Condition length and complexity for the solution of polynomial systems

Diego Armentano*
Universidad de La República
URUGUAY
diego@cmat.edu.uy

Carlos Beltrán†
Universidad de Cantabria
SPAIN
beltranc@unican.es

Peter Bürgisser‡
Technische Universität Berlin
GERMANY
pbuerg@math.tu-berlin.de

Felipe Cucker§
City University of Hong Kong
HONG KONG
macucker@cityu.edu.hk

Michael Shub
City University of New York
U.S.A.
shub.michael@gmail.com

Abstract

Smale's 17th problem asks for an algorithm which finds an approximate zero of polynomial systems in average polynomial time (see [21]). The main progress on Smale's problem is [6] and [10]. In this paper we will improve on both approaches and prove an interesting intermediate result on the

*Partially supported by Agencia Nacional de Investigación e Innovación (ANII), Uruguay, and by CSIC group 618

†Partially suported by the research projects MTM2010-16051 and MTM2014-57590-P from Spanish Ministry of Science MICINN

‡Partially funded by DFG research grant BU 1371/2-2

§Partially funded by a GRF grant from the Research Grants Council of the Hong Kong SAR (project number CityU 100813).

AMS classification: Primary 65H10, 65H20. Secondary 58C35

Key words and phrases: Polynomial systems, homotopy methods, complexity estimates

Communicated by Teresa Krick

average value of the condition number. Our main results are Theorem 1 on the complexity of a randomized algorithm which improves the result of [6], Theorem 2 on the average of the condition number of polynomial systems which improves the estimate found in [10], and Theorem 3 on the complexity of finding a single zero of polynomial systems. This last theorem is similar to the main result of [10] but relies only on homotopy methods, thus removing the need for the elimination theory methods used in [10]. We build on methods developed in [2].

1 Introduction

Homotopy or continuation methods to solve a problem which might depend on parameters start with a problem instance and known solution and try to continue the solution along a path in parameter space ending at the problem we wish to solve. We recall how this works for the solutions of polynomial systems using a variant of Newton's method to accomplish the continuation.

Let $\mathcal{H}_d$ be the complex vector space of degree d complex homogeneous polynomials in $n+1$ variables. For $\alpha = (\alpha_0, \ldots, \alpha_n) \in \mathbb{N}^{n+1}$ satisfying $\sum_{j=0}^n \alpha_j = d$, and the monomial $z^\alpha = z_0^{\alpha_0} \cdots z_n^{\alpha_n}$, the Weyl Hermitian structure on $\mathcal{H}_d$ makes $\langle z^\alpha, z^\beta \rangle := 0$, for $\alpha \neq \beta$ and

$$\langle z^\alpha, z^\alpha \rangle := \binom{d}{\alpha}^{-1} = \left(\frac{d!}{\alpha_0! \cdots \alpha_n!} \right)^{-1}.$$

Now for $(d) = (d_1, \ldots, d_n)$ we let $\mathcal{H}_{(d)} = \prod_{k=1}^n \mathcal{H}_{d_k}$. This is a complex vector space of dimension

$$N := \sum_{i=1}^n \binom{n+d_i}{n}.$$

That is, N is the *size* of a system $f \in \mathcal{H}_{(d)}$, understood as the number of coefficients needed to describe f.

We endow $\mathcal{H}_{(d)}$ with the product Hermitian structure

$$\langle f, g \rangle := \sum_{k=1}^n \langle f_i, g_i \rangle,$$

where $f = (f_1, \ldots, f_n)$, and $g = (g_1, \ldots, g_n)$. This Hermitian structure is sometimes called the Weyl, Bombieri-Weyl, or Kostlan Hermitian structure. It is invariant under unitary substitution $f \mapsto f \circ U^{-1}$, where U is a unitary transformation of $\mathbb{C}^{n+1}$ (see [9, p. 118] for example).

On $\mathbb{C}^{n+1}$ we consider the usual Hermitian structure

$$\langle x, y \rangle := \sum_{k=0}^n x_k \, \overline{y_k}.$$

Given $0 \neq \zeta \in \mathbb{C}^{n+1}$, let $\zeta^\perp$ denote the Hermitian complement of ζ,

$$\zeta^\perp := \{v \in \mathbb{C}^{n+1} : \langle v, \zeta \rangle = 0\}.$$

For any nonzero $\zeta \in \mathbb{C}^{n+1}$, the subspace $\zeta^\perp$ is a model for the tangent space, $T_\zeta \mathbb{P}(\mathbb{C}^{n+1})$, of the projective space $\mathbb{P}(\mathbb{C}^{n+1})$ at the equivalence class of ζ (which we also denote by ζ). The space $T_\zeta \mathbb{P}(\mathbb{C}^{n+1})$ inherits an Hermitian structure from $\langle \cdot, \cdot \rangle$ given by

$$\langle v, w \rangle_\zeta := \frac{\langle v, w \rangle}{\langle \zeta, \zeta \rangle}.$$

See for example [9, Sec. 12.2] for more details on this standard metric structure of $\mathbb{P}(\mathbb{C}^{n+1})$.

The group of unitary transformations $\mathbb{U}$ acts naturally on $\mathbb{C}^{n+1}$ by $\zeta \mapsto U\zeta$ for $U \in \mathbb{U}$, and the Hermitian structure of $\mathbb{C}^{n+1}$ is invariant under this action.

A *zero* of the system of equations f is a point $x \in \mathbb{C}^{n+1}$ such that $f_i(x) = 0$, $i = 1, \ldots, n$. If we think of f as a mapping $f : \mathbb{C}^{n+1} \to \mathbb{C}^n$, it is a point x such that $f(x) = 0$.

For a *generic* system (that is, for a Zariski open set of $f \in \mathcal{H}_{(d)}$), Bézout's theorem states that the set of zeros consists of $\mathcal{D} := \prod_{k=1}^n d_k$ complex lines through 0. These $\mathcal{D}$ lines are $\mathcal{D}$ points in projective space $\mathbb{P}(\mathbb{C}^{n+1})$. So our goal is to approximate one of these points, and we will use *homotopy* or *continuation methods*.

These methods for the solution of a system $f \in \mathcal{H}_{(d)}$ proceed as follows. Choose $g \in \mathcal{H}_{(d)}$ and a zero $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$ of g (we denote by the same symbol an affine point and its projective class). Connect g to f by a path f_t, $0 \leq t \leq 1$, in $\mathcal{H}_{(d)}$ such that $f_0 = g$, $f_1 = f$, and try to continue $\zeta_0 = \zeta$ to ζ_t such that $f_t(\zeta_t) = 0$, so that $f_1(\zeta_1) = 0$ (see [7] for details or [12] for a complete discussion).

So homotopy methods numerically approximate the path (f_t, ζ_t). One way to accomplish the approximation is via (projective) Newton's method. Given an approximation x_t to ζ_t, define

$$x_{t+\Delta t} := N_{f_{t+\Delta t}}(x_t),$$

where for $h \in \mathcal{H}_{(d)}$ and $y \in \mathbb{P}(\mathbb{C}^{n+1})$ we define the *projective Newton's method* $N_h(y)$ following [17]:

$$N_h(y) := y - (Dh(y)|_{y^\perp})^{-1} h(y).$$

Note that N_h is defined on $\mathbb{P}(\mathbb{C}^{n+1})$ at those points where $Dh(y)|_{y^\perp}$ is invertible.

That x_t is an *approximate zero of f_t with associated (exact) zero ζ_t* means that the sequence of Newton iterations $N_{f_t}^k(x_t)$ converges immediately and quadratically to ζ_t.

Let us assume that $\{f_t\}_{t\in[0,1]}$ is a path in the sphere $\mathbb{S}(\mathcal{H}_{(d)}) := \{h \in \mathcal{H}_{(d)} : \|h\| = 1\}$. The main result of [16][1] is that the Δt_k may be chosen so that $t_0 = 0$, $t_k = t_{k-1} + \Delta t_k$ for $k = 1, \ldots, K$ with $t_K = 1$, such that for all k, x_{t_k} is an approximate zero of f_{t_k} with associated zero ζ_{t_k}, and the number K of steps can be bounded as follows:

$$K = K(f, g, \zeta) \leq C\, D^{3/2} \int_0^1 \mu(f_t, \zeta_t)\, \|(\dot{f}_t, \dot{\zeta}_t)\|\, dt. \tag{1.1}$$

Here C is a universal constant, $D = \max_i d_i$,

$$\mu(f, \zeta) := \begin{cases} \|f\| \left\|(Df(\zeta)|_{\zeta^\perp})^{-1} \mathrm{diag}(\|\zeta\|^{d_i - 1}\sqrt{d_i})\right\| & \text{if } Df(\zeta)|_{\zeta^\perp} \text{ is invertible} \\ \infty & \text{otherwise} \end{cases}$$

is the condition number of $f \in \mathcal{H}_{(d)}$ at $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$, $\mathrm{diag}(v)$ is the diagonal matrix whose diagonal entries are the coordinates of the vector v), and

$$\|(\dot{f}_t, \dot{\zeta}_t)\| = (\|\dot{f}_t\|^2 + \|\dot{\zeta}_t\|^2_{\zeta_t})^{1/2}$$

is the norm of the tangent vector to the curve in (f_t, ζ_t). The result in [16] is not fully constructive, but specific constructions have been given, see [3] and [14], and even programmed [4]. These constructions are similar to those given in [20] and [2] (this last, for the eigenvalue-eigenvector problem case).

