$. We let $D = \max_i d_i$ and $N = \dim_{\mathbb{C}} \mathcal{H}_{\mathbf{d}}$. We endow this space with the unitarily invariant Bombieri-Weyl Hermitian product and consider the unit sphere $S(\mathcal{H}_{\mathbf{d}})$ with respect to the norm induced by this product. We then make this sphere a probability space by considering the uniform measure on it. The expression "on the average" refers to expectation on this probability space. Also, the expression "approximate zero" refers to a point for which Newton's method, starting at it, converges immediately, quadratically fast. Finally, the notion of "uniform polynomial time algorithm" refers to the so-called BSS-model [5], which is essentially a model of a random access machine operating with real numbers with infinite precision and at unit cost.

This is the setting underlying the series of papers [14, 15, 16, 17, 18] —commonly referred to as "the Bézout series"— written by Shub and Smale during the first half of the 1990s, a collection of ideas, methods, and results that pervade all the research done in Smale's 17th problem since this was proposed. The overall idea in the Bézout series is to use a

*A full version of this paper is available at `arxiv.org/abs/0909.2114`

†Partially supported by DFG grant BU 1371/2-1 and Paderborn Institute for Scientific Computation (PaSCo).

‡Partially supported by GRF grant CityU 100808.

STOC'10, June 5–8, 2010, Cambridge, Massachusetts, USA.

linear homotopy. That is, one starts with a system g and a zero ζ of g and considers the segment $E_{f,g}$ with extremities f and g. Here f is the system whose zero we want to compute. Almost surely, when one moves from g to f, the zero ζ of g follows a curve in projective space to end in a zero of f. The homotopy method consists of dividing the segment $E_{f,g}$ in a number, say k, of subsegments E_i small enough to ensure that an approximate zero x_i of the system at the origin of E_i can be made into an approximate zero x_{i+1} of the system at its end (via one step of Newton's method). The difficulty of this overall idea lies in the following issues:

1. How does one choose the initial pair (g, ζ)?
2. How does one choose the subsegments E_i? In particular, how large k should be?

The state of the art at the end of the Bézout series, i.e., in [18], showed an incomplete picture. For (2), an adaptive algorithm is described which determines the interval E_{i+1} from data computed from the interval E_i. The analysis of the size of k, however, reveals a dependence on the initial pair (g, ζ) through a quantity involving the whole path $E_{f,g}$. Which brings us to issue (1). Concerning this issue, Shub and Smale proved that good initial pairs (g, ζ) (in the sense that the average number of iterations for the algorithm above was polynomial in the size of f) existed for each degree pattern $\mathbf{d}$, but they could not exhibit a procedure to generate one such pair.

The next breakthrough took a decade to come. Beltrán and Pardo proposed in [1, 2, 3] that the initial pair (g, ζ) should be randomly chosen. The consideration of randomized algorithms departs from the formulation of Smale's 17th problem but it is widely accepted that, in practical terms, such algorithms are as good as their deterministic siblings. And in the case at hand this departure turned out to pay off. The average (over f) of the expected (over (g, ζ)) number of iterations of the algorithm proposed in [3] is $\mathcal{O}(nND^{3/2})$. One of the most notable features of the ideas introduced by Beltrán and Pardo is the use of a measure on the space of pairs (g, ζ) which is friendly enough to perform a probabilistic analysis while, at the same time, does allow for an efficient sampling.

A first goal of this paper is to perform a smoothed analysis of a randomized algorithm (essentially Beltrán and Pardo's), which we call LV, that computes an approximate zero of f. The precise details of this smoothed analysis are in §3.3.

Average (or smoothed) complexity results do not provide information on the running time of an algorithm for the instance at hand. The second goal of this paper shows a bound of this kind for the expected running time of LV. This bound depends on the condition number of the input system f. The precise statement, Theorem 3.5, is in §3.5 below.

Last but not least, to close this introduction, we return to its opening theme: Smale's 17th problem. Even though randomized algorithms are efficient in theory and reliable in practice they do not offer an answer to the question of the existence of a *deterministic* algorithm computing approximate zeros of complex polynomial systems in average polynomial time. The third main result in this paper exhibits a deterministic algorithm computing approximate zeros in average time $N^{\mathcal{O}(\log\log N)}$. To do so we design and analyze a deterministic homotopy algorithm, call it MD, whose average complexity is polynomial in n and N and exponential in D. This already yields a polynomial-time algorithm when one restricts the degree D to be at most $n^{1-\varepsilon}$ for any fixed $\varepsilon > 0$ (and, in particular, when D is fixed as in a system of quadratic or cubic equations). Algorithm MD is fast when D is small. We complement it with an algorithm that uses a procedure proposed by Jim Renegar [10] and which computes approximate zeros similarly fast when D is large.

In order to prove the results described above we have relied on a number of ideas and techniques. Some of them — e.g., the use of the coarea formula or of the Bombieri-Weyl Hermitian inner product— are taken from the Bézout series and are pervasive in the literature on the subject. Some others —notably the use of the Gaussian distribution and its truncations in Euclidean space instead of the uniform distribution on a sphere or a projective space— are less common. The blending of these ideas has allowed us a development, which unifies the treatment of the several situations we consider for zero finding in this paper.

Acknowledgments. We thank Carlos Beltrán, Jean-Pierre Dedieu and Mike Shub for helpful criticism and comments. Part of the manuscript was written while the authors stayed at the Fields Institute in Toronto during the thematic program on the Foundations of Computational Mathematics. The stimulating working conditions there and the financial support are gratefully acknowledged.

2. PRELIMINARIES

2.1 Setting and Notation

Fix a degree pattern $\mathbf{d} = (d_1, \ldots, d_n)$. The input space is the vector space $\mathcal{H}_\mathbf{d}$ of polynomial systems $f = (f_1, \ldots, f_n)$ with $f_i = \sum_\alpha a^i_\alpha X^\alpha \in \mathbb{C}[X_0, \ldots, X_n]$ homogeneous of degree d_i. We endow $\mathcal{H}_\mathbf{d}$ with the *Bombieri-Weyl Hermitian inner product* that is associated with the norm

$$\|f\|^2 := \sum_{|\alpha|=d_i} |a^i_\alpha|^2 \binom{d_i}{\alpha}^{-1}.$$

The reason to do so is that this inner product is invariant under the natural action of the unitary group $U(n+1)$, cf. [5, Chap. 12]. The quantity $N := \dim_\mathbb{C} \mathcal{H}_\mathbf{d}$ measures the size of the input system f and we further put $D := \max_i d_i$ and let $\mathcal{D} = \prod_i d_i$ be the *Bézout number*.

We look for solutions ζ of the equation $f(\zeta) = 0$ in the complex projective space $\mathbb{P}^n := \mathbb{P}(\mathbb{C}^{n+1})$. For $f, g \in \mathcal{H}_\mathbf{d} \setminus \{0\}$, we denote by $d_\mathbb{S}(f, g)$ the angle between f and g. Similarly we define $d_\mathbb{P}(x, y)$ for $x, y \in \mathbb{P}^n$. We define the *solution variety* to be

$$V_\mathbb{P} := \big\{(f, \zeta) \in \mathcal{H}_\mathbf{d} \times \mathbb{P}^n \mid f(\zeta) = 0\big\}.$$

This is a smooth submanifold of $\mathcal{H}_\mathbf{d} \times \mathbb{P}^n$. We denote by $V_\mathbb{P}(f)$ the zero set of $f \in \mathcal{H}_\mathbf{d}$ in $\mathbb{P}^n$. By Bézout's Theorem, it contains $\mathcal{D}$ points for almost all f. Let $Df(\zeta)_{|T_\zeta}$ denote the restriction of the derivative of $f\colon \mathbb{C}^{n+1} \to \mathbb{C}^n$ at ζ to the tangent space $T_\zeta := \{v \in \mathbb{C}^{n+1} \mid \langle v, \zeta\rangle = 0\}$ of $\mathbb{P}^n$ at ζ. The *discriminant variety* Σ is the set of systems f for which there exists a zero $\zeta \in \mathbb{P}^n$ such that $\mathsf{rank}\, Df(\zeta)_{|T_\zeta} < n$. Note that $f \notin \Sigma$ means all the zeros of f are simple.

The distribution of inputs will be modelled with isotropic Gaussians. We recall that the *Gaussian distribution* $N(\overline{f}, \sigma^2 \mathrm{I})$ on $\mathcal{H}_\mathbf{d}$ with mean $\overline{f} \in \mathcal{H}_\mathbf{d}$ and covariance matrix $\sigma^2\mathrm{I}$, $\sigma > 0$,

is given by the density

$$\rho_{\overline{f},\sigma}(f) = \Big(\frac{1}{\sigma\sqrt{2\pi}}\Big)^{2N} \exp\big(-\frac{\|f-\overline{f}\|^2}{2\sigma^2}\big).$$

2.2 Approximate zeros

In [12], Mike Shub introduced the following *projective version of Newton's method*. We associate to $f \in \mathcal{H}_{\mathbf{d}}$ a map $N_f : \mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{C}^{n+1} \setminus \{0\}$ defined (almost everywhere) by

$$N_f(z) = z - Df(z)^{-1}_{|T_z} f(z).$$

Note that $N_f(z)$ is homogeneous of degree 0 in f so that N_f induces a rational map from $\mathbb{P}^n$ to $\mathbb{P}^n$.

