FAST LINEAR HOMOTOPY TO FIND APPROXIMATE ZEROS OF POLYNOMIAL SYSTEMS

CARLOS BELTRÁN AND LUIS MIGUEL PARDO

Abstract. We prove a new complexity bound, polynomial on the average, for the problem of finding an approximate zero of systems of polynomial equations. The average number of Newton steps required by this method is almost linear in the size of the input. We show that the method can also be used to approximate several or all the solutions of non–degenerate systems, and prove that this last task can be done in running time which is linear in the Bézout number of the system, on the average.

Contents

Key words and phrases. Approximate zero, homotopy method, average complexity.

Research was partially supported by the a postdoctoral grant from FECYT (Fundación Española para la Ciencia y la Tecnología) and MTM2007-62799, Spanish Ministry of Science (MICINN).

1. Introduction

In recent papers [BePa2, BePa3], we have presented an Average Las Vegas algorithm, which is a probabilistic solution to Smale's 17^{th} Problem [Sma]:

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

In the present – pretty self contained – paper, we present a new algorithm which admits a much simpler analysis. The complexity results of our previous papers are greatly improved, constants and exponents are sharpened, and we provide a shorter, more general proof. Moreover, the new proof allows us to extend the result for the search of more than one solution. Our model of computation is the Blum–Shub–Smale model (see [BSS] or [BCSS, Ch. 2 and 17]), so exact arithmetic computations are assumed. As in [Sma], we consider the homogeneous version of the problem, so our systems consist on n homogeneous polynomial equations with $n+1$ unknowns.

Our algorithm is, as some of the most popular polynomial system solvers, based in the general idea of homotopy (or continuation) methods: let f be the system we want to solve, and let g be another system which has a known solution ζ_0. Let f_t, $0 \leq t \leq T$ be a path in the vector space of systems with extremes $g = f_0$ and $f = f_T$. Then, one might attempt to lift the path f_t to the solution path (f_t, ζ_t) with extremes (f_0, ζ_0) and (f_T, ζ_T) where ζ_t is a zero of f_t for all t. In particular, the method should produce an approximation to ζ_T, a zero of f_T.

The reader may note that there are two key ingredients for this general approach:

- *The choice of a path f_t which can actualy be lifted.*
- *The choice of the path–following method used to lift the path and thus produce an approximation to ζ_T.*

When studying the complexity of such an algorithm, both ingredients must be carefully designed and one must try to get a path f_t such that the chosen path–following method requires a small number of arithmetic operations to lift f_t.

There are several approaches to the choice of the path f_t. Among the most popular ones is linear homotopy, that is the path f_t is the linear path $f_t = (1 - t)g + tf$. In this paper we use the great circles homotopy that we define now: [1] first, note that the zeros of f do not change if f is multiplied by a non–zero number. Hence, it suffices to consider systems in the unit sphere $\mathbb{S}$ contained in the vector space of systems, see Section 2.1 for a detailed definition. Then, on input f, the algorithm will choose some $g \in \mathbb{S}$ with a known zero ζ_0, and f_t will then be the shortest portion of the great circle in $\mathbb{S}$ joining g and f. There is just one thing to decide:

- *Where to begin the homotopy? Namely, how to choose (g, ζ_0)?*

The literature contains different approaches to solving this problem. One of them is Shub & Smale's conjecture in [ShSm5] where a particularly simple initial pair is conjectured to be a good starting point for the linear homotopy, in the sense that it allows to perform the path–following procedure in average polynomial time. This is still an open conjecture. Other initial systems were considered in

[1] From a theoretical point of view, geodesics in the so–called condition metric (see [Shu2, BeSh]) are known to have much better properties for path–following methods, but up to know there is no practical way to construct these geodesics.

[Ren, AlGe, Ver1, LLT] (see also references in [BCSS, SoWa]). Among the most popular ones is the system having as zeros the roots of the unity.

In [BePa2, BePa3] we used a different approach: instead of introducing a deterministic algorithm we introduced a randomized one. Thus, no precise initial pair is given, but we admitted randomized guessing in some special set of pairs that we called *questor set*.