The constructive versions cited above have slightly different criteria to choose the step length, which is the backbone of the continuation algorithm. However, all these algorithms satisfy a unitary invariance in the sense that if U is a unitary matrix of size $n + 1$ then

$$K(f, g, \zeta) = K(f \circ U^*, g \circ U^*, U\zeta). \tag{1.2}$$

The right-hand side in expression (1.1) is known as the *condition length* of the path (f_t, ζ_t). We will call (1.1) the *condition length estimate* of the number of steps.

Taking derivatives w.r.t. t in the equality $f_t(\zeta_t) = 0$ it is easily seen that

$$\dot{\zeta}_t = -(Df_t(\zeta_t)|_{\zeta_t^\perp})^{-1} \dot{f}_t(\zeta_t), \tag{1.3}$$

and with some work (see [9, Lemma 12, p. 231] one can prove that

$$\|\dot{\zeta}\|_{\zeta_t} \leq \mu(f_t, \zeta_t)\|\dot{f}_t\|.$$

It is known that $\mu(f, \zeta) \geq \sqrt{n} \geq 1$, so the estimate (1.1) may be bounded from above by

$$K(f, g, \zeta) \leq C'\, D^{3/2} \int_0^1 \mu^2(f_t, \zeta_t)\, \|\dot{f}_t\|\, dt, \tag{1.4}$$

[1] In [16] the theorem is actually proven in the projective space instead of the sphere, which is sharper, but we only use the sphere version in this paper.

where $C' = \sqrt{2}C$ is another constant. Let us call this estimate the μ^2-*estimate*.

The condition length estimate is better than the μ^2-estimate, but algorithms achieving the smaller number of steps are more subtle and the proofs of correctness more difficult.

Indeed in [5] and [10] the authors rely on the μ^2-estimate. At the times of these papers the algorithms achieving the condition length bound where in development, and [10] includes a construction which achieves the μ^2-estimate.

Yet, in a random situation, one might expect the improvement to be similar to the improvement given by the average of $\|A(x)\|$, in all possible directions, compared with $\|A\|$ (here, $A\colon \mathbb{C}^n \to \mathbb{C}^n$ denotes a linear operator), which according to [1] should give an improvement by a factor of the square root of the domain dimension. We have accomplished this for the eigenvalue-eigenvector problem in [2]. Here we use an argument similar to that of [2] to improve the estimate for the randomized algorithm in [6].

The Beltrán-Pardo randomized algorithm works as follows (see [6], and also [10]): on input $f \in \mathcal{H}_{(d)}$,

1. Choose f_0 at random and then a zero ζ_0 of f_0 at random. [6] describes a general scheme to do so (roughly speaking, one first draws the "linear" part of f_0, computes ζ_0 from it, and then draws the "nonlinear" part of f_0). An efficient implementation of this scheme, having running time $\mathcal{O}(nDN)$, is fully described and analyzed in [12, Section 17.6].

2. Then connect $f_0/\|f_0\|$ to $f/\|f\|$ by an arc of a great circle in the sphere, and invoke the continuation strategy above.

The main result of [6] is that the average number of steps of this procedure is bounded by $O(D^{3/2}nN)$, and its total average complexity is then $\mathcal{O}(D^{3/2}nN^2)$ (since the cost of an iteration of Newton's method, assuming all $d_i \geq 2$, is $\mathcal{O}(N)$, see [12, Proposition 16.32] and [11, Remark 7.8(1)]).

Our first main result is the following improvement of this last bound.

Theorem 1 (Randomized algorithm) *The average number of steps of the randomized algorithm with the condition length estimate is bounded by*

$$CD^{3/2}nN^{1/2},$$

where C is a universal constant.

The constant C can be taken as $\frac{\pi}{\sqrt{2}}C'$ with C' not more than 400 even accounting for input and round-off error, cf. [14].

The randomized algorithm has a nice property as proved in [6]: for every input system f with no singular zeros, the probability that the algorithm outputs an approximate zero associated to each of the $\mathcal{D}$ zeros of f is exactly $1/\mathcal{D}$. It can thus be used to generate a zero of f with the uniform distribution.

Remark 1 Theorem 1 is an improvement by a factor of $1/N^{1/2}$ of the bound in [6], which results from using the condition length estimate in place of the μ^2-estimate.

Before proceeding with the proof of Theorem 1, we introduce some useful notation. We define the *solution variety*

$$\mathcal{V} := \{(f, \zeta) \in \mathcal{H}_{(d)} \times \mathbb{P}(\mathbb{C}^{n+1}) \mid f(\zeta) = 0\},$$

and consider the projections

$$\begin{array}{ccccc} & & \mathcal{V} & & \\ & \swarrow^{\pi_1} & & \searrow^{\pi_2} & \\ \mathcal{H}_{(d)} & & & & \mathbb{P}(\mathbb{C}^{n+1}). \end{array} \tag{1.5}$$

The set of *ill-posed pairs* is the subset

$$\Sigma' := \{(f, \zeta) \in \mathcal{V} \mid Df(\zeta)|_{\zeta^\perp} \text{ is not invertible}\} = \{(f, \zeta) \in \mathcal{V} \mid \mu(f, \zeta) = \infty\}$$

and its projection $\Sigma := \pi_1(\Sigma')$ is the set of *ill-posed systems*. The number of iterations of the homotopy algorithm, $K(f, g, \zeta)$, is finite if and only if the lifting $\{(f_t, \zeta_t)\}_{t\in[0,1]}$ of the segment $\{f_t\}_{t\in[0,1]}$ does not cut Σ'.

1.1 Note added in proof

This manuscript was submitted to J. FoCM on July 14th, 2015. Just six days later, we received a note from Pierre Lairez who had found a way to derandomize the main result of [6], thus finding a deterministic answer to Smale's 17th problem [15]. The total complexity $O(n^2D^{3/2}N^2)$ of Lairez's algorithm is very similar to that of the original randomized version. According to our Theorem 1, the complexity bound of the randomized version can be lowered by a factor of $1/\sqrt{N}$. We think that the same improvement should apply to Lairez's deterministic algorithm.

We want to thank three anonymous referees for helpful comments.

2 Proof of Theorem 1

2.1 Preliminaries

Let us start this section with a few general facts we will use from Gaussian measures.

Given a finite dimensional real vector space V of dimension m, with an inner product, we define two natural objects.

- The unit sphere $\mathbb{S}(V)$ with the induced Riemannian structure and volume form: the volume of $\mathbb{S}(V)$ is $\frac{2\pi^{m/2}}{\Gamma(\frac{m}{2})}$.

- The Gaussian measure centered at $c \in V$, with variance $\frac{\sigma^2}{2} > 0$, whose density is

$$\frac{1}{\sigma^m \pi^{m/2}} e^{-\|x-c\|^2/\sigma^2}. \tag{2.6}$$

We will denote by $N_V(c, \sigma^2 \mathrm{Id})$ the density given in (2.6). We will skip the notation of the underlying space when it is understood. Furthermore, we will denote by $\mathbb{E}_{x \in V}$ the average in the case $\sigma = 1$ (that is, variance $1/2$).

The following lemma is well known, we however provide a proof because a similar argument is used later in the manuscript.

Lemma 2 *If $\varphi \colon V \to [0, +\infty]$ is measurable and homogeneous of degree $p > -m$, then*

$$\mathop{\mathbb{E}}_{x \in V}(\varphi(x)) = \frac{\Gamma(\frac{m+p}{2})}{\Gamma(\frac{m}{2})} \mathop{\mathbb{E}}_{u \in \mathbb{S}(V)}(\varphi(u)),$$

where

$$\mathop{\mathbb{E}}_{u \in \mathbb{S}(V)}(\varphi(u)) = \frac{1}{\mathrm{vol}(\mathbb{S}(V))} \int_{\mathbb{S}(V)} \varphi(u)\, du.$$

Proof. Integrating in polar coordinates we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \in V}(\varphi(x)) &= \frac{1}{\pi^{m/2}} \int_{x \in V} \varphi(x)\, e^{-\|x\|^2}\, dx \\
&= \frac{1}{\pi^{m/2}} \int_0^{+\infty} \rho^{m+p-1} e^{-\rho^2}\, d\rho \cdot \int_{u \in \mathbb{S}(V)} \varphi(u)\, du \\
&= \frac{\Gamma(\frac{m+p}{2})}{2\pi^{m/2}} \int_{u \in \mathbb{S}(V)} \varphi(u)\, du = \frac{\Gamma(\frac{m+p}{2})}{\Gamma(\frac{m}{2})} \mathop{\mathbb{E}}_{u \in \mathbb{S}(V)}(\varphi(u))
\end{aligned}$$

where we have used that $\int_0^{+\infty} \rho^k e^{-\rho^2}\, d\rho = \frac{1}{2}\Gamma(\frac{k+1}{2})$. □

The next results follows immediately from Fubini's theorem.