It is well-known that when z is sufficiently close to a simple zero ζ of f, the sequence of Newton iterates beginning at z will converge quadratically fast to ζ. This property lead Steve Smale to define the following intrinsic notion of approximate zero.

DEFINITION 2.1. *By an* approximate zero *of $f \in \mathcal{H}_{\mathbf{d}}$ associated with a zero $\zeta \in \mathbb{P}^n$ of f we understand a point $z \in \mathbb{P}^n$ such that the sequence of Newton iterates $z_{i+1} := N_f(z_i)$ with initial point $z_0 := z$ converges immediately quadratically to ζ, i.e., $d_{\mathbb{P}}(z_i,\zeta) \le 2^{-(2^i-1)}\, d_{\mathbb{P}}(z_0,\zeta)$ for all $i \in \mathbb{N}$.*

2.3 Condition Numbers

The so called α-theory of Smale [19] shows that the size of the basin of attraction of approximate zeros for a zero ζ of f is controlled by the following quantity:

$$\mu(f,\zeta) = \|f\| \cdot \|M^\dagger\|, \tag{1}$$

where (choosing a representative of ζ with $\|\zeta\| = 1$)

$$M := \Delta^{-1} Df(\zeta) \in \mathbb{C}^{n\times(n+1)},$$

with the diagonal matrix $\Delta := \mathsf{diag}(\sqrt{d_1},\ldots,\sqrt{d_n})$. Here $M^\dagger = M^*(MM^*)^{-1}$ denotes the Moore-Penrose inverse of M, and $\|M^\dagger\|$ its spectral norm. We note that the *condition number* $\mu(f,\zeta)$ is well defined for $(f,\zeta) \in \mathcal{H}_{\mathbf{d}} \times \mathbb{P}^n$.

Smale's α-theory [19] shows that, for z being an approximate zero of f associated with ζ, it is sufficient to have (cf. [13, 6])

$$d_{\mathbb{P}}(z,\zeta) \le \frac{0.3}{D^{3/2}\mu(f,\zeta)}.$$

2.4 Coarea Formula

Suppose that X, Y are Riemannian manifolds of dimensions m, n, respectively such that $m \ge n$. Let $\varphi\colon X \to Y$ be differentiable. By definition, the derivative $D\varphi(x)\colon T_xX \to T_{\varphi(x)}Y$ at a regular point $x \in X$ is surjective. Hence the restriction of $D\varphi(x)$ to the orthogonal complement of its kernel yields a linear isomorphism. The absolute value of its determinant is called the *normal Jacobian* of φ at x and denoted $\mathrm{NJ}\varphi(x)$. We set $\mathrm{NJ}\varphi(x) := 0$ if x is not a regular point. We note that the fiber $F_y := \varphi^{-1}(y)$ is a Riemannian submanifold of X of dimension $m-n$ if y is a regular value of φ. Sard's lemma states that almost all $y \in Y$ are regular values.

The following result is the *coarea formula*, sometimes also called Fubini's Theorem for Riemannian manifolds. It tells us how probability distributions on Riemannian manifolds transform. A proof can be found e.g., in [9, Appendix].

PROPOSITION 2.2. *Suppose that X, Y are Riemannian manifolds of dimensions m, n, respectively, and let $\varphi\colon X \to Y$ be a surjective differentiable map. Put $F_y = \varphi^{-1}(y)$. Then we have for any function $\chi\colon X \to \mathbb{R}$ that is integrable with respect to the volume measure of X that*

$$\int_X \chi\, dX = \int_{y\in Y} \left(\int_{F_y} \frac{\chi}{\mathrm{NJ}\varphi}\, dF_y \right) dY.$$

3. STATEMENT OF MAIN RESULTS

3.1 The Homotopy Continuation Routine ALH

Suppose that we are given an input system $f \in \mathcal{H}_{\mathbf{d}}$ and an initial pair (g,ζ) in the solution variety $V_{\mathbb{P}}$ such that f and g are $\mathbb{R}$-linearly independent. Let $\alpha = d_{\mathbb{S}}(f,g)$ and consider the line segment $E_{f,g}$ in $\mathcal{H}_{\mathbf{d}}$ with endpoints f and g. We parameterize this segment by writing $E_{f,g} = \{q_\tau \in \mathcal{H}_{\mathbf{d}} \mid \tau \in [0,1]\}$ with q_τ being the only point in $E_{f,g}$ such that $d_{\mathbb{S}}(g,q_\tau) = \tau\alpha$. If $E_{f,g}$ does not intersect the discriminant variety Σ, there is a unique continuous map $\gamma\colon [0,1] \to V_{\mathbb{P}}, \tau \mapsto (q_\tau,\zeta_\tau)$ such that $(q_0,\zeta_0) = (g,\zeta)$, called the *lifting* of $E_{f,g}$ with origin (g,ζ).

The idea is to follow the path γ numerically: we adaptively choose a partition $\tau_0 = 0, \tau_1, \ldots, \tau_k = 1$ and, writing $q_i := q_{\tau_i}$ and $\zeta_i := \zeta_{t_i}$, we successively compute approximations z_i of ζ_i by Newton's method starting with $z_0 := \zeta$. More specifically, we compute

$$z_{i+1} := N_{q_{i+1}}(z_i).$$

More precisely, we consider the following algorithm ALH (Adaptive Linear Homotopy) following an idea in Shub [13]. The stepsize parameter is set to $\lambda = 7.53 \cdot 10^{-3}$.

Algorithm ALH
input $f, g \in \mathcal{H}_{\mathbf{d}}$ and $\zeta \in \mathbb{P}^n$ such that $g(\zeta) = 0$
  $\alpha := d_{\mathbb{S}}(f,g)$, $r := \|f\|$, $s := \|g\|$
  $\tau := 0$, $q := g$, $z := \zeta$
  `repeat`
    $\Delta\tau := \frac{\lambda}{\alpha D^{3/2}\mu(q,z)^2}$
    $\tau := \min\{1, \tau + \Delta\tau\}$
    $t := \frac{s}{r\sin\alpha\cot(\tau\alpha) - r\cos\alpha + s}$
    $q := tf + (1-t)g$
    $z := N_q(z)$
  `until` $\tau = 1$
  `RETURN` z

In [13, 6] the following analysis of ALH was provided.

THEOREM 3.1. *The number of iterations $K(f,g,\zeta)$ of algorithm ALH in input f, g, ζ is bounded as*

$$K(f,g,\zeta) \;\le\; 217\, D^{3/2}\, d_{\mathbb{S}}(f,g) \int_0^1 \mu(q_\tau,\zeta_\tau)^2\, d\tau.$$

The returned point z is an approximate zero of f with associated zero ζ_1.

3.2 Randomization and Complexity: the Algorithm LV

ALH will serve as the basic routine for a number of algorithms computing zeros of polynomial systems in different

contexts. In these contexts both the input system f and the origin (g,ζ) of the homotopy may be randomly chosen: in the case of (g,ζ) as a computational technique and in the case of f in order to perform a probabilistic analysis of the algorithm's running time.

In both cases, a probability measure is needed: one for f and one for the pair (g,ζ). The measure for f will depend on the kind of probabilistic analysis (standard average-case or smoothed analysis) we perform. In contrast, we will consider only one measure on $V_{\mathbb{P}}$—which we denote by $\rho_{\mathbf{st}}$— for the initial pair (g,ζ). It consists of drawing g from $\mathcal{H}_{\mathbf{d}}$ from the standard Gaussian distribution (defined via the isomorphism $\mathcal{H}_{\mathbf{d}} \simeq \mathbb{R}^{2N}$ given by the Bombieri-Weyl basis) and then choosing one of the (almost surely) $\mathcal{D}$ zeros of g from the uniform distribution on $\{1,\dots,\mathcal{D}\}$. The above procedure is clearly non-constructive as computing a zero of a system is the problem we wanted to solve in the first place. One of the major contributions in [1] was to show that this drawback can be repaired.

PROPOSITION 3.2. *We can compute a random pair $(g,\zeta) \in V_{\mathbb{P}}$ according to the density $\rho_{\mathbf{st}}$ with $\mathcal{O}(N)$ choices of random real numbers from the standard Gaussian distribution and $\mathcal{O}(DnN+n^3)$ arithmetic operations (including square roots of positive numbers).*

Consider the following *Las Vegas algorithm* LV:

input $f \in \mathcal{H}_{\mathbf{d}}$
  `draw` $(g,\zeta) \in V_{\mathbb{P}}$ `from` $\rho_{\mathbf{st}}$
  `run` ALH `on input` (f,g,ζ)

For an input $f \in \mathcal{H}_{\mathbf{d}}$ algorithm LV either outputs an approximate zero z of f or loops forever. We write $t(f) := \mathbb{E}_{(g,\zeta)\sim\rho_{\mathbf{st}}}(t(f,g,\zeta))$ for the *expected running time* of LV on input f.