Our new algorithm here also admits randomized guessing but in a set which is different from the one used in [BePa2, BePa3]. This is the first novelty in our pages here. The precise description of this new questor set is in Section 2.3.

There also exist different ways to define the path–following method. For instance, in [BePa2] we used a constant–step Newton procedure to approximate the lifted path. In [BePa3] we chose the path–following method described in [ShSm5], thus yielding, for the first time, an average polynomial time procedure. Other path–following methods have been described in the above cited literature, some times accompained by complexity analysis.

It was recently pointed out by Shub [Shu2] the existence of a fast path–following method which admits a complexity analysis, see equation (2.3) below. This method has been explicitely constructed in [Bel], and is the one that we choose here. It performs a number of "homotopy steps", each of them being one application of projective Newton's method.

Our algorithm thus uses the two ingredients outlined above (choice of the path f_t with randomized (g, ζ_0) and the path–following method of [Shu2, Bel]). We may informaly write our main outcome as follows.

Theorem 1. *The average running time of the algorithm* AHMR *described in these pages is $O\tilde{}(N^2)$, where N is the input size (dense enconding). The average number of homotopy steps is linear in N. In particular, it is an Average Las Vegas algorithm that answers affirmatively Smale's 17^{th} problem.*

See Section 2.5 for the description of AHMR. Here we use the $O\tilde{}$ (soft-Oh) notation, that is $R_k \leq O\tilde{}(k)$ means that there exist constants C, c such that $R_k \leq Ck(\log k)^c \, \forall k \in \{1, 2, \ldots\}$. The technical version of this result (Corollary 9 below) computes explicitly the bound and is thus more precise.

Our algorithm computes an approximation to some zero of the input system f. We may then ask: *what kind of questions can be answered with this information?* A framework to study this question is that of universal system solvers. Consider the following "Nullstellensatz–like" question:

Problem (Approximate Nullstellensatz)**.** *Given $\varepsilon > 0$ and complex polynomials $f_1, \ldots, f_n, g$ in variables $X_1, \ldots, X_n$, decide whether there is some solution x of $f_1 = 0, \ldots, f_n = 0$ such that $|g(x)| < \varepsilon$.*

This problem deals with affine (not projective) solutions, while in this paper we center our attention in the homogeneous case. The relation between these two cases is however well understood, see for example [BePa3].

A polynomial system solver is called universal (cf. [CGH$^+$03] or [BePa1]) if the information contained in its output suffices to answer questions like this one. Symbolic solvers are typically universal, and so are the numerical solvers which compute all the solutions of f. On the other hand, a system solver is called non–universal if it attempts to approximate only one or few solutions of f, thus producing less information but doing it faster.

Some solvers can be used as universal or non–universal depending on the needs of the user. For example, most procedures (as, for instance, those quoted in [SoWa]) can be used to find only one, maybe a few or perhaps all solutions of f. On the other hand, some non–universal solvers are originally designed to find just one solution, and it is not clear that we can modify any of them to become universal. No general method is known and, perhaps, it does not exists. Up to our knowledge, the unique solvers with proven average polynomial running time are the one introduced in [BePa3] and the one we introduce here. Both of them were initially designed to find just one zero. We do not know whether we could transform the algorithm in [BePa3] to get a universal version of it.

In the case of probabilistic algorithms, an obvious way to get more zeros of f is running the algorithm several times, and hoping that it will produce different solutions. We claim the following result, which proves that our algorithm admits this strategy.

Theorem 2. *Let f be a non–degenerate system (i.e. f has no singular zeros). Then, every zero ζ is equally probable as an output of our algorithm* AHMR.

Note that the set of non–degenerate systems is a Zariski open set in $\mathbb{S}$. In particular, its complement has zero measure. Theorem 2 is proved in Section 10.

An inmediate consequence is the following (see corollaries 26 and 27 for details.)