Lemma 3 *Let E be a linear subspace of V, and let $\Pi : V \to E$ be the orthogonal projection. Then, for any integrable function $\psi : E \to \mathbb{R}$ and for any $c \in V$, $\sigma > 0$, we have*

$$\mathop{\mathbb{E}}_{x \sim N_V(c, \sigma^2 \mathrm{Id})}(\psi(\Pi(x))) = \mathop{\mathbb{E}}_{y \sim N_E(\Pi(c), \sigma^2 \mathrm{Id})}(\psi(y)). \qquad \square$$

When V is a finite dimensional Hermitian vector space of complex dimension m, then the *complex Gaussian measure on V* with variance σ^2 is defined by the real Gaussian measure with variance $\sigma^2/2$ on the $2m$-dimensional real vector

space associated to V, whose inner product is the real part of the Hermitian product.

In this fashion, for any fixed $g \in \mathcal{H}_{(d)}$ and $\sigma > 0$, the Hermitian space $(\mathcal{H}_{(d)}, \langle \cdot, \cdot \rangle)$ is equipped with the complex Gaussian measure $N(g, \sigma^2 \mathrm{Id})$. The expected value of a function $\phi : \mathcal{H}_{(d)} \to \mathbb{R}$ with respect to this measure is given by

$$\mathop{\mathbb{E}}_{f \sim N(g, \sigma^2 \mathrm{Id})}(\phi) = \frac{1}{\sigma^{2N} \pi^N} \int_{f \in \mathcal{H}_{(d)}} \phi(f) e^{-\|f-g\|^2/\sigma^2} \, df. \tag{2.7}$$

Fix any $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$. Following [18, Sect. I-4], the space $\mathcal{H}_{(d)}$ is orthogonally decomposed into the sum $C_\zeta \oplus \mathcal{V}_\zeta$, where

$$\mathcal{V}_\zeta = \pi_2^{-1}(\zeta) = \{f \in \mathcal{H}_{(d)} : \, f(\zeta) = 0\}$$

is the fiber over ζ and

$$C_\zeta = \left\{ \mathrm{diag}\left(\frac{\langle \cdot, \zeta \rangle^{d_i}}{\langle \zeta, \zeta \rangle^{d_i}} \right) a : \, a \in \mathbb{C}^n \right\} \tag{2.8}$$

is the set of polynomial systems $f \in \mathcal{H}_{(d)}$ parametrized by $a \in \mathbb{C}^n$ such that for $z \in \mathbb{C}^{n+1}$, $f_i(z) = \langle z, \zeta \rangle^{d_i} / \|\zeta\|^{2d_i} \, a_i$, $1 \le i \le n$. Note that $\mathcal{V}_\zeta$ and C_ζ are linear subspaces of $\mathcal{H}_{(d)}$ of respective (complex) dimensions $N - n$ and n. Note also that

$$f_0 = f - \mathrm{diag}\left(\frac{\langle \cdot, \zeta \rangle^{d_i}}{\langle \zeta, \zeta \rangle^{d_i}} \right) f(\zeta)$$

is the orthogonal projection $\Pi_\zeta(f)$ of f onto the fiber $\mathcal{V}_\zeta$.

2.2 Average condition numbers

In this section we revisit the average value of the operator and Frobenius condition numbers on $\mathcal{H}_{(d)}$. The *Frobenius condition number* of f at ζ is given by

$$\mu_F(f, \zeta) := \|f\| \left\| \left(Df(\zeta)|_{\zeta^\perp} \right)^{-1} \mathrm{diag}(\|\zeta\|^{d_i - 1} d_i^{1/2}) \right\|_F, \tag{2.9}$$

that is, μ_F is defined as μ but using Frobenius instead of operator norm. Note that $\mu \le \mu_F \le \sqrt{n} \, \mu$. This version of the condition number was studied in depth in [8], where it was denoted $\widetilde{\mu}$ instead of μ_F.

Given $f \in \mathcal{H}_{(d)} \setminus \Sigma$, the average of the condition numbers over the fiber is

$$\mu^2_{\mathrm{av}}(f) := \frac{1}{\mathcal{D}} \sum_{\zeta : \, f(\zeta) = 0} \mu^2(f, \zeta), \qquad \mu^2_{F, \mathrm{av}}(f) := \frac{1}{\mathcal{D}} \sum_{\zeta : \, f(\zeta) = 0} \mu^2_F(f, \zeta)$$

(or ∞ if $f \in \Sigma$). For simplicity, in what follows we write $\mathbb{S} := \mathbb{S}(\mathcal{H}_{(d)})$.

Estimates on the probability distribution of the condition number μ are known since [19]. The exact expected value of $\mu^2_{\mathrm{av}}(f)$ when f is in the sphere $\mathbb{S}$ was found

in [6] and the following estimate for the expected value of $\mu^2_{\mathrm{av}}(f)$ when f is non-centered Gaussian was proved in [10]: for all $\widehat{f} \in \mathcal{H}_{(d)}$ and all $\sigma > 0$,

$$\mathop{\mathbb{E}}_{f\sim N(\widehat{f},\sigma^2\mathrm{Id})} \frac{\mu^2_{\mathrm{av}}(f)}{\|f\|^2} \;\leq\; \frac{e(n+1)}{2\sigma^2}. \tag{2.10}$$

The following result slightly improves (2.10), even though it is computed for μ_F.

Theorem 2 (Average condition number) *For every $\widehat{f} \in \mathcal{H}_{(d)}$ and $\sigma > 0$,*

$$\mathop{\mathbb{E}}_{f\sim N(\widehat{f},\sigma^2\mathrm{Id})} \frac{\mu^2_{F,\mathrm{av}}(f)}{\|f\|^2} \;\leq\; \frac{n}{\sigma^2},$$

and equality holds in the centered case.

Remark 4 The equality (in the centered case) of Theorem 2 implies from Lemma 2 with $p = -2$ that

$$\mathop{\mathbb{E}}_{f\in\mathbb{S}} \mu^2_{F,\mathrm{av}}(f) = (N-1)n.$$

Remark 5 In the proof of Theorem 2 we use the double fibration technique, a strategy based on the use of the classical coarea formula, see for example [9, p. 241]. In order to integrate some real-valued function over $\mathcal{H}_{(d)}$ whose value at some point f is an average over the fiber $\pi_1^{-1}(f)$, we lift it to $\mathcal{V}$ and then pushforward to $\mathbb{P}(\mathbb{C}^{n+1})$ using the projections given in (1.5). The original expected value in $\mathcal{H}_{(d)}$ is then written as an integral over $\mathbb{P}(\mathbb{C}^{n+1})$ which involves the quotient of normal Jacobians of the projections π_1 and π_2. More precisely,

$$\int_{f\in\mathcal{H}_{(d)}} \sum_{\zeta\colon f(\zeta)=0} \phi(f,\zeta)\,df = \int_{\zeta\in\mathbb{P}(\mathbb{C}^{n+1})} \int_{(f,\zeta)\in\pi_2^{-1}(\zeta)} \phi(f,\zeta) \frac{\mathrm{NJ}_{\pi_1}}{\mathrm{NJ}_{\pi_2}}(f,\zeta)\,d\pi_2^{-1}(\zeta)\,d\zeta, \tag{2.11}$$

where

$$\frac{\mathrm{NJ}_{\pi_1}}{\mathrm{NJ}_{\pi_2}}(f,\zeta) = |\det(Df(\zeta)|_{\zeta^\perp})|^2$$

(see [9, Section 13.2], [12, Section 17.3], or [2, Theorem 6.2] for further details and other examples of use).

We point out that the proof of Theorem 2 can also be achieved using the (slightly) different method of [6] and [12, Chapter 18] based on the mapping taking (f,ζ) to $(Df(\zeta),\zeta)$ whose normal Jacobian is known to be constant (see [6, Main Lemma]).