For all f,g,ζ_0, the running time $t(f,g,\zeta)$ is given by the *number of iterations* $K(f,g,\zeta)$ of ALH with input this triple, times the cost of an iteration, the latter being dominated by that of computing one Newton iterate (which is $\mathcal{O}(N+n^3)$ independently of the triple (f,g,ζ)). It therefore follows that analyzing the expected running time of LV amounts to do so for the expected value of $K(f,g,\zeta)$, over $(g,\zeta) \in V_{\mathbb{P}}$ drawn from $\rho_{\mathbf{st}}$. We denote this expectation by

$$K(f) := \mathop{\mathbb{E}}_{(g,\zeta)\sim\rho_{\mathbf{st}}}(K(f,g,\zeta)).$$

Beltrán and Pardo [3] performed an average-case analysis of LV showing that

$$E_{f\in S(\mathcal{H}_{\mathbf{d}})} K(f) = \mathcal{O}(D^{3/2}Nn).$$

3.3 Smoothed Analysis of LV

A smoothed analysis of an algorithm consists of bounding, for all possible input data $\overline{f}$, the average of its running time (its expected running time if it is a Las Vegas algorithm) over small perturbations of $\overline{f}$. For the latter, for technical simplicity, we assume a truncated Gaussian defined as follows. For $\overline{f} \in \mathcal{H}_{\mathbf{d}}$ and $\sigma > 0$ we shall denote by $N(\overline{f},\sigma^2\mathrm{I})$ the Gaussian distribution on $\mathcal{H}_{\mathbf{d}}$ with mean $\overline{f}$ and covariance matrix $\sigma^2\mathrm{I}$ (defined with respect to the Bombieri-Weyl basis). Further, let $A = \sqrt{2N}$ and $P_{A,\sigma} := \mathrm{Prob}\{\|f\| \le A \mid f \sim N(0,\sigma^2\mathrm{I})\}$ (it is known that $P_{A,\sigma} \ge \frac{1}{2}$ for all $\sigma \le 1$, cf. [8]). The *truncated Gaussian* $N_A(\overline{f},\sigma^2\mathrm{I})$ with center $\overline{f} \in \mathcal{H}_{\mathbf{d}}$ is the probability measure on $\mathcal{H}_{\mathbf{d}}$ with density

$$\rho(f) = \begin{cases} \frac{\rho_{\overline{f},\sigma}(f)}{P_{A,\sigma}} & \text{if } \|f-\overline{f}\| \le A \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\rho_{\overline{f},\sigma}$ denotes the density of $N(\overline{f},\sigma^2\mathrm{I})$. We now state our smoothed analysis result for LV.

THEOREM 3.3. *For any $0 < \sigma \le 1$, Algorithm* LV *satisfies*

$$\sup_{\overline{f}\in S(\mathcal{H}_{\mathbf{d}})} \mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} K(f) \le 3707\, D^{3/2}\big(N+\sqrt{\frac{N}{2}}\big)(n+1)\,\frac{1}{\sigma}.$$

3.4 The Main Technical Result

The technical heart of the proof of Theorem 3.3 is the following smoothed analysis of the *mean square condition number* $\mu_2(q)$ of $q \in \mathcal{H}_{\mathbf{d}}$, defined as

$$\mu_2(q) \;:=\; \Big(\frac{1}{\mathcal{D}} \sum_{\zeta|q(\zeta)=0} \mu(q,\zeta)^2\Big)^{1/2}.$$

THEOREM 3.4. *Let $\overline{q} \in \mathcal{H}_{\mathbf{d}}$ and $\sigma > 0$. For $q \in \mathcal{H}_{\mathbf{d}}$ drawn from $N(\overline{q},\sigma^2\mathrm{I})$ we have*

$$\mathop{\mathbb{E}}_{\mathcal{H}_{\mathbf{d}}}\Big(\frac{\mu_2(q)^2}{\|q\|^2}\Big) \;\le\; \frac{e(n+1)}{2\sigma^2}.$$

3.5 Condition-based Analysis of LV

We are here interested in estimating $K(f)$ for a fixed input system $f \in S(\mathcal{H}_{\mathbf{d}})$. Such an estimate will have to depend on, besides N, n, and D, the condition of f. We measure the latter using Shub and Smale's [14] condition number $\mu_{\max}(f)$ defined by $\mu_{\max}(f) := \max_{\zeta|f(\zeta)=0}\mu(f,\zeta)$. Our condition-based analysis of LV is summarized as follows.

THEOREM 3.5. *The expected number of iterations of Algorithm* LV *with input $f \in S(\mathcal{H}_{\mathbf{d}}) \setminus \Sigma$ is bounded as*

$$K(f) \le 157109\, D^3 N(n+1)\mu_{\max}(f)^2.$$

3.6 A Near Solution of Smale's 17th Problem

We finally want to consider *deterministic* algorithms finding approximate zeros of polynomial systems. Our goal is to exhibit one such algorithm working in nearly-polynomial average time, more precisely in average time $N^{\mathcal{O}(\log\log N)}$. A first ingredient to do so is a deterministic homotopy algorithm which is fast when D is small. This consists of algorithm ALH plus the initial pair $(\overline{U},\boldsymbol{z}_1)$, where $\overline{U} = (\overline{U}_1,\dots,\overline{U}_n) \in S(\mathcal{H}_{\mathbf{d}})$ with $\overline{U}_i = \frac{1}{\sqrt{2n}}(X_0^{d_i} - X_i^{d_i})$ and $\boldsymbol{z}_1 = (1:1:\dots:1)$.

We consider the following algorithm MD (Moderate Degree):

input $f \in \mathcal{H}_{\mathbf{d}}$
  `run` ALH `on input` $(f,\overline{U},\boldsymbol{z}_1)$

We write $K_{\overline{U}}(f) := K(f,\overline{U},\boldsymbol{z}_1)$ for the number of iterations of algorithm MD with input f. We are interested in computing the average over f of $K_{\overline{U}}(f)$ for f randomly chosen in $S(\mathcal{H}_{\mathbf{d}})$ from the uniform distribution.

THEOREM 3.6. *The average number of iterations of Algorithm* MD *is bounded as*

$$\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) \leq 314217\, D^3\, N(n+1)^{D+1}.$$

Algorithm MD is efficient when D is small, say, when $D \leq n$. For $D > n$ we use another approach, namely, a real number algorithm designed by Jim Renegar [10] which in this case has a performance similar to that of MD when $D \leq n$. Putting both pieces together we reach our last main result.

THEOREM 3.7. *There is a* deterministic *real number algorithm that on input $f \in \mathcal{H}_{\mathbf{d}}$ computes an approximate zero of f in average time $N^{\mathcal{O}(\log\log N)}$, where $N = \dim \mathcal{H}_{\mathbf{d}}$ measures the size of the input f. Moreover, if we restrict data to polynomials satisfying*

$$D \leq n^{\frac{1}{1+\varepsilon}} \quad \text{or} \quad D \geq n^{1+\varepsilon},$$

for some fixed $\varepsilon > 0$, then the average time of the algorithm is polynomial in the input size N.

4. SMOOTHED ANALYSIS OF LV (PROOF)

We first state a consequence of Theorem 3.1.

PROPOSITION 4.1. *The expected number of iterations of* ALH *on input $f \in \mathcal{H}_{\mathbf{d}} \setminus \Sigma$ is bounded as*

$$K(f) \leq 217\, D^{3/2} \mathop{\mathbb{E}}_{g\sim N(0,\mathrm{I})} \left(d_{\mathbb{S}}(f,g) \int_0^1 \mu_2(q_\tau)^2\, d\tau \right).$$

PROOF. Fix $g \in \mathcal{H}_{\mathbf{d}}$ such that the segment $E_{f,g}$ does not intersect the discriminant variety Σ (which is the case for almost all g, as $f \notin \Sigma$). To each of the zeros $\zeta^{(i)}$ of g there corresponds a lifting $[0,1] \to V, \tau \mapsto (q_\tau, \zeta_\tau^{(i)})$ of $E_{f,g}$ such that $\zeta_0^{(i)} = \zeta^{(i)}$. Theorem 3.1 states that

$$K(f,g,\zeta^{(i)}) \leq 217\, D^{3/2}\, d_{\mathbb{S}}(f,g) \int_0^1 \mu(q_\tau, \zeta_\tau^{(i)})^2\, d\tau.$$

Since $\zeta_\tau^{(1)}, \ldots, \zeta_\tau^{(\mathcal{D})}$ are the zeros of q_τ, we have by the definition of $\mu_2(q_\tau)$,

$$\frac{1}{\mathcal{D}} \sum_{i=1}^{\mathcal{D}} K(f,g,\zeta^{(i)}) \leq 217\, D^{3/2}\, d_{\mathbb{S}}(f,g) \int_0^1 \mu_2(q_\tau)^2\, d\tau. \quad (3)$$

Taking into account that

$$K(f) = \mathop{\mathbb{E}}_{(g,\zeta)\sim\rho_{\mathbf{st}}} (K(f,g,\zeta)) = \mathop{\mathbb{E}}_{g\sim N(0,\mathrm{I})} \Big(\frac{1}{\mathcal{D}} \sum_{i\leq\mathcal{D}} K(f,g,\zeta^{(i)}) \Big).$$

the assertion follows. □

All our main results involve expectations —over random f and/or g— of the integral $\int_0^1 \mu_2(q_\tau)^2\, d\tau$. In all cases, we will eventually deal with such an expectation with f and g Gaussian. Since a linear combination (with fixed coefficients) of two such Gaussian systems is Gaussian as well, it is convenient to parameterize the interval $E_{f,g}$ by a parameter $t \in [0,1]$ representing a ratio of Euclidean distances (instead of a ratio of angles as τ does). Thus we write, abusing notation, $q_t = tf + (1-t)g$. For fixed t, as noted before, q_t follows a Gaussian law.