Corollary 3. *Algorithm* AHMR *may be adapted to compute k approximate zeros of k different solutions of a non–degenerate input system, in average time $\tilde{O}(kN^2)$, accepting a small probability that less than k are found. In particular, finding all the $\mathcal{D}$ (Bézout number) solutions of a non–degenerate system can be done in average time $\tilde{O}(\mathcal{D}N^2)$, with probability of success greater than $1 - 1/\mathcal{D}$.*

The proof of Theorem 2 will easily follow from our analysis of the algorithm. There exists no similar result for the algorithms we introduced in [BePa2] and [BePa3]: these last two methods may produce different answers each time they run, but the exact behavior of the output is not well understood, see for example the comment in the Review [Ver2]. A natural concept to study how close the distribution of the output is to the uniform distribution is Shannon's entropy as introduced in Section 9.

As another example, the linear homotopy method with the initial pair from Shub & Smale's conjecture (which is a non–universal solver) produces always the same zero of any fixed input system f. This algorithm can be modified by a random unitary transformation applied to the initial pair, but no result in the lines of Theorem 2 is known for this modification. Equidistribution of the output with this randomized version of Shub–Smale's conjecture has been experimentaly confirmed (see [BeLe]) for degree 2 systems by computing its Shannon's entropy.

There are several remarkable technical results in this paper, one of them being the Main Lemma of Section 5 below, which claims that the Normal Jacobian of certain natural mapping from the solution variety (i.e. the set of pairs $(system, solution)$ defined in Section 2) onto its linear counterpart is constant. Thanks to this fact we can greatly simplify the computations of many integrals, state the precise relation between the expectation and moments of the linear and non–linear condition numbers, and produce a short and elegant proof of the existence and generation of good starting pairs. The reader familiar to the Bézout series and our previous papers may find in this result the greatest novelty of these pages.

The manuscript is structured as follows. In Section 2 we introduce the basic notation and the precise statements of our main results. In Section 3 we recall the geometrical background of the problem and we prove some preliminary results which are essentially contained in the literature. In Section 4 we prove an Integral Geometry result needed for our proofs. In Section 5 we prove the Main Lemma. In Section 6 we use the Main Lemma to prove a formula that helps to compute many integrals in the solution variety. In Section 7 the exact moments of the condition number in the linear case are computed. In Section 8 we combine the reduction formulas of Section 6 and the results of Section 7 to compute the exact moments of the condition number in the non–linear case, and the average complexity of linear homotopy with random initial pair. In Section 9 we use again the reduction formulas of Section 6 to prove that random initial pairs can be obtained by a simple procedure. Finally, in Section 10 we prove equidistribution in the output of the algorithm.

Acknowledgments. Thanks to Michael Shub for many insightful comments and long discussions, some of which lead to the simplification of several proofs in this manuscript. Thanks to Alan Edelman for pointing out to us that exact computation of the moments of the linear condition number was possible. Thanks to the referees for very helpful comments and suggestions.

2. Description of the main results

2.1. **Metrics, solution variety and condition number.** Let $\mathcal{H}_{(d)}$ be the vector space of all systems of n homogeneous polynomial equations of degrees $(d_1,\ldots,d_n)=(d)$ with complex coefficients and variables $X_0,\ldots,X_n$. Elements in $\mathcal{H}_{(d)}$ are n–tuples $f=(f_1,\ldots,f_n)$ where f_i is a homogeneous polynomial of degree d_i. Sometimes we think on f as a vector in a high-dimensional vector space, containing the coefficients of the monomials of the f_i.

Let $\mathcal{H}_{(d)}$ be equipped with the unitarily-invariant, Bombieri-Weyl Hermitian product, sometimes called Kostlan product (cf. for example [ShSm1] or [BCSS, Sec. 12.1]). Namely, if $f=(f_1,\ldots,f_n)\in\mathcal{H}_{(d)}$ and $g=(g_1,\ldots,g_n)\in\mathcal{H}_{(d)}$ with