Proof of Theorem 2. By the definition of non-centered Gaussian, and the double-fibration formula (2.11), we have

$$\mathop{\mathbb{E}}_{f\sim N(\widehat{f},\sigma^2\mathrm{Id})}\frac{\mu^2_{F,\mathrm{av}}(f)}{\|f\|^2}=\frac{1}{\mathcal{D}}\int_{f\in\mathcal{H}_{(d)}}\left(\sum_{\zeta:\,f(\zeta)=0}\frac{\mu^2_F(f,\zeta)}{\|f\|^2}\right)\frac{e^{\frac{-\|f-\widehat{f}\|^2}{\sigma^2}}}{\sigma^{2N}\pi^N}\,df \tag{2.12}$$

$$=\frac{1}{\mathcal{D}}\frac{1}{(\sigma^2\pi)^n}\int_{\zeta\in\mathbb{P}(\mathbb{C}^{n+1})}e^{\frac{-\|\widehat{f}-\Pi_\zeta(\widehat{f})\|^2}{\sigma^2}}\int_{f\in\mathcal{V}_\zeta}\frac{\mu^2_F(f,\zeta)}{\|f\|^2}\left|\det\left(Df(\zeta)|_{\zeta^\perp}\right)\right|^2\frac{e^{\frac{-\|f-\Pi_\zeta(\widehat{f})\|^2}{\sigma^2}}}{(\sigma^2\pi)^{N-n}}\,df\,d\zeta,$$

where we have used that $\|f-\widehat{f}\|^2=\|f-\Pi_\zeta(\widehat{f})\|^2+\|\widehat{f}-\Pi_\zeta(\widehat{f})\|^2$ for every $f\in\mathcal{V}_\zeta$ (note that $\Pi_\zeta(\hat{f})=\hat{f}$ if $\hat{f}\in\mathcal{V}_\zeta$).

We simplify now the integral $I_\zeta(\hat{f})$ over the fiber $\mathcal{V}_\zeta$, that is

$$I_\zeta(\hat{f}):=\int_{f\in\mathcal{V}_\zeta}\frac{\mu^2_F(f,\zeta)}{\|f\|^2}\left|\det\left(Df(\zeta)|_{\zeta^\perp}\right)\right|^2\frac{e^{\frac{-\|f-\Pi_\zeta(\widehat{f})\|^2}{\sigma^2}}}{(\sigma^2\pi)^{N-n}}\,df.$$

Let U_ζ be a unitary transformation of $\mathbb{C}^{n+1}$ such that $U_\zeta(\zeta/\|\zeta\|)=e_0$. Then, by the invariance under unitary substitution of each term under the integral sign, we have by the change of variable formula with $h=f\circ U_\zeta^*$ that

$$\begin{aligned}I_\zeta(\hat{f})&=\int_{h\in\mathcal{V}_{e_0}}\frac{\mu^2_F(h\circ U_\zeta,\zeta)}{\|h\circ U_\zeta\|^2}\left|\det\left(D(h\circ U_\zeta)(\zeta)|_{\zeta^\perp}\right)\right|^2\frac{e^{\frac{-\|h\circ U_\zeta-\Pi_\zeta(\widehat{f})\|^2}{\sigma^2}}}{(\sigma^2\pi)^{N-n}}\,dh\\&=\int_{h\in\mathcal{V}_{e_0}}\frac{\mu^2_F(h,e_0)}{\|h\|^2}\left|\det\left(Dh(e_0)|_{e_0^\perp}\right)\right|^2\frac{e^{\frac{-\|h-\Pi_{e_0}(\widehat{h_\zeta})\|^2}{\sigma^2}}}{(\sigma^2\pi)^{N-n}}\,dh\\&=\mathop{\mathbb{E}}_{h\sim N(\Pi_{e_0}(\widehat{h_\zeta}),\sigma^2\mathrm{Id})}\left(\frac{\mu^2_F(h,e_0)}{\|h\|^2}\left|\det\left(Dh(e_0)|_{e_0^\perp}\right)\right|^2\right),\end{aligned}$$

where $\widehat{h}_\zeta:=\widehat{f}\circ U_\zeta^{-1}$. We project now $h\in\mathcal{V}_{e_0}$ orthogonally onto the vector space

$$L_{e_0}:=\{g\in\mathcal{H}_{(d)}:g(e_0)=0,D^kg(e_0)=0\text{ for }k\geq 2\},$$

obtaining $g\in L_{e_0}$. Since $Dh(e_0)|_{e_0^\perp}$ coincides with $Dg(e_0)|_{e_0^\perp}$ (see for example [12, Prop. 16.16]), which implies indeed that $\mu_F(h,e_0)/\|h\|^2=\mu_F(g,e_0)/\|g\|^2$, we conclude by Lemma 3 that

$$I_\zeta(\hat{f})=\mathop{\mathbb{E}}_{g\sim N(\Pi_{L_0}(\widehat{h}_\zeta),\sigma^2\mathrm{Id})}\left(\frac{\mu^2_F(g,e_0)}{\|g\|^2}\left|\det\left(Dg(e_0)|_{e_0^\perp}\right)\right|^2\right).$$

By the change of variables given by

$$L_{e_0}\to\mathbb{C}^{n\times n},\ g\mapsto A:=\mathrm{diag}(d_i^{-1/2})Dg(e_0)|_{e_0^\perp},$$

which is a linear isometry (see [9, Lemma 17, Ch. 12]), we have $\frac{\mu^2_{F,\mathrm{av}}(g)}{\|g\|^2} = \|A^{-1}\|_F^2$ and denoting by $\widehat{A}_\zeta$ the image of $\Pi_{L_0}(\widehat{h}_\zeta)$, we obtain that

$$I_\zeta(\hat{f}) = \mathop{\mathbb{E}}_{A\in N(\widehat{A}_\zeta,\sigma^2\mathrm{Id}_n)} \left(\|A^{-1}\|_F^2\,|\det(A)|^2\right).$$

We thus conclude from (2.12) that

$$\mathop{\mathbb{E}}_{f\sim N(\widehat{f},\sigma^2\mathrm{Id})}\left(\frac{\mu^2_{F,\mathrm{av}}(f)}{\|f\|^2}\right) = \tag{2.13}$$
$$\frac{1}{\mathcal{D}}\frac{1}{(\sigma^2\pi)^n}\int_{\zeta\in\mathbb{P}(\mathbb{C}^{n+1})} e^{\frac{-\|\widehat{f}-\Pi_\zeta(\widehat{f})\|^2}{\sigma^2}} \mathop{\mathbb{E}}_{A\in N(\widehat{A}_\zeta,\sigma^2\mathrm{Id}_n)} \left(\|A^{-1}\|_F^2\,|\det(A)|^2\right) d\zeta.$$

If we replace $\mu_{F,\mathrm{av}}(f)^2/\|f\|^2$ by the constant function 1 on $\mathcal{H}_{(d)}$, the same argument leading to (2.13) now leads to

$$1 \;=\; \frac{1}{\mathcal{D}}\frac{1}{(\sigma^2\pi)^n}\int_{\zeta\in\mathbb{P}(\mathbb{C}^{n+1})} e^{\frac{-\|\widehat{f}-\Pi_\zeta(\widehat{f})\|^2}{\sigma^2}} \mathop{\mathbb{E}}_{A\in N(\widehat{A}_\zeta,\sigma^2\mathrm{Id}_n)} \left(|\det(A)|^2\right) d\zeta. \tag{2.14}$$

From Proposition 7.1 of [2], we can bound

$$\mathop{\mathbb{E}}_{A\in N(\widehat{A}_\zeta,\sigma^2\mathrm{Id}_n)} \left(\|A^{-1}\|_F^2\,|\det(A)|^2\right) \;\le\; \frac{n}{\sigma^2} \mathop{\mathbb{E}}_{A\in N(\widehat{A}_\zeta,\sigma^2\mathrm{Id}_n)} \left(|\det(A)|^2\right), \tag{2.15}$$

with equality if $\widehat{A}_\zeta = 0$. By combining (2.13), (2.15), and (2.14) we obtain

$$\mathop{\mathbb{E}}_{f\sim N(\widehat{f},\sigma^2\mathrm{Id})}\left(\frac{\mu^2_{F,\mathrm{av}}(f)}{\|f\|^2}\right) \;\le\; \frac{n}{\sigma^2},$$

as claimed by the theorem. Moreover, equality holds if $\widehat{f} = 0$ (which implies $\widehat{A}_\zeta = 0$ for all ζ). □

2.3 Complexity of the randomized algorithm

The goal of this section is to prove Theorem 1. To do so, we begin with some preliminaries.

For $f \in \mathbb{S}$ we denote by $T_f\mathbb{S}$ the tangent space at f of $\mathbb{S}$. This space is equipped with the real part of the Hermitian structure of $\mathcal{H}_{(d)}$, and coincides with the (real) orthogonal complement of $f \in \mathcal{H}_{(d)}$.

We consider the map $\phi\colon \mathbb{S}\times\mathcal{H}_{(d)} \to [0,\infty]$ defined for $f\notin\Sigma$ by

$$\phi(f,\dot{f}) := \frac{1}{\mathcal{D}}\sum_{\zeta:\, f(\zeta)=0} \mu(f,\zeta)\big\|\big(\dot{f},\dot{\zeta}\big)\big\|, \tag{2.16}$$

where $\dot{\zeta} = -(Df(\zeta)|_{\zeta^\perp})^{-1}\dot{f}(\zeta)$, and by $\phi(f,\dot{f}) := \infty$ if $f \in \Sigma$. Note that ϕ satisfies $\phi(f,\lambda\dot{f}) = \lambda\phi(f,\dot{f})$ for $\lambda \geq 0$.