For this new parametrization we have the following result, whose elementary proof is omitted.

PROPOSITION 4.2. *Let $f, g \in \mathcal{H}_{\mathbf{d}}$ be $\mathbb{R}$-linearly independent and $\tau_0 \in [0,1]$. Then*

$$d_{\mathbb{S}}(f,g) \int_{\tau_0}^1 \mu_2(q_\tau)^2\, d\tau \leq \int_{t_0}^1 \|f\|\, \|g\|\, \frac{\mu_2(q_t)^2}{\|q_t\|^2}\, dt,$$

where $\alpha = d_{\mathbb{S}}(f,g)$ and

$$t_0 = \frac{\|g\|}{\|g\| + \|f\|(\sin\alpha\cot(\tau_0\alpha) - \cos\alpha)}$$

is the fraction of the Euclidean distance $\|f-g\|$ corresponding to the fraction τ_0 of α.

As we will see, the factor $\|f\|\,\|g\|$ can be easily bounded and factored out the expectation. We will ultimately face the problem of estimating expectations of $\mu_2(q_t)^2/\|q_t\|^2$, where q_t follows a Gaussian law. This is provided by the crucial Theorem 3.4, whose sketch of proof is postponed to §5.

PROOF OF THEOREM 3.3. Fix $\overline{f} \in S(\mathcal{H}_{\mathbf{d}})$. We use Proposition 4.1 to obtain

$$\mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} K(f) \leq$$
$$217\, D^{3/2} \mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} \mathop{\mathbb{E}}_{g\sim N_A(0,\mathrm{I})} \left(d_{\mathbb{S}}(f,g) \int_0^1 \mu_2(q_\tau)^2\, d\tau \right).$$

This follows from the fact that, since both $d_{\mathbb{S}}(f,g)$ and $\mu_2(q_\tau)^2$ are homogeneous of degree 0 in both f and g, we may replace the Gaussian $N(0,\mathrm{I})$ by any rotationally invariant distribution on $\mathcal{H}_{\mathbf{d}}$, in particular by the centered truncated Gaussian $N_A(0,\mathrm{I})$ defined in (2). Now we use Proposition 4.2 (with $\tau_0 = 0$) and $\|f\| \leq \|\overline{f}\| + \|f-\overline{f}\| \leq 1 + A$ to get

$$\mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} K(f) \leq \quad (4)$$
$$217 D^{3/2}(A+1)A \mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} \mathop{\mathbb{E}}_{g\sim N_A(0,\mathrm{I})} \left(\int_0^1 \frac{\mu_2(q_t)^2}{\|q_t\|^2}\, dt \right).$$

Denoting by $\rho_{\overline{f},\sigma}$ and $\rho_{0,1}$ the densities of $N(\overline{f},\sigma^2\mathrm{I})$ and $N(0,\mathrm{I})$, respectively, the right-hand side of (4) equals

$$\begin{aligned}
&217\, D^{3/2} \frac{(A+1)A}{P_{A,\sigma} P_{A,1}} \cdot \\
&\qquad \int_{\|f-\overline{f}\|\leq A} \int_{\|g\|\leq A} \int_0^1 \frac{\mu_2(q_t)^2}{\|q_t\|^2}\, \rho_{0,1}(g)\, \rho_{\overline{f},\sigma}(f)\, dt\, dg\, df \\
&\leq 217 D^{3/2} \frac{(A+1)A}{P_{A,\sigma} P_{A,1}} \mathop{\mathbb{E}}_{f\sim N(\overline{f},\sigma^2\mathrm{I})} \mathop{\mathbb{E}}_{g\sim N(0,\mathrm{I})} \left(\int_0^1 \frac{\mu_2(q_t)^2}{\|q_t\|}\, dt \right) \\
&= 217 D^{3/2} \frac{(A+1)A}{P_{A,\sigma} P_{A,1}} \int_0^1 \mathop{\mathbb{E}}_{q_t\sim N(\overline{q}_t,\sigma_t^2\mathrm{I})} \left(\frac{\mu_2(q_t)^2}{\|q_t\|} \right) dt,
\end{aligned}$$

where the last equality follows from the fact that, for fixed t, the random polynomial system $q_t = tf + (1-t)g$ has a Gaussian distribution with law $N(\overline{q}_t, \sigma_t^2\mathrm{I})$ (with $\overline{q}_t = t\overline{f}$ and $\sigma_t^2 = (1-t)^2 + \sigma^2 t^2$). We now apply Theorem 3.4 to deduce

$$\begin{aligned}
&\int_0^1 \mathop{\mathbb{E}}_{q_t\sim N(\overline{q}_t,\sigma_t^2\mathrm{I})} \left(\frac{\mu_2(q_t)^2}{\|q_t\|^2} \right) dt \\
&\qquad \leq \frac{e(n+1)}{2} \int_0^1 \frac{dt}{(1-t)^2 + \sigma^2 t^2} = \frac{e\pi(n+1)}{4\sigma}.
\end{aligned}$$

Consequently, using that $\frac{(A+1)A}{P_{A,\sigma}P_{A,1}} \leq 4(2N + \sqrt{2N})$, we get

$$\mathop{\mathbb{E}}_{f\sim N_A(\overline{f},\sigma^2\mathrm{I})} K(f) \leq 217\, D^{3/2}\, 4(2N + \sqrt{2N})\, \frac{e\pi(n+1)}{4\sigma}. \quad \square$$

5. PROOF OF THEOREM 3.4

5.1 Reduction to Matrix Condition Number

We distinguish points $[\zeta] \in \mathbb{P}^n$ from their representatives ζ in the sphere $\mathbb{S}^n := \{\zeta \in \mathbb{C}^{n+1} \mid \|\zeta\| = 1\}$. Note that $[\zeta] \cap \mathbb{S}^n$ is a circle with radius one. We work with the "lifting"

$$V := \{(q, \zeta) \in \mathcal{H}_{\mathbf{d}} \times \mathbb{S}^n \mid q(\zeta) = 0\}$$

of the solution variety $V_{\mathbb{P}}$, which is a vector bundle over $\mathbb{S}^n$ with respect to the projection $\pi_2 \colon V \to \mathbb{S}^n, (q, \zeta) \mapsto \zeta$.

For $\zeta \in \mathbb{S}^n$ we consider the following subspace R_ζ of $\mathcal{H}_{\mathbf{d}}$ consisting of systems h that vanish at ζ of higher order:

$$R_\zeta := \{h \in \mathcal{H}_{\mathbf{d}} \mid h(\zeta) = 0, Dh(\zeta) = 0\}.$$

We further decompose the orthogonal complement $R_\zeta^\perp$ of R_ζ in $\mathcal{H}_{\mathbf{d}}$ (defined with respect to the Bombieri-Weyl Hermitian inner product). Let L_ζ denote the subspace of $R_\zeta^\perp$ consisting of the systems vanishing at ζ and let C_ζ denote its orthogonal complement in $R_\zeta^\perp$. Then we have an orthogonal decomposition

$$\mathcal{H}_{\mathbf{d}} = C_\zeta \oplus L_\zeta \oplus R_\zeta \tag{5}$$

parameterized by $\zeta \in \mathbb{S}^n$. In fact, this can be interpreted as an orthogonal decomposition of the trivial Hermitian vector bundle $\mathcal{H}_{\mathbf{d}} \times \mathbb{S}^n \to \mathbb{S}^n$ into subbundles C, L, and R over $\mathbb{S}^n$. Moreover, the vector bundle V is the orthogonal sum of L and R: we have $V_\zeta = L_\zeta \oplus R_\zeta$ for all ζ.

The following lemma is easily verified at $\zeta = (1, 0, \ldots, 0)$. (By unitary invariance this is sufficient.)

LEMMA 5.1. *The space C_ζ consists of $(c_i \langle X, \zeta\rangle^{d_i})$ with $c_i \in \mathbb{C}$. The space L_ζ consists of the systems*

$$g = (\sqrt{d_i}\, \langle X, \zeta\rangle^{d_i - 1} \ell_i)$$

where ℓ_i is a linear form vanishing at ζ. Moreover, if $\ell_i = \sum_{j=0}^n m_{ij} X_j$ with $M = (m_{ij})$, then $\|g\| = \|M\|_F$.

Let $\mathscr{M}$ denote the space $\mathbb{C}^{n \times (n+1)}$ of matrices. In the special case, where all the degrees d_i are one, the solution manifold V specializes to the manifold

$$W := \big\{\big(M, \zeta\big) \in \mathscr{M} \times \mathbb{S}^n \mid M\zeta = 0\big\}$$

and π_2 specializes to the vector bundle $p_2 \colon W \to \mathbb{S}^n, (M, \zeta) \mapsto \zeta$ with the fibers

$$W_\zeta := \{M \in \mathscr{M} \mid M\zeta = 0\}.$$

Lemma 5.1 implies that we have *isometrical* linear maps

$$W_\zeta \to L_\zeta,\ M \mapsto g_{M,\zeta} := \big(\sqrt{d_i}\, \langle X, \zeta\rangle^{d_i - 1} \textstyle\sum_j m_{ij} X_j\big), \tag{6}$$

where $M = (m_{ij})$. In other words, the Hermitean vector bundles W and L over $\mathbb{S}^n$ are isometric.

We compose the orthogonal bundle projection $V_\zeta = L_\zeta \oplus R_\zeta \to L_\zeta$ with the bundle isometry $L_\zeta \simeq W_\zeta$ obtaining the map of vector bundles

$$\Psi \colon V \to W,\ (g_{M,\zeta} + h, \zeta) \mapsto (M, \zeta) \tag{7}$$

whose fibers $\Psi^{-1}(M, \zeta)$ are isometric to R_ζ. Another characterization of Ψ is $\Psi(q, \zeta) = (M, \zeta)$, where $M = \Delta^{-1} D g_{M,\zeta}(\zeta)$. This shows that Ψ provides the link to the condition number: by the definition (1) we have

$$\frac{\mu(q, \zeta)}{\|q\|} \;=\; \|M^\dagger\|, \quad \text{where } (M, \zeta) = \Psi(q, \zeta). \tag{8}$$

Let $\rho_{\mathcal{H}_{\mathbf{d}}}$ denote the density of the Gaussian $N(\overline{q}, \sigma^2 \mathrm{I})$ on $\mathcal{H}_{\mathbf{d}}$, where $\overline{q} \in \mathcal{H}_{\mathbf{d}}$ and $\sigma > 0$. For fixed $\zeta \in \mathbb{S}^n$ we decompose the mean $\overline{q}$ as

$$\overline{q} = \overline{k}_\zeta + \overline{g}_\zeta + \overline{h}_\zeta \in C_\zeta \oplus L_\zeta \oplus R_\zeta$$

according to (5). If we denote by ρ_{C_ζ}, ρ_{L_ζ}, and ρ_{R_ζ} the densities of the Gaussian distributions in the spaces C_ζ, L_ζ, and R_ζ with covariance matrices $\sigma^2 \mathrm{I}$ and means $\overline{k}_\zeta, \overline{M}_\zeta$, and $\overline{h}_\zeta$, respectively, then the density $\rho_{\mathcal{H}_{\mathbf{d}}}$ factors as

$$\rho_{\mathcal{H}_{\mathbf{d}}}(k + g + h) = \rho_{C_\zeta}(k) \cdot \rho_{L_\zeta}(g) \cdot \rho_{R_\zeta}(h). \tag{9}$$

The Gaussian density ρ_{L_ζ} on L_ζ induces a Gaussian density ρ_{W_ζ} on the fiber W_ζ with the covariance matrix $\sigma^2 \mathrm{I}$ via the isometrical linear map (6), so that we have $\rho_{W_\zeta}(M) = \rho_{L_\zeta}(g_{M,\zeta})$.

By Bézout's theorem, the fiber $V(q)$ of the projection $\pi_1 \colon V \to \mathcal{H}_{\mathbf{d}}$ at $q \in \mathcal{H}_{\mathbf{d}}$ is a disjoint union of $\mathcal{D} = d_1 \cdots d_n$ unit circles and therefore has the volume $2\pi\mathcal{D}$, provided q does not lie in the discriminant variety.

Think now of choosing (q, ζ) at random from V by first choosing $q \in \mathcal{H}_{\mathbf{d}}$ from $N(\overline{q}, \sigma^2 \mathrm{I})$, then choosing one of its $\mathcal{D}$ zeros $[\zeta] \in \mathbb{P}^n$ at random from the uniform distribution on $\{1, \ldots, \mathcal{D}\}$, and finally choosing a representative ζ in the unit circle $[\zeta] \cap \mathbb{S}^n$ uniformly at random. The resulting probability density ρ_V on V is a natural extension of $\rho_{\mathbf{st}}$. With the coarea formula (Proposition 2.2) one can show that ρ_V has the following form

$$\rho_V(q, \zeta) := \frac{1}{2\pi\mathcal{D}}\, \rho_{\mathcal{H}_{\mathbf{d}}}(q)\, \mathrm{NJ}\pi_1(q, \zeta). \tag{10}$$

The plan to show Theorem 3.4 is now as follows. We have

$$\mathop{\mathbb{E}}_{\mathcal{H}_{\mathbf{d}}} \Big(\frac{\mu_2(q)^2}{\|q\|^2}\Big) \;=\; \mathop{\mathbb{E}}_{V} \Big(\frac{\mu(q, \zeta)^2}{\|q\|^2}\Big), \tag{11}$$

where $\mathbb{E}_{\mathcal{H}_{\mathbf{d}}}$ and $\mathbb{E}_V$ refer to the expectations with respect to the distribution $N(\overline{q}, \sigma^2 \mathrm{I})$ on $\mathcal{H}_{\mathbf{d}}$ and the probability density ρ_V on V, respectively. Moreover, by Equation (8),

$$\mathop{\mathbb{E}}_{V} \Big(\frac{\mu(q, \zeta)^2}{\|q\|^2}\Big) \;=\; \mathop{\mathbb{E}}_{\mathscr{M}} \big(\|M^\dagger\|^2\big),$$

where $\mathbb{E}_{\mathscr{M}}$ denotes the expectation with respect to the *push-forward density* $\rho_{\mathscr{M}}$ of ρ_V with respect to the map

$$\psi := p_1 \circ \Psi \colon V \to \mathscr{M},\ (q, \zeta) \mapsto M = \Delta^{-1} D q(\zeta).$$

PROPOSITION 5.2. *1. For $M \in \mathscr{M}$ of rank n and $\zeta \in \mathbb{S}^n$ with $M\zeta = 0$ we have*

$$\rho_{\mathscr{M}}(M) = \rho_{C_\zeta}(0) \cdot \frac{1}{2\pi} \int_{\lambda \in S^1} \rho_{W_{\lambda\zeta}}(M)\, dS^1. \tag{12}$$

2. If we put $c_\zeta := \mathbb{E}_{M \sim \rho_{W_\zeta}}(\det(MM^*))$, then $\rho_{\mathbb{S}^n}(\zeta) = \frac{c_\zeta}{2\pi}\, \rho_{C_\zeta}(0)$ is a probability density on $\mathbb{S}^n$ and

$$\widetilde{\rho}_{W_\zeta}(M) := c_\zeta^{-1}\, \rho_{W_\zeta}(M)\, \det(MM^*)$$

defines a probability density on W_ζ.

3. The expectation of $\|M^\dagger\|^2$ with respect to the density $\rho_{\mathscr{M}}$ satisfies

$$\mathop{\mathbb{E}}_{M \sim \rho_{\mathscr{M}}} \big(\|M^\dagger\|^2\big) \;=\; \mathop{\mathbb{E}}_{\zeta \sim \rho_{\mathbb{S}^n}} \Big(\mathop{\mathbb{E}}_{M \sim \widetilde{\rho}_{W_\zeta}} \big(\|M^\dagger\|^2\big)\Big).$$

The formal proof of Proposition 5.2 is quite involved, and will be given in §5.2. However, Equation (12) has an easy heuristic explanation. We decompose a random $q \in \mathcal{H}_{\mathbf{d}}$ according to the decomposition $\mathcal{H}_{\mathbf{d}} = C_\zeta \oplus L_\zeta \oplus R_\zeta$ as $q = k + g + h$. Choose $\lambda \in \mathbb{C}$ with $|\lambda| = 1$ uniformly at random in the unit circle. Then we have $\Psi(q, \lambda\zeta) = (M, \lambda\zeta)$ iff $k = 0$ and g is mapped to M under the isometry in (6). The probability density for the event $k = 0$ equals $\rho_{C_\zeta}(0)$. The second event, conditioned on λ, has the probability density $\rho_{W_{\lambda\zeta}}(M)$.

For the proof of Theorem 3.4, according to Prop. 5.2(3), it is sufficient to show that for all ζ

$$\mathop{\mathbb{E}}_{M\sim\widetilde{\rho}_{W_\zeta}} \left(\|M^\dagger\|^2\right) \leq \frac{e(n+1)}{2\sigma^2}. \tag{13}$$

This is achieved in §5.3. We note that the density $\widetilde{\rho}_{W_\zeta}$ on the fiber W_ζ is closely related to a Gaussian. By unitary invariance we may assume that $\zeta = (1, 0, \ldots, 0)$. The matrices $M \in W_\zeta$ can then be identified with square matrices $A \in \mathbb{C}^{n\times n}$ and $M^\dagger$ corresponds to A^{-1}.

5.2 Proof of Proposition 5.2

We have already seen that the condition number $\mu(q, \zeta)$ can be described in terms of the map Ψ introduced in (7). Beltrán and Pardo in [3] proved that the normal Jacobian of the related bundle map

$$\Phi\colon V \to W,\ (q, \zeta) \mapsto (N, \zeta) = (Dq(\zeta), \zeta),$$

is constant. This is a crucial observation.

PROPOSITION 5.3. $\mathrm{NJ}\Phi(q, \zeta) = \mathcal{D}^n$ *for all* $(q, \zeta) \in V$.

We also need the fact that $\mathrm{NJ}\pi_1(q, \zeta) = \mathrm{NJ}p_1(N, \zeta)$ where $N = Dq(\zeta)$ and $p_1\colon W \to \mathscr{M}, (N', \zeta) \mapsto N'$, cf. [15].