Suppose that $f_0, f \in \mathbb{S}$ are such that $f_0 \neq \pm f$ and denote by $\mathcal{L}_{f_0,f}$ the shorter great circle segment with endpoints f_0 and f. Moreover, let $\alpha = d_{\mathbb{S}}(f_0,f)$ denote the angle between f_0 and f. If $[0,1] \to \mathbb{S}$, $t \mapsto f_t$ is the constant speed parametrization of $\mathcal{L}_{f_0,f}$ with endpoints f_0 and $f_1 = f$, then $\|\dot{f}_t\| = \alpha$. We may also parametrize $\mathcal{L}_{f_0,f}$ by the arc-length $s = \alpha t$, setting $F_s := f_{\alpha^{-1}s}$, in which case $\dot{F}_s = \alpha^{-1}\dot{f}_t$ is the unit tangent vector (in the direction of the parametrization) to $\mathcal{L}_{f_0,f}$ at F_s. Moreover,

$$\int_0^1 \phi(f_t,\dot{f}_t)\,dt = \int_0^\alpha \phi(F_s,\dot{F}_s)\,ds.$$

Consider the compact submanifold $\mathcal{S}$ of $\mathbb{S}\times\mathbb{S}$ given by

$$\mathcal{S} = \{(f,\dot{f}) \in \mathbb{S}\times\mathbb{S} : \dot{f} \in T_f\mathbb{S}\},$$

which inherits a Riemannian structure from the product $\mathbb{S}\times\mathbb{S}$.

Lemma 6 *Let $V \equiv \mathbb{R}^m$ be a finite–dimensional Hilbert space. Let $\mathbb{S}(V)$ be the unit sphere and*

$$\mathcal{S}(V) = \{(x,y) \in \mathbb{S}(V)\times\mathbb{S}(V) : y \in T_x\mathbb{S}(V)\}$$

Then, the projection $\pi_V : \mathcal{S}(V) \to \mathbb{S}(V)$, $(x,y) \mapsto x$, has normal Jacobian $1/\sqrt{2}$.

Proof. Note that $\mathcal{S}(V) = \{(x,y) \in \mathbb{S}(V)\times\mathbb{S}(V) : y^Tx = 0\}$ and from the regular mapping theorem $\mathcal{S}(V)$ is a hypersurface of $\mathbb{S}(V)\times\mathbb{S}(V)$ with tangent space

$$T_{(x,y)}\mathcal{S}(V) = \{(\dot{x},\dot{y}) \in x^\perp \times y^\perp : \dot{y}^Tx + y^T\dot{x} = 0\}.$$

The kernel of the derivative is easy to compute: $Ker(D\pi_V(x,y)) = \{(0,\dot{y}) \in x^\perp\times y^\perp : \dot{y}^Tx = 0\}$. The orthogonal complement X of the kernel is then

$$X = (KerD\pi_V(x,y))^\perp = \{(\dot{x},\dot{y}) \in T_{(x,y)}\mathcal{S}(V) : \dot{y} = \lambda x\} = \{(\dot{x},-(y^T\dot{x})x) : \dot{x} \in x^\perp\}.$$

Let $y = \dot{x}_1, \dot{x}_2, \ldots, \dot{x}_{m-1}$ be an orthogonal basis of $x^\perp$. The linear mapping $D\pi_V(x,y)\,|_X$ in the associated orthogonal basis

$$\{(y,-x)/\sqrt{2}, (\dot{x}_2,0), \ldots, (\dot{x}_{m-1},0)\}$$

of X and $\{\dot{x}_1,\ldots,\dot{x}_{m-1}\}$ of $T_x\mathbb{S}$ is diagonal with entries $1/\sqrt{2}, 1, \ldots, 1$. The normal Jacobian is thus $1/\sqrt{2}$ as claimed.

□

The following lemma has been proven in [2].

Lemma 7 *Let*

$$I_\phi := \mathop{\mathbb{E}}_{f_0, f \in \mathbb{S}} \left(\int_0^1 \phi(f_t, \dot{f}_t)\, dt \right).$$

Then, we have

$$I_\phi = \frac{\pi}{2} \mathop{\mathbb{E}}_{(f,\dot{f}) \in \mathcal{S}} \big(\phi(f, \dot{f})\big),$$

where the expectation on the right hand-side refers to the uniform distribution on $\mathcal{S}$.

We proceed with a further auxiliary result. For $f \in \mathbb{S}$ we consider the unit sphere $\mathcal{S}_f := \{\dot{f} \in T_f\mathbb{S} : (f, \dot{f}) \in \mathcal{S}\}$ in $T_f\mathbb{S}$.

Lemma 8 *Fix $f \in \mathbb{S}$ and $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$ with $f(\zeta) = 0$. For $\dot{f} \in \mathcal{S}_f$ let $\dot{\zeta} = \dot{\zeta}(\dot{f})$ be the function of $(f, \dot{f})$ and ζ given by $\dot{\zeta} = (-Df(\zeta)|_{\zeta^\perp})^{-1}\dot{f}(\zeta)$, that is, ζ is as in* (2.16)*. Then we have*

$$\mathop{\mathbb{E}}_{\dot{f} \in \mathcal{S}_f} (\|\dot{\zeta}\|^2) = \frac{1}{N - \frac{1}{2}} \left\| (Df(\zeta)|_{\zeta^\perp})^{-1} \right\|_F^2,$$

where the expectation is with respect to the uniform probability distribution on $\mathcal{S}_f$.

Proof. Since the map $T_f\mathbb{S} \to \mathbb{R}$, $\dot{f} \mapsto \|\dot{\zeta}(\dot{f})\|^2$ is quadratic, we get from Lemma 2 (recall that $\dim T_f\mathbb{S} = 2N - 1$)

$$\mathop{\mathbb{E}}_{\dot{f} \in T_f\mathbb{S}} (\|\dot{\zeta}(\dot{f})\|^2) = \left(N - \frac{1}{2}\right) \mathop{\mathbb{E}}_{\dot{f} \in \mathcal{S}_f} \left(\left\|\dot{\zeta}(\dot{f})\right\|^2\right).$$

Recall the definition of C_ζ given in (2.8). Note that the mapping $\mathcal{H}_{(d)} \to C_\zeta$ given by $\dot{f} \mapsto \Pi_{C_\zeta}\dot{f}$ is an orthogonal projection, and furthermore $C_\zeta \to \mathbb{C}^n$ given by $\dot{f} \mapsto \dot{f}(\zeta)$ is a linear isometry. Then from Lemma 3, and the change of variables formula we obtain

$$\mathop{\mathbb{E}}_{\dot{f} \in T_f\mathbb{S}} \left(\left\|\dot{\zeta}(\dot{f})\right\|^2\right) = \mathop{\mathbb{E}}_{\dot{w} \in \mathbb{C}^n} \left\|(Df(\zeta)|_{\zeta^\perp})^{-1}\dot{w}\right\|^2 = \left\|(Df(\zeta)|_{\zeta^\perp})^{-1}\right\|_F^2,$$

where the last equality is straightforward looking at the singular value decomposition of $(Df(\zeta)|_{\zeta^\perp})^{-1}$. □

Proof of Theorem 1. The average number of homotopy steps of the randomized algorithm is given by the following integral:

$$\mathop{\mathbb{E}}_{f, f_0 \in \mathbb{S}} \Big(\frac{1}{\mathcal{D}} \sum_{\zeta_0 \colon f_0(\zeta_0) = 0} K(f, f_0, \zeta_0) \Big).$$

From (1.1), using the notation there, we know that the number of Newton steps of the homotopy with starting pair (f_0, ζ_0) and target system f is bounded as

$$K(f, f_0, \zeta_0) \;\le\; CD^{3/2} \int_0^1 \mu(f_t, \zeta_t)\, \|(\dot f_t, \dot\zeta_t)\|\, dt.$$

Hence we get for $f, f_0 \in \mathbb{S}$,

$$\begin{aligned}\frac{1}{\mathcal{D}} \sum_{\zeta_0 \colon f_0(\zeta_0)=0} K(f, f_0, \zeta_0) \;&\le CD^{3/2} \int_0^1 \frac{1}{\mathcal{D}} \sum_{\zeta_0 \colon f_0(\zeta_0)=0} \mu(f_t, \zeta_t)\, \|(\dot f_t, \dot\zeta_t)\|\, dt \\ &= CD^{3/2} \int_0^1 \phi(f_t, \dot f_t)\, dt.\end{aligned}$$

Therefore, by Lemma 7,

$$\mathop{\mathbb{E}}_{f, f_0 \in \mathbb{S}} \Big(\frac{1}{\mathcal{D}} \sum_{\zeta_0 \colon f_0(\zeta_0)=0} K(f, f_0, \zeta_0) \Big) \;\le\; C\, D^{3/2}\, \frac{\pi}{2} \mathop{\mathbb{E}}_{(f, \dot f) \in \mathcal{S}} \big(\phi(f, \dot f) \big). \tag{2.17}$$