Using this fact, the factorization (9) of Gaussians, and (10), the density ρ_V can be written as

$$\rho_V(g_{M,\zeta} + h, \zeta) = \frac{1}{2\pi\mathcal{D}}\, \rho_{C_\zeta}(0)\, \rho_{L_\zeta}(g_{M,\zeta})\, \rho_{R_\zeta}(h)\, \mathrm{NJ}p_1(N, \zeta),$$

where $N = \Delta M$.

Applying the coarea formula (Proposition 2.2) to Φ and using $\mathrm{NJ}\Phi = \mathcal{D}^n$ (Proposition 5.3), we obtain for any measurable function $\chi\colon \mathscr{M} \to [0, \infty)$ that

$$\int_V (\chi \circ \psi)\, \rho_V\, dV = \frac{1}{\mathcal{D}^n} \int_{(N,\zeta)\in W} \chi(M)\, \frac{\rho_{C_\zeta}(0)}{2\pi\mathcal{D}}\, \rho_{L_\zeta}(g_{M,\zeta})\, \mathrm{NJ}p_1(N, \zeta)\, dW,$$

where we have used that $\int_{R_\zeta} \rho_{R_\zeta}\, dR_\zeta = 1$.

We next apply the coarea formula to the projection p_1. For $N \in \mathscr{M}$ of rank n let $\zeta_N \in \mathbb{P}^n$ be given by $N\zeta_N = 0$. The fiber $V(N) = p_1^{-1}(N)$ is a circle consisting of all representations ζ of ζ_N with norm 1. Recall that ρ_{C_ζ} and ρ_{L_ζ} are independent of the choice of the representation of ζ. Moreover, $\rho_{L_{\zeta_N}}(g_{M,\zeta}) = \rho_{W_\zeta}(M)$. We can thus express the above integral as

$$\frac{1}{\mathcal{D}^n} \int_{N\in\mathscr{M}} \chi(M)\, \frac{\rho_{C_{\zeta_N}}(0)}{\mathcal{D}}\, I(M)\, d\mathscr{M},$$

where

$$I(M) := \frac{1}{2\pi} \int_{\zeta\in V(N)} \rho_{W_\zeta}(M)\, dV(N).$$

We finally perform the linear change of variables $\mathscr{M} \to \mathscr{M}, N \mapsto M$, that has the Jacobian determinant $\mathcal{D}^{n+1}$, and conclude that (note $\zeta_N = \zeta_M$)

$$\int_V (\chi \circ \psi)\, \rho_V\, dV = \int_{M\in\mathscr{M}} \chi(M)\, \rho_{C_{\zeta_M}}(0)\, I(M)\, d\mathscr{M}.$$

This proves that $\rho_{\mathscr{M}}(M) = \rho_{C_{\zeta_M}}(0)\, I(M)$ is the pushforward density of ρ_V with respect to ψ and shows the first assertion of Proposition 5.2.

For the second assertion, we apply the coarea formula first to $p_1\colon W \to \mathscr{M}$ and then to the projection $p_2\colon W \to \mathbb{S}^n, (M, \zeta) \mapsto \zeta$ with the fibers W_ζ and obtain for any measurable function $\chi\colon \mathscr{M} \to [0, \infty)$

$$\begin{aligned} 2\pi \int_{\mathscr{M}} \chi\, d\mathscr{M} &= \int_W (\chi \circ p_1)\, \mathrm{NJ}p_1\, dW \\ &= \int_{\zeta\in\mathbb{S}^n} \Big(\int_{W_\zeta} \chi\, \frac{\mathrm{NJ}p_1}{\mathrm{NJ}p_2}\, dW_\zeta \Big)\, d\mathbb{S}^n. \end{aligned}$$

In [15] (see also [5, Section 13.2, Lemmas 2-3]) it was shown that

$$\frac{\mathrm{NJ}p_1}{\mathrm{NJ}p_2}(M, \zeta) = \det(MM^*).$$

Using this, we obtain

$$\int_{\mathscr{M}} \chi\, d\mathscr{M} = \int_{\zeta\in\mathbb{S}^n} \Big(\int_{M\in W_\zeta} \frac{\chi(M)}{2\pi}\, \det(MM^*)\, d\mathscr{M}_\zeta \Big)\, d\mathbb{S}^n. \tag{14}$$

Applying this to $\chi = \rho_{\mathscr{M}}$, exchanging the order of the integrals and using S^1-invariance, we get

$$1 = \int_{\mathscr{M}} \rho_{\mathscr{M}}\, d\mathscr{M}_\zeta = \int_{\zeta\in\mathbb{S}^n} \frac{c_\zeta}{2\pi}\, \rho_{C_\zeta}(0)\, d\mathbb{S}^n,$$

where $c_\zeta := \mathbb{E}_{M\sim\rho_{W_\zeta}}(\det(MM^*))$. Hence $\frac{1}{2\pi} c_\zeta \rho_{\mathbb{C}_\zeta}(0)$ defines a density on $\mathbb{S}^n$, showing the second assertion of Proposition 5.2. Finally, the third assertion follows by applying (14) to $\chi(M) = \|M^\dagger\|^2\, \rho_{\mathscr{M}}(M)$. □

5.3 Smoothed Analysis of a Matrix Condition Number

In the following we fix $\overline{A} \in \mathbb{C}^{n\times n}$, $\sigma > 0$ and denote by ρ the Gaussian density of $N(\overline{A}, \sigma^2\mathrm{I})$ on $\mathbb{C}^{n\times n}$. Moreover, we consider the related density

$$\widetilde{\rho}(A) = c^{-1}\, |\det A|^2\, \rho(A) \quad \text{where } c := \mathop{\mathbb{E}}_{A\sim\rho} (|\det A|^2).$$

The following result is akin to a smoothed analysis of the matrix condition number $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ (compare Sankar et al. [11, §3]) with respect to the probability densities $\widetilde{\rho}$ that are not Gaussian, but closely related to Gaussians.

PROPOSITION 5.4. *For any* $t > 0$ *we have*

$$\mathop{\mathrm{Prob}}_{A\sim\widetilde{\rho}} \Big\{ \|A^{-1}\| \geq t \Big\} \ \leq\ \frac{e^2(n+1)^2}{16\sigma^4}\, \frac{1}{t^4}.$$

This tail bound easily implies $\mathbb{E}_{A\sim\widetilde{\rho}} \left(\|A^{-1}\|^2\right) \leq \frac{e(n+1)}{2\sigma^2}$, which proves inequality (13) and thus completes the proof of Theorem 3.4.

Recall that $\mathbb{S}^{n-1}$ denotes the unit sphere in $\mathbb{C}^n$.

LEMMA 5.5. *For any $v \in \mathbb{S}^{n-1}$ and any $t > 0$ we have*

$$\operatorname*{Prob}_{A\sim\widetilde{\rho}}\left\{\|A^{-1}v\| \geq t\right\} \;\leq\; \frac{1}{4\sigma^4 t^4}.$$

PROOF. We first claim that, because of unitary invariance, we may assume that $v = e_n := (0,\ldots,0,1)$. To see this, take $S \in U(n)$ such that $v = Se_n$. Consider the isometric map $A \mapsto B = S^{-1}A$ which transforms the density $\widetilde{\rho}(A)$ to a density of the same form, namely

$$\widetilde{\rho}'(B) = \widetilde{\rho}(A) = c^{-1}|\det A|^2\rho(A) = c^{-1}|\det B|^2\rho'(B),$$

where $\rho'(B)$ denotes the density of $N(S^{-1}\overline{A}, \sigma^2\mathrm{I})$ and $c = \mathbb{E}_\rho(|\det A|^2) = \mathbb{E}_{\rho'}(|\det B|^2)$. Thus the assertion for e_n and random B (chosen from any isotropic Gaussian distribution) implies the assertion for v and A, noting that $A^{-1}v = B^{-1}e_n$. This proves the claim.

Let a_i denote the ith row of A. Almost surely, the rows $a_1,\ldots,a_{n-1}$ are linearly independent. We are going to characterize $\|A^{-1}e_n\|$ in a geometric way. Let S_n denote the span of $a_1,\ldots,a_{n-1}$ and denote by $a_n^\perp$ the orthogonal projection of a_n onto $S_n^\perp$. Consider $w := A^{-1}e_n$, which is the nth column of A^{-1}. Since $AA^{-1} = \mathrm{I}$ we have $\langle w, a_i\rangle = 0$ for $i = 1,\ldots,n-1$ and hence $w \in S_n^\perp$. Moreover, $\langle w, a_n\rangle = 1$, so $\|w\|\,\|a_n^\perp\| = 1$ and we arrive at

$$\|A^{-1}e_n\| = \frac{1}{\|a_n^\perp\|}. \tag{15}$$

Let $A_n \in \mathbb{C}^{(n-1)\times n}$ denote the matrix obtained from A by omitting a_n. We shall write $\mathsf{vol}(A_n) = \det(AA^*)^{1/2}$ for the $(n-1)$-dimensional volume of the parallelepiped spanned by the rows of A_n. Similarly, $|\det A|$ can be interpreted as the n-dimensional volume of the parallelepiped spanned by the rows of A.