From the coarea formula and Lemma 6, we obtain

$$\begin{aligned}\mathop{\mathbb{E}}_{(f, \dot f) \in \mathcal{S}} \big(\phi(f, \dot f) \big) &= \sqrt{2} \mathop{\mathbb{E}}_{f \in \mathbb{S}} \mathop{\mathbb{E}}_{\dot f \in \mathcal{S}_f} \big(\phi(f, \dot f) \big) \\ &= \sqrt{2} \mathop{\mathbb{E}}_{f \in \mathbb{S}} \Big(\frac{1}{\mathcal{D}} \sum_{\zeta \colon f(\zeta)=0} \mu(f, \zeta) \mathop{\mathbb{E}}_{\dot f \in \mathcal{S}_f} \big(\|(\dot f, \dot\zeta)\| \big) \Big).\end{aligned}$$

In order to estimate this last quantity, note first that from the Cauchy–Schwarz inequality, for $f \in \mathbb{S}$,

$$\begin{aligned}\mathop{\mathbb{E}}_{\dot f \in \mathcal{S}_f} \Big(\|(\dot f, \dot\zeta)\| \Big) = \mathop{\mathbb{E}}_{\dot f \in \mathcal{S}_f} \big((1 + \|\dot\zeta\|^2)^{\frac12} \big) \;&\le\; \Big(1 + \mathop{\mathbb{E}}_{\dot f \in \mathcal{S}_f} (\|\dot\zeta\|^2) \Big)^{1/2} \\ &\le\; \left(1 + \frac{1}{N - \frac12} \left\| (Df(\zeta)|_{\zeta^\perp})^{-1} \right\|_F^2 \right)^{1/2},\end{aligned}$$

the last by Lemma 8. Now we use $\|(Df(\zeta)|_{\zeta^\perp})^{-1}\|_F \le \mu_F(f, \zeta)$ and $\mu(f, \zeta) \le \mu_F(f, \zeta)$ to deduce

$$\begin{aligned}\frac{1}{\sqrt{2}} \mathop{\mathbb{E}}_{(f, \dot f) \in \mathcal{S}} \big(\phi(f, \dot f) \big) &\le \mathop{\mathbb{E}}_{f \in \mathbb{S}} \left(\frac{1}{\mathcal{D}} \sum_{\zeta \colon f(\zeta)=0} \mu_F(f, \zeta) \Big(1 + \frac{\mu_F^2(f, \zeta)}{N - \frac12} \Big)^{\frac12} \right) \\ &\le \mathop{\mathbb{E}}_{f \in \mathbb{S}} \left(\frac{1}{\mathcal{D}} \sum_{\zeta \colon f(\zeta)=0} \left(\frac{(N - \frac12)^{\frac12}}{2} + \frac{\mu_F^2(f, \zeta)}{(N - \frac12)^{\frac12}} \right) \right) \\ &= \frac{(N - \frac12)^{\frac12}}{2} + \mathop{\mathbb{E}}_{f \in \mathbb{S}} \left(\frac{\mu_{F,\mathrm{av}}^2(f)}{(N - \frac12)^{\frac12}} \right)\end{aligned}$$

the second inequality since for all $x \geq 0$ and $a > 0$ we have

$$x^{1/2}(1 + a^2 x)^{1/2} \leq \frac{1}{2a} + ax.$$

A call to Remark 4 finally yields

$$\frac{1}{\sqrt{2}} \mathop{\mathbb{E}}_{(f,\dot{f})\in\mathcal{S}} \left(\phi(f, \dot{f})\right) \;\leq\; \frac{(N - \frac{1}{2})^{\frac{1}{2}}}{2} + \frac{(N-1)n}{(N-\frac{1}{2})^{\frac{1}{2}}} \;\leq\; \sqrt{N}\left(\frac{1}{2} + n\right).$$

Replacing this bound in (2.17) finishes the proof. □

3 A Deterministic Algorithm

A deterministic solution for Smale's 17th problem is yet to be found (added in proof: see Section 1.1). The state of the art for this theme is given in [10] where the following result is proven.

Theorem 3 *There is a deterministic real-number algorithm that on input $f \in \mathcal{H}_{(d)}$ computes an approximate zero of f in average time $N^{\mathcal{O}(\log\log N)}$. 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.

The algorithm exhibited in [10] uses two algorithmic strategies according to whether $D \leq n$ or $D > n$. In the first case, it applies a homotopy method and in the second an adaptation of a method coming from symbolic computation.

The goal of this section is to show that a more unified approach, where homotopy methods are used in both cases, yields a proof of Theorem 3 as well. Besides a gain in expositional simplicity, this approach can claim for it the well-established numerical stability of homotopy methods.

In all what follows we assume the simpler homotopy algorithm in [10] (as opposed to those in [3, 14]). Its choice of step length at the kth iteration is proportional to $\mu^{-2}(f_{t_k}, x_{t_k})$ (which, in turn, is proportional to $\mu^{-2}(f_{t_k}, \zeta_{t_k})$). For this algorithm, we have the μ^2-estimate (1.4) but not the finer estimate (1.1).

To understand the technical requirements of the analysis of a deterministic algorithm, let us summarize an analysis (simpler than the one in the preceding section because of the assumption above) for the randomized algorithm. Recall, the latter draws an initial pair (g, ζ) from a distribution which amounts to first draw g from the distribution on $\mathbb{S}$ and then draw ζ uniformly among the $\mathcal{D}$ zeros $\{\zeta^{(1)}, \ldots, \zeta^{(\mathcal{D})}\}$ of g. The μ^2-estimate (1.4) provides an upper bound for

the number of steps needed to continue ζ to a zero of f following the great circle from g to f (assuming $\|f\| = \|g\| = 1$ and $f \neq \pm g$). Now (1.4) does not change if we reparametrize $\{f_t\}_{t\in[0,1]}$ by arc–length, so we can also write it as

$$K(f,g,\zeta) \leq C'\, D^{3/2} \int_0^{d_{\mathbb{S}}(g,f)} \mu^2(f_s,\zeta_s)\, ds,$$

where $d_{\mathbb{S}}(g,f)$ is the spherical distance from g to f. Thus, the average number of homotopy iterations satisfies

$$\begin{aligned}\mathop{\mathbb{E}}_{f\in\mathbb{S}}\mathop{\mathbb{E}}_{g\in\mathbb{S}}\frac{1}{\mathcal{D}}\sum_{i=1}^{\mathcal{D}} K(f,g,\zeta^{(i)}) &\leq\ C'\, D^{3/2}\ \mathop{\mathbb{E}}_{f\in\mathbb{S}}\mathop{\mathbb{E}}_{g\in\mathbb{S}}\frac{1}{\mathcal{D}}\sum_{i=1}^{\mathcal{D}}\int_0^{d_{\mathbb{S}}(g,f)} \mu^2(f_s,\zeta_s^{(i)})\, ds \\ &\leq\ C'\, D^{3/2}\ \mathop{\mathbb{E}}_{f\in\mathbb{S}}\mathop{\mathbb{E}}_{g\in\mathbb{S}}\int_0^{d_{\mathbb{S}}(g,f)} \mu^2_{F,\mathrm{av}}(f_s)\, ds. \qquad (3.18)\end{aligned}$$

Let P_s denote the set of pairs $(f,g) \in \mathbb{S}^2$ such that $d_{\mathbb{S}}(g,f) \geq s$. Rewriting the above integral using Fubini, we get

$$\mathop{\mathbb{E}}_{f\in\mathbb{S}}\mathop{\mathbb{E}}_{g\in\mathbb{S}}\int_0^{d_{\mathbb{S}}(g,f)} \mu^2_{F,\mathrm{av}}(f_s)\, ds = \int_0^{\pi}\int_{P_s} \mu^2_{F,\mathrm{av}}(f_s)\, df\, dg\, ds = \frac{\pi}{2}\mathop{\mathbb{E}}_{h\in\mathbb{S}}\mu^2_{F,\mathrm{av}}(h),$$

the second equality holding since for a fixed $s \in [0,\pi]$ and uniformly distributed $(f,g) \in P_s$, one can show that the system f_s is uniformly distributed on $\mathbb{S}$. Summarizing, we get

$$\mathop{\mathbb{E}}_{f\in\mathbb{S}}\mathop{\mathbb{E}}_{g\in\mathbb{S}}\frac{1}{\mathcal{D}}\sum_{i=1}^{\mathcal{D}} K(f,g,\zeta^{(i)})\ \leq\ C'\, D^{3/2}\,\frac{\pi}{2}\mathop{\mathbb{E}}_{h\in\mathbb{S}}\mu^2_{F,\mathrm{av}}(h) \underset{\text{Rmk. 4}}{=} C'\, D^{3/2}\,\frac{\pi}{2}(N-1)n.$$

This constitutes an elegant derivation of the previous $\mathcal{O}(nD^{3/2}N)$ bound (but not of the sharper bound of our Theorem 1).