Now we write $\rho(A) = \rho_1(A_n)\rho_2(a_n)$ where ρ_1 and ρ_2 are the density functions of $N(\overline{A}_n, \sigma^2\mathrm{I})$ and $N(\overline{a}_n, \sigma^2\mathrm{I})$, respectively (the meaning of $\overline{A}_n$ and $\overline{a}_n$ being clear). Moreover, note that

$$\mathsf{vol}(A)^2 = \mathsf{vol}(A_n)^2\,\|a_n^\perp\|^2.$$

Fubini's Theorem combined with (15) yields for $t > 0$

$$\begin{aligned}\int_{\|A^{-1}e_n\|\geq t} \mathsf{vol}(A)^2\rho(A)\,dA \;=\; & \\ \int_{A_n\in\mathbb{C}^{(n-1)\times n}} & \mathsf{vol}(A_n)^2\,\rho_1(A_n) \\ & \cdot\left(\int_{\|a_n^\perp\|\leq 1/t}\|a_n^\perp\|^2\rho_2(a_n)\,da_n\right)dA_n. \end{aligned}\tag{16}$$

We next show that for fixed, linearly independent $a_1,\ldots,a_{n-1}$ and $\lambda > 0$

$$\int_{\|a_n^\perp\|\leq\lambda}\|a_n^\perp\|^2\rho_2(a_n)\,da_n \leq \frac{\lambda^4}{2\sigma^2}. \tag{17}$$

For this, note that $a_n^\perp \sim N(\overline{a}_n^\perp, \sigma^2\mathrm{I})$ in $S_n^\perp \simeq \mathbb{C}$ where $\overline{a}_n^\perp$ is the orthogonal projection of $\overline{a}_n$ onto $S_n^\perp$. Thus, proving (17) amounts to showing

$$\int_{|z|\leq\lambda}|z|^2\rho_{\overline{z}}(z)dz \leq \frac{\lambda^4}{2\sigma^2}$$

for the Gaussian density $\rho_{\overline{z}}(z) = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}|z-\overline{z}|^2}$ of $z \in \mathbb{C}$, where $\overline{z} \in \mathbb{C}$. Clearly, it is enough to show that

$$\int_{|z|\leq\lambda}\rho_{\overline{z}}(z)dz \leq \frac{\lambda^2}{2\sigma^2}.$$

Without loss of generality we may assume that $\overline{z} = 0$, since the integral in the left-hand side is maximized at this value of $\overline{z}$. Then, writing $z = \sigma w$, we have

$$\begin{aligned}\int_{|z|\leq\lambda}\rho_0(z)dz &= \int_{|w|\leq\frac{\lambda}{\sigma}}\frac{1}{2\pi}e^{-\frac{1}{2}|w|^2}\,dw \\ &= \int_0^{\frac{\lambda}{\sigma}}\frac{1}{2\pi}e^{-\frac{1}{2}r^2}2\pi r\,dr \\ &= -e^{-\frac{1}{2}r^2}\Big|_0^{\frac{\lambda}{\sigma}} = 1 - e^{-\frac{\lambda^2}{2\sigma^2}} \leq \frac{\lambda^2}{2\sigma^2},\end{aligned}$$

which proves inequality (17).

A similar argument shows that

$$2\sigma^2 \;\leq\; \int|z|^2\rho_{\overline{z}}(z)dz = \int\|a_n^\perp\|^2\rho_2(a_n)\,da_n. \tag{18}$$

Plugging in this inequality into (16) (with $t = \infty$) we conclude that

$$2\sigma^2\;\mathop{\mathbb{E}}_{\rho_1}\left(\mathsf{vol}(A_n)^2\right) \;\leq\; \mathop{\mathbb{E}}_{\rho}\left(\mathsf{vol}(A)^2\right). \tag{19}$$

On the other hand, plugging in (17) with $\lambda = \frac{1}{t}$ into (16), we obtain

$$\int_{\|A^{-1}e_n\|\geq t}\mathsf{vol}(A)^2\rho(A)\,dA \;\leq\; \frac{1}{2\sigma^2t^4}\;\mathop{\mathbb{E}}_{\rho_1}\left(\mathsf{vol}(A_n)^2\right).$$

Combined with (19) this yields

$$\int_{\|A^{-1}e_n\|\geq t}\mathsf{vol}(A)^2\rho(A)\,dA \;\leq\; \frac{1}{4\sigma^4t^4}\;\mathop{\mathbb{E}}_{\rho}\left(\mathsf{vol}(A)^2\right).$$

By the definition of the density $\widetilde{\rho}$, this means that

$$\operatorname*{Prob}_{A\sim\widetilde{\rho}}\left\{\|A^{-1}e_n\| \geq t\right\} \;\leq\; \frac{1}{4\sigma^4t^4},$$

which was to be shown. □

The following is a consequence of [7, Lemma 2.1].

LEMMA 5.6. *For fixed $u \in \mathbb{S}^{n-1}$, $0 \leq s \leq 1$, and v uniformly chosen at random in $\mathbb{S}^{n-1}$, we have*

$$\operatorname*{Prob}_{v}\left\{|u^{\mathrm{T}}v| \geq s\right\} = (1-s^2)^{n-1}.$$

PROOF OF PROPOSITION 5.4. We use an idea in Sankar et al. [11, §3]. For any invertible $A \in \mathbb{C}^{n\times n}$ there exists $u \in \mathbb{S}^{n-1}$ such that $\|A^{-1}u\| = \|A^{-1}\|$. For almost all A, the vector u is uniquely determined up to a scaling factor θ of modulus 1. We shall denote by u_A a representative of such u.

The following is an easy consequence of the singular value decomposition of $\|A^{-1}\|$: for any $v \in \mathbb{S}^{n-1}$ we have

$$\|A^{-1}v\| \;\geq\; \|A^{-1}\|\cdot|u_A^{\mathrm{T}}v|. \tag{20}$$

We choose now a random pair (A, v) with A following the law $\widetilde{\rho}$ and, independently, $v \in \mathbb{S}^{n-1}$ from the uniform distribution. Lemma 5.5 implies that

$$\operatorname*{Prob}_{A,v}\left\{\|A^{-1}v\| \geq t\sqrt{\frac{2}{n+1}}\right\} \;\leq\; \frac{(n+1)^2}{16\sigma^4t^4}.$$

On the other hand, we have by (20)

$$\begin{aligned}
&\operatorname*{Prob}_{A,v}\left\{\|A^{-1}v\| \geq t\sqrt{2/(n+1)}\right\}\\
&\quad\geq \operatorname*{Prob}_{A,v}\left\{\|A^{-1}\| \geq t \text{ and } |u_A^{\mathrm{T}}\, v| \geq \sqrt{2/(n+1)}\right\}\\
&\quad\geq \operatorname*{Prob}_{A}\left\{\|A^{-1}\| \geq t\right\}\\
&\qquad\cdot \operatorname*{Prob}_{A,v}\left\{|u_A^{\mathrm{T}}\, v| \geq \sqrt{2/(n+1)} \;\Big|\; \|A^{-1}\| \geq t\right\}.
\end{aligned}$$

Lemma 5.6 tells us that for any fixed $u \in \mathbb{S}^{n-1}$ we have

$$\operatorname*{Prob}_{v}\left\{|u^{\mathrm{T}}\, v| \geq \sqrt{2/(n+1)}\right\} = (1-2/(n+1))^{n-1} \geq e^{-2},$$

the last inequality as $(\frac{n+1}{n-1})^{n-1} = (1+\frac{2}{n-1})^{n-1} \leq e^2$. We thus obtain

$$\begin{aligned}
\operatorname*{Prob}_{A}\left\{\|A^{-1}\| \geq t\right\} &\leq e^2 \operatorname*{Prob}_{A,v}\left\{\|A^{-1}v\| \geq t\sqrt{\frac{2}{n+1}}\right\}\\
&\leq \frac{e^2(n+1)^2}{16\sigma^4 t^4},
\end{aligned}$$

as claimed. □

6. SKETCH OF REMAINING PROOFS

6.1 Homotopies with a Fixed Extremity

The next two cases we wish to analyze (the condition-based analysis of LV and a solution for Smale's 17th problem with moderate degrees) share the feature that one endpoint of the homotopy segment is fixed, not randomized. This sharing actually allows one to derive both corresponding results (Theorems 3.5 and 3.6, respectively) as a consequence of the following statement.

THEOREM 6.1. *For $g \in S(\mathcal{H}_{\mathbf{d}}) \setminus \Sigma$ we have*

$$\begin{aligned}
&\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})}\left(d_{\mathbb{S}}(f,g)\int_0^1 \mu_2(q_\tau)^2\, d\tau\right)\\
&\quad\leq 724\, D^{3/2} N(n+1)\mu_{\max}(g)^2 + 0.01.
\end{aligned}$$

The idea to prove Theorem 6.1 is simple. For small values of τ the system q_τ is close to g and therefore, the value of $\mu_2(q_\tau)^2$ can be bounded by a small multiple of $\mu_{\max}(g)^2$. For the remaining values of τ, the corresponding $t = t(\tau)$ is bounded away from 0 and therefore so is the variance σ_t^2 in the distribution $N(\overline{q}_t, \sigma_t^2 \mathrm{I})$ for q_t. This allows one to control the denominator in the right-hand side of Theorem 3.4 when using this result.