Proof of Theorem 3. If the initial pair (g,ζ) is not going to be random we face two difficulties. First —as g is not random— the intermediate systems f_t are not going to be uniformly distributed on $\mathbb{S}$. Second —as ζ is not random— we will need a bound on a given $\mu^2(f_t,\zeta_t)$ rather than one on the mean of these quantities (over the $\mathcal{D}$ possible zeros of f_t), as provided by Theorem 2.

Consider a fixed initial pair (g,ζ) with $g \in \mathbb{S}$ and let s_1 be the step length of the first step of the algorithm (see for example the definition of Algorithm ALH in [10]), which satisfies

$$s_1 \geq \frac{c}{D^{3/2}\mu^2(g,\zeta)} \quad (c \text{ a constant}). \qquad (3.19)$$

Note that this bound on the length s_1 of the first homotopy step depends on the condition $\mu(g,\zeta)$ only and is thus independent of the condition at the other zeros of g.

Consider also the short portion of great circle contained in $\mathbb{S}$ with endpoints g and $f/\|f\|$, which we parametrize by arc length and call h_s (that is, $h_0 = g$ and $h_\alpha = f/\|f\|$ where $\alpha = d_{\mathbb{S}}(g, f/\|f\|)$), defined for $s \in [0, \alpha]$. Thus, after the first step of the homotopy, the current pair is (h_{s_1}, x_1) and we denote by ζ' the zero of h_{s_1} associated to x_1. We will focus on bounding the quantity

$$H := H(g, \zeta) := \mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} \frac{1}{\mathcal{D}} \sum_{i=1}^{\mathcal{D}} K\big(f/\|f\|, h_{s_1}, \zeta^{(i)}\big),$$

where the sum is over all the zeros $\zeta^{(i)}$ of h_{s_1}. This is the average of the number of homotopy steps over *both* the system f and the $\mathcal{D}$ zeros of h_{s_1}. We will be interested in this average even though we will not consider algorithms following a path randomly chosen: the homotopy starts at the pair (g, ζ), moves to (h_{s_1}, x_1) and proceeds following this path.

From (1.4) applied to $(h_{s_1}, \zeta^{(i)})$,

$$K\big(f/\|f\|, h_{s_1}, \zeta^{(i)}\big) \;\le\; C'\, D^{3/2} \int_{s_1}^{\alpha} \mu^2(h_s, \zeta_s^{(i)})\, \|\dot{h}_s\|\, ds, \tag{3.20}$$

Reparametrizing $\{h_s : s_1 \le s \le \alpha\}$ by $\{f_t/\|f_t\| : t_1 \le t \le 1\}$ where $f_t = (1-t)g + tf$ and t_1 is such that $f_{t_1}/\|f_{t_1}\| = h_{s_1}$ does not change the value of the path integral in (3.20).

Lemma 9 *With the notations above we have*

$$t_1 = \frac{1}{\|f\| \sin\alpha \cot(s_1\alpha) - \|f\| \cos\alpha + 1} \ge \frac{c'}{D^{3/2}\sqrt{N}\mu^2(g,\zeta)},$$

c' a constant.

Proof. The formula for t_1 is shown in [10, Prop. 5.2]. For the bound, we have

$$\begin{aligned}
\|f\| \sin\alpha \cot(s_1\alpha) - \|f\| \cos\alpha + 1 &\le \|f\| \sin\alpha\, (s_1\alpha)^{-1} + \|f\| + 1 \\
&\le \sqrt{2N}\frac{1}{s_1} + \sqrt{2N} + 1 \\
&\le \sqrt{2N}\left(\frac{D^{3/2}\mu^2(g,\zeta)}{c} + 1 + \frac{1}{\sqrt{2N}}\right) \\
&\le \frac{\sqrt{N} D^{3/2}\mu^2(g,\zeta)}{c'}
\end{aligned}$$

for an appropriately chosen c'. □

We continue with the proof of Theorem 3. A simple computation shows that

$$\|\dot{h}_t\| = \left\| \frac{d}{dt}\left(\frac{f_t}{\|f_t\|}\right)\right\| \;\le\; \frac{\|f\|\|g\|}{\|f_t\|^2} = \frac{\|f\|}{\|f_t\|^2},$$

so we have

$$K\big(f/\|f\|, h_{s_1}, \zeta^{(i)}\big) \;\le\; C'\, D^{3/2}\, \|f\| \int_{t_1}^{1} \frac{\mu^2(f_t, \zeta_t^{(i)})}{\|f_t\|^2}\, dt. \tag{3.21}$$

Because of scale invariance, the quantity H satisfies

$$H = \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}^{\sqrt{2N}}} \frac{1}{\mathcal{D}} \sum_{i=1}^{\mathcal{D}} K(f, h_{s_1}, \zeta^{(i)}),$$

where the second expectation is taken over a truncated Gaussian (that only draws systems f with $\|f\| \le \sqrt{2N}$) with density function given by

$$\rho(f) := \begin{cases} \frac{1}{P}\varphi(f) & \text{if } \|f\| \le \sqrt{2N} \\ 0 & \text{otherwise.} \end{cases}$$

Here φ is the density function of the standard Gaussian on $\mathcal{H}_{(d)}$ and $P := \text{Prob}\{\|f\| \le \sqrt{2N}\}$. Note that (following the same arguments as in the proof of Lemma 2):

$$\begin{aligned} P =& \frac{1}{\pi^N} \int_{\|f\|\le\sqrt{2N}} e^{-\|f\|^2}\, df \\ =& \frac{\text{vol}(\mathbb{S}(\mathbb{R}^{2N}))}{\pi^N} \int_0^{\sqrt{2N}} t^{2N-1} e^{-t^2}\, dt \\ \overset{s=t^2}{=}& \frac{1}{\Gamma(N)} \int_0^{2N} s^{N-1} e^{-s}\, ds \ge \frac{1}{2}, \end{aligned}$$

the last inequality from [13, Th. 1]. We thus have

$$\rho(f) \le 2\varphi(f). \tag{3.22}$$

Then, using (3.21),

$$H \;\le\; \sqrt{2N} C'\, D^{3/2} \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}^{\sqrt{2N}}} \frac{1}{\mathcal{D}} \sum_{i=1}^{\mathcal{D}} \int_{t_1}^{1} \frac{\mu^2(f_t, \zeta_t^{(i)})}{\|f_t\|^2}\, dt.$$

From Lemma 9 we have

$$t_1 \ge \frac{c'}{D^{3/2}\sqrt{N}\mu^2(g,\zeta)}$$

for a constant c' (different from, but close to, c). We thus have proved that there are constants C'', c' such that

$$\begin{aligned} H \;&\le C''\sqrt{N}\, D^{3/2} \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}^{\sqrt{2N}}} \frac{1}{\mathcal{D}} \sum_{i=1}^{\mathcal{D}} \int_{\frac{c'}{D^{3/2}\sqrt{N}\mu^2(g,\zeta)}}^{1} \frac{\mu^2(f_t, \zeta_t^{(i)})}{\|f_t\|^2}\, dt \\ &= C''\sqrt{N}\, D^{3/2} \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}^{\sqrt{2N}}} \int_{\frac{c'}{D^{3/2}\sqrt{N}\mu^2(g,\zeta)}}^{1} \frac{\mu_{\text{av}}^2(f_t)}{\|f_t\|^2}\, dt. \end{aligned}$$

Using (3.22) we deduce that

$$\begin{aligned} H &\leq 2C''\sqrt{N}\,D^{3/2} \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}} \int_{\frac{c'}{D^{3/2}\sqrt{N}\mu^2(g,\zeta)}}^1 \frac{\mu^2_{\mathrm{av}}(f_t)}{\|f_t\|^2}\,dt \\ &\leq 2C''\sqrt{N}\,D^{3/2} \int_{\frac{c'}{D^{3/2}\sqrt{N}\mu(g,\zeta)^2}}^1 \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}} \frac{\mu^2_{F,\mathrm{av}}(f_t)}{\|f_t\|^2}\,dt. \end{aligned}$$

We next bound the expectation in the right-hand side using Theorem 2 and the fact that $f_t \sim N((1-t)g, t^2\mathrm{Id})$ and obtain

$$\begin{aligned} H &\leq 2nC''\sqrt{N}\,D^{3/2} \int_{\frac{c'}{D^{3/2}\sqrt{N}\mu^2(g,\zeta)}}^1 \frac{1}{t^2}\,dt \\ &\leq C'''\,D^3 nN\mu^2(g,\zeta), \end{aligned} \tag{3.23}$$

with C''' yet another constant.

Having reached thus far, the major obstacle we face is that the quantity H, for which we derived the bound (3.23), is an average over all initial zeros of h_{s_1} (as well as over f). None of the two solutions below is fully satisfactory but together they can handle a broad range of pairs (n, D) with a moderate complexity.

Case 1: $D > n$**.** Consider any $g \in \mathbb{S}$, ζ a well-posed zero of g, and let $\zeta^{(1)}, \ldots, \zeta^{(\mathcal{D})}$ be the zeros of h_{s_1}. Note that when f is Gaussian, these are $\mathcal{D}$ different zeros almost surely. Clearly,

$$\begin{aligned} \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}} K(f,g,\zeta) &\leq 1 + \mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}} \sum_{i=1}^{\mathcal{D}} K(f, h_{s_1}, \zeta^{(i)}) \\ &= 1 + \mathcal{D}\,H = \mathcal{O}(\mathcal{D}D^3 Nn\mu^2(g,\zeta)) \end{aligned}$$

the last by (3.23). We now take as initial pair (g,ζ) the pair $(\overline{g}, e_0)$ where $\overline{g} = (\overline{g}_1, \ldots, \overline{g}_n)$ is given by

$$\overline{g}_i = \sqrt{\frac{d_i}{n}} X_0^{d_i-1} X_i, \qquad \text{for } i = 1, \ldots, n$$

(the scaling factor guaranteeing that $\|\overline{g}\| = 1$) and $e_0 = (1, 0, \ldots, 0) \in \mathbb{C}^{n+1}$. It is easy to see that $\mu(\overline{g}, e_0) = \sqrt{n}$ (and that all other zeros of $\overline{g}$ are ill-posed, but this is not relevant for our argument). Replacing this equality in the bound above we obtain

$$\mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}} K(f, \overline{g}, e_0) = \mathcal{O}(\mathcal{D}D^3 Nn^2), \tag{3.24}$$

which implies an average cost of $\mathcal{O}(\mathcal{D}D^3N^2n^2)$ since the number of operations at each iteration of the homotopy algorithm is $\mathcal{O}(N)$ (see [12, Proposition 16.32]).

For any $\varepsilon > 0$ this quantity is polynomially bounded in N provided $D \geq n^{1+\varepsilon}$ and is bounded as $N^{\mathcal{O}(\log\log N)}$ when D is in the range $[n, n^{1+\varepsilon}]$ ([10, Lemma 11.1]).

Case 2: $D \leq n$**.** The occurrence of $\mathcal{D}$ makes the bound in (3.24) too large when D is small. In this case, we consider the initial pair $(\overline{U}, \mathbf{z}_1)$ where $\overline{U} \in \mathcal{H}_{(d)}$ is given by

$$\overline{U}_1 = \frac{1}{\sqrt{2n}}(X_0^{d_1} - X_1^{d_1}), \ldots, \overline{U}_n = \frac{1}{\sqrt{2n}}(X_0^{d_n} - X_n^{d_n}),$$

(the scaling factor guaranteeing that $\|\overline{U}\| = 1$) and $\mathbf{z}_1 = (1, 1, \ldots, 1)$. We denote by $\mathbf{z}_1, \ldots, \mathbf{z}_{\mathcal{D}}$ the zeros of $\overline{U}$.

The reason for this choice is a strong presence of symmetries. More exactly, for any $i \neq j$ there exists a unitary matrix U_{ij} of size $n+1$ such that $U_{ij}\mathbf{z}_i = \mathbf{z}_j$ and $\overline{U} \circ (U_{ij})^* = \overline{U}$. That is,

$$(\overline{U} \circ (U_{ij})^*, U_{ij}\mathbf{z}_i) = (\overline{U}, \mathbf{z}_j).$$

In particular, from (1.2) and the unitary change of variables $f \mapsto f \circ (U_{ij})^*$ we have

$$\mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} K(f, \overline{U}, \mathbf{z}_1) = \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} \mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} K(f, \overline{U}, \mathbf{z}_j).$$

These symmetries also guarantee that, for all $1 \leq i, j \leq \mathcal{D}$,

$$\mu(\overline{U}, \mathbf{z}_i) = \mu(\overline{U}, \mathbf{z}_j), \tag{3.25}$$

and, consequently, that the value of s_1 is the same for all the zeros of $\overline{U}$. Hence,

$$\begin{aligned}
\mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} K(f, \overline{U}, \mathbf{z}_1) &= \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} \mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} K(f, \overline{U}, \mathbf{z}_j) = \mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} K(f, \overline{U}, \mathbf{z}_j) \\
&\leq \mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} \Big(1 + K(f, h_{s_1}, \zeta^{(j)})\Big) \qquad (3.26) \\
&= 1 + \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} H(\overline{U}, \mathbf{z}_i) = 1 + H(\overline{U}, \mathbf{z}_1),
\end{aligned}$$

the last equality because the unique dependence on j of $H(\overline{U}, \mathbf{z}_j)$ is in the value of s_1 and as said above this value is independent of j.

Note now that for $i \neq j$, the isometric change of variables $f \mapsto f \circ (U_{ij})^*$ gives

$$\mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} \Big(K(f, h_{s_1}, \zeta^{(j)})\Big) = \mathop{\mathbb{E}}_{f \in \mathcal{H}_{(d)}} \frac{1}{\mathcal{D}} \sum_{j=1}^{\mathcal{D}} \Big(K(f \circ U_{ij}, h_{s_1}, \zeta^{(j)})\Big)$$

That is, the average (w.r.t. f) number of homotopy steps with initial system $\overline{U}$ is the same no matter whether the zero of $\overline{U}$ is taken at random or set to be $\mathbf{z}_1$. Also,

$$\mu^2(\overline{U}, \mathbf{z}_1) \leq 2\,(n+1)^D \tag{3.27}$$

(actually such bound holds for all zeros of $\overline{U}$ but, again, this is not relevant for our argument). Both (3.25) and (3.27) are proved in [10, Section 10.2]. It follows from (3.26), (3.23), and (3.27) that

$$\mathop{\mathbb{E}}_{f\in\mathcal{H}_{(d)}} K(f, \overline{U}, \mathbf{z}_1) = \mathcal{O}(D^3 N n^{D+1}). \tag{3.28}$$

As above, for any fixed $\varepsilon > 0$ this bound is polynomial in N provided $D \leq n^{\frac{1}{1+\varepsilon}}$ and is bounded by $N^{\mathcal{O}(\log\log N)}$ when $D \in [n^{\frac{1}{1+\varepsilon}}, n]$. □

References

[1] D. Armentano. *Stochastic perturbation and smooth condition numbers*. Journal of Complexity 26 (2010) 161–171.

[2] D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, M. Shub. *A stable, polynomial-time algorithm for the eigenpair problem.* Preprint, available at `arXiv:1410.0116`.

[3] C. Beltrán. *A continuation method to solve polynomial systems and its complexity*. Numer. Math. 117 (2011), no. 1, 89–113.

[4] C. Beltrán and A. Leykin. Robust certified numerical homotopy tracking. *Found. Comput. Math.*, 13(2):253–295, 2013.

[5] C. Beltrán and L. M. Pardo. *Smale's 17th problem: average polynomial time to compute affine and projective solutions*. J. Amer. Math. Soc. 22 (2009), no. 2, 363–385.

[6] C. Beltrán and L. M. Pardo. *Fast linear homotopy to find approximate zeros of polynomial systems.* Found. Comput. Math. 11 (2011), no. 1, 95–129.

[7] C. Beltrán, M. Shub. *The complexity and geometry of numerically solving polynomial systems*. Contemporary Mathematics, volume 604, 2013, pp. 71–104.

[8] C. Beltrán, M. Shub. *On the Geometry and Topology of the Solution Variety for Polynomial System Solving*. Found. Comput. Math. 12 (2012), 719–763.

[9] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and real computation*, Springer-Verlag, New York, 1998.

[10] P. Bürgisser and F. Cucker. *On a problem posed by Steve Smale*, Ann. of Math. (2) 174 (2011), no. 3, 1785–1836.

[11] P. Bürgisser, M. Clausen and A. Shokrollahi. *Algebraic Complexity Theory*, volume 315 of *Grundlehren der mathematischen Wissenschaften*, Springer-Verlag, Berlin, 1996.

[12] P. Bürgisser and F. Cucker. *Condition*, volume 349 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2013.

[13] K.P. Choi. *On the medians of gamma distributions and an equation of Ramanujan*, Proc. Amer. Math. Soc. 121 (1994), no. 1, 245–251.

[14] J-P. Dedieu, G. Malajovich, and M. Shub. *Adaptative step size selection for homotopy methods to solve polynomial equations*. IMA Journal of Numerical Analysis 33 (2013), no. 1, 1–29.

[15] P. Lairez. *A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time.* Preprint, available at `arXiv:1507.05485`.

[16] M. Shub. *Complexity of Bezout's theorem. VI. Geodesics in the condition (number) metric*. Found. Comput. Math. 9 (2009), no. 2, 171–178.

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

[18] M. Shub and S. Smale. *Complexity of Bézout's theorem. I. Geometric aspects.* J. Amer. Math. Soc. 6 (1993), no. 2, 459-501.

[19] M. Shub and S. Smale. *Complexity of Bézout's theorem. II: volumes and probabilities.* Computational Algebraic Geometry. Progress in Mathematics Volume 109, 1993, pp 267–285

[20] M. Shub and S. Smale. *Complexity of Bézout's theorem. V: Polynomial time.* Theoretical Computing Science, Vol 133, 1994, pag 141–164.

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