6.2 Condition-based Analysis of LV (proof of Theorem 3.5)

It follows immediately by combining Proposition 4.1 with Theorem 6.1, with the roles of f and g swapped. □

6.3 The Complexity of a Deterministic Homotopy Continuation

We next prove Theorem 3.6. The unitary group $\mathcal{U}(n+1)$ naturally acts on $\mathbb{P}^n$ as well as on $\mathcal{H}_{\mathbf{d}}$ via $(\nu, f) \mapsto f \circ \nu^{-1}$. The following lemma results from the unitary invariance of our setting. The proof is immediate.

LEMMA 6.2. *Let $g \in \mathcal{H}_{\mathbf{d}}$, $\zeta \in \mathbb{P}^n$ be a zero of g, and $\nu \in \mathcal{U}(n+1)$. Then $\mu(g,\zeta) = \mu(g\circ\nu^{-1}, \nu\zeta)$. Moreover, for $f \in \mathcal{H}_{\mathbf{d}}$, we have $K(f,g,\zeta) = K(f\circ\nu^{-1}, g\circ\nu^{-1}, \nu\zeta)$.*

Denote by $z_{(i)}$ a d_ith primitive root of unity. The $\mathcal{D}$ zeros of $\overline{U} = (\overline{U}_1, \ldots, \overline{U}_n)$ are the points $\boldsymbol{z}_j = \big(1 : z_{(1)}^{j_1} : \ldots : z_{(n)}^{j_n}\big) \in \mathbb{P}^n$ for all the possible tuples $j = (j_1, \ldots, j_n)$ with $j_i \in \{0, \ldots, d_i - 1\}$. Clearly, each $\boldsymbol{z}_j$ can be obtained from $\boldsymbol{z}_1 := (1 : 1 : \ldots : 1)$ by a unitary transformation ν_j, which leaves $\overline{U}$ invariant, that is, $\nu_j \boldsymbol{z}_1 = \boldsymbol{z}_j$, $\overline{U}\circ\nu_j^{-1} = \overline{U}$. Hence Lemma 6.2 implies $\mu(\overline{U}, \boldsymbol{z}_j) = \mu(\overline{U}, \boldsymbol{z}_1)$ for all j. In particular, $\mu_{\max}(\overline{U}) = \mu(\overline{U}, \boldsymbol{z}_1)$. It also implies the following result.

PROPOSITION 6.3. *$K_{\overline{U}}(f) = K(f, \overline{U}, \boldsymbol{z}_1)$ satisfies*

$$\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) = \mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})}\Big(\frac{1}{\mathcal{D}}\sum_{j=1}^{\mathcal{D}} K(f, \overline{U}, \boldsymbol{z}_j)\Big).$$

The following lemma follows in a straightforward way from the definition of μ.

LEMMA 6.4. *We have*

$$\mu_{\max}(\overline{U})^2 \leq 2n \max_i \frac{1}{d_i}(n+1)^{d_i - 1} \leq 2\,(n+1)^D.$$

PROOF OF THEOREM 3.6. Equation (3) in the proof of Proposition 4.1 implies for $g = \overline{U}$ that

$$\frac{1}{\mathcal{D}}\sum_{i=1}^{\mathcal{D}} K(f, \overline{U}, \boldsymbol{z}_i) \leq 217\, D^{3/2}\, d_{\mathbb{S}}(f, \overline{U}) \int_0^1 \mu_2(q_\tau)^2\, d\tau.$$

Using Proposition 6.3 we get

$$\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) \leq 217\, D^{3/2} \mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})}\Big(d_{\mathbb{S}}(f, \overline{U}) \int_0^1 \mu_2(q_\tau)^2\, d\tau\Big).$$

Applying Theorem 6.1 with $g = \overline{U}$ we obtain

$$\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) \leq 217\, D^{3/2}\left(724\, D^{3/2} N(n+1)\, \mu_{\max}(\overline{U})^2 + 0.01\right).$$

We now plug in the bound $\mu_{\max}(\overline{U})^2 \leq 2(n+1)^D$ of Lemma 6.4 to obtain

$$\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) \leq 314216\, D^3\, N(n+1)^{D+1} + 2.17\, D^{3/2}.$$

This is bounded from above by $314217\, D^3\, N(n+1)^{D+1}$, which completes the proof. □

6.4 A Near Solution to Smale's 17th Problem

We finally proceed with the proof of Theorem 3.7. The algorithm we will exhibit uses different routines for $D \leq n$ and $D > n$.

6.4.1 The case $D \leq n$

Theorem 3.6 bounds the number of iterations of Algorithm MD as

$$\mathop{\mathbb{E}}_{f\in S(\mathcal{H}_{\mathbf{d}})} K_{\overline{U}}(f) = \mathcal{O}(D^3 N n^{D+1}).$$

For comparing the order of magnitude of this upper bound to the input size $N = \sum_{i=1}^n \binom{n+d_i}{n}$ we need the following technical lemma (which will be useful for the case $D > n$ as well).

LEMMA 6.5. *For $D \leq n$ we have $n^D \leq N^{2\ln\ln N+\mathcal{O}(1)}$.*

Theorem 3.6 combined with Lemma 6.5 implies that

$$\mathop{\mathbb{E}}_{f} K_{\overline{U}}(f) = N^{2\ln\ln N+\mathcal{O}(1)} \quad \text{if } D \leq n. \tag{21}$$

Note that this bound is nearly polynomial in N.

6.4.2 The case $D > n$

The homotopy continuation algorithm MD is not efficient for large degrees—the main problem being that we do not know how to deterministically compute a starting system g with small $\mu_{\max}(g)$. However, it turns out that an iterative procedure ItRen using an algorithm due to Jim Renegar [10] is fast for large degrees. We prove the following result.

PROPOSITION 6.6. *Let $T(f)$ denote the running time of algorithm* ItRen *on input f. Then, for standard Gaussian $f \in \mathcal{H}_{\mathbf{d}}$, $\mathbb{E}_f\, T(f) = (nN\mathcal{D})^{\mathcal{O}(1)}$.*

PROOF OF THEOREM 3.7. We use Algorithm MD if $D \leq n$ and Algorithm ItRen if $D > n$. We have already shown (see (21)) that the assertion holds if $D \leq n$. For the case $D > n$ we use Proposition 6.6 together with the inequality $\mathcal{D}^{\mathcal{O}(1)} \leq D^{\mathcal{O}(n)} \leq N^{\mathcal{O}(\log\log N)}$ which follows from Lemma 6.5 with the roles of D and n swapped. □

7. REFERENCES

[1] C. Beltrán and L. M. Pardo. On Smale's 17th problem: a probabilistic positive solution. *Found. Comput. Math.*, 8(1):1–43, 2008.

[2] C. Beltrán and L.M. Pardo. Smale's 17th problem: average polynomial time to compute affine and projective solutions. *J. Amer. Math. Soc.*, 22(2):363–385, 2009.

[3] C. Beltrán and L. M. Pardo. Fast linear homotopy to find approximate zeros of polynomial systems. Manuscript 2008.

[4] C. Beltrán and M. Shub. Complexity of Bézout's theorem VII: Distance estimates in the condition metric. *Found. Comput. Math.*, 9(2):179–195, 2009.

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

[6] P. Bürgisser, and F. Cucker. On a problem posed by Steve Smale. `arxiv.org/abs/0909.2114`

[7] P. Bürgisser, F. Cucker, and M. Lotz. Smoothed analysis of complex conic condition numbers. *J. Math. Pures et Appl.*, 86:293–309, 2006.

[8] K. P. Choi. On the medians of gamma distributions and an equation of Ramanujan. *Proc. Amer. Math. Soc.*, 121(1):245–251, 1994.

[9] R. Howard. The kinematic formula in Riemannian homogeneous spaces. *Mem. Amer. Math. Soc.*, 106(509):vi+69, 1993.

[10] J. Renegar. On the worst-case arithmetic complexity of approximating zeros of systems of polynomials. *SIAM J. Comput.*, 18:350–370, 1989.

[11] A. Sankar, D. A. Spielman, and S.-H. Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476 (electronic), 2006.

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

[13] M. Shub. Complexity of Bézout's theorem VI: Geodesics in the condition (number) metric. *Found. Comput. Math.*, 9(2):171–178, 2009.

[14] M. Shub and S. Smale. Complexity of Bézout's theorem. I. Geometric aspects. *J. Amer. Math. Soc.*, 6(2):459–501, 1993.

[15] M. Shub and S. Smale. Complexity of Bézout's theorem II: volumes and probabilities. In F. Eyssette and A. Galligo, editors, *Computational Algebraic Geometry*, volume 109 of *Progress in Mathematics*, pages 267–285. Birkhäuser, 1993.

[16] M. Shub and S. Smale. Complexity of Bézout's theorem III: condition number and packing. *Journal of Complexity*, 9:4–14, 1993.

[17] M. Shub and S. Smale. Complexity of Bézout's theorem IV: probability of success; extensions. *SIAM J. of Numer. Anal.*, 33:128–148, 1996.

[18] M. Shub and S. Smale. Complexity of Bézout's theorem V: polynomial time. *Theoretical Computer Science*, 133:141–164, 1994.

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

[20] S. Smale. Mathematical problems for the next century. In *Mathematics: frontiers and perspectives*, pages 271–294. Amer. Math. Soc., Providence, RI, 2000.

[21] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians*, volume I, pages 597–606, 2002.