$$f_i(X)=\sum_{\substack{(\alpha)=(\alpha_0,\ldots,\alpha_n)\\ \alpha_0+\cdots+\alpha_n=d_i}} a_{(\alpha),i}X_0^{\alpha_0}\cdots X_n^{\alpha_n},\quad g_i(X)=\sum_{\substack{(\alpha)=(\alpha_0,\ldots,\alpha_n)\\ \alpha_0+\cdots+\alpha_n=d_i}} b_{(\alpha),i}X_0^{\alpha_0}\cdots X_n^{\alpha_n},$$

then

$$\langle f,g\rangle=\sum_{i=1}^n\langle f_i,g_i\rangle,\quad \langle f_i,g_i\rangle=\sum_{\substack{(\alpha)=(\alpha_0,\ldots,\alpha_n)\\ \alpha_0+\cdots+\alpha_n=d_i}}\binom{d_i}{(\alpha)}^{-1}a_{(\alpha),i}\overline{b}_{(\alpha),i},$$

where

$$\binom{d_i}{(\alpha)}=\frac{d_i!}{\alpha_0!\cdots\alpha_n!}\text{ is the multinomial coefficient,}$$

and $\overline{\cdot}$ denotes complex conjugation.

One of the main properties of the Bombieri-Weyl product is the invariance under unitary changes of coordinates. Namely, if U is a $(n+1)\times(n+1)$ unitary matrix then the map $\mathcal{H}_{(d)}\to\mathcal{H}_{(d)}$ sending f to $f\circ U^*$ is an isometry[2]. With this product, the space $\mathcal{H}_{(d)}$ is a complex Hilbert space. We denote by $\mathbb{S}=\{f\in\mathcal{H}_{(d)}:\|f\|=1\}$

[2] It is a common practice to write $f\circ U^*$ instead of $f\circ U$. The reason is that the mapping $(U,f)\to f\circ U^*$ defines a left action of the unitary group in $\mathcal{H}_{(d)}$, which is useful in some contexts.

the unit sphere in $\mathcal{H}_{(d)}$, and equip it with Riemannian structure inherited from $\mathcal{H}_{(d)}$.

The projective solution set $V(f) = \{\zeta \in \mathbb{P}(\mathbb{C}^{n+1}) : f(\zeta) = 0\}$ of a generic system f consists of $\mathcal{D} = d_1 \cdots d_n$ points. The set of solutions $V_{\mathbb{S}}(f) = \{\hat{\zeta} \in \mathbb{S}(\mathbb{C}^{n+1}) : f(\hat{\zeta}) = 0\}$ where $\mathbb{S}(\mathbb{C}^{n+1})$ is the unit sphere of $\mathbb{C}^{n+1}$, consists of $\mathcal{D}$ great circles.

Here we take the solution variety as the set $V = V_{(d)} = \{(f, \zeta) \in \mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1}) : \zeta \in V(f)\}$, which is a smooth (real) submanifold of $\mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1})$ (cf. [BCSS, p. 193]). Note that in some papers, the solution variety is considered as a subset of $\mathbb{P}(\mathcal{H}_{(d)}) \times \mathbb{P}(\mathbb{C}^{n+1})$ instead of $\mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1})$ as here. We equip V with the metric induced by the product metric in $\mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1})$. The unitary invariance of the Bombieri-Weyl Hermitian product implies that for any unitary matrix U, the mapping $(f, \zeta) \mapsto (f \circ U^*, U\zeta)$ is an isometry of V (cf. [BCSS, Lemma 3, p. 222]). Let $\Sigma' = \{(f, \zeta) \in V : \det(Df(\zeta)Df(\zeta)^*) = 0\}$.

Recall the condition number of [ShSm1, BCSS],

$$\mu_{\text{norm}}(f, \zeta) = \|(Df(\zeta) \mid_{\zeta^\perp})^{-1} Diag(\|\zeta\|^{d_i - 1} d_i^{1/2})\|, \quad (f, \zeta) \in V \setminus \Sigma',$$

or $\mu_{\text{norm}}(f, \zeta) = \infty$ if $(f, \zeta) \in \Sigma'$. Here, $\| \cdot \|$ is the operator norm of a linear map. An equivalent formula is

$$\mu_{\text{norm}}(f, \zeta) = \|(Diag(\|\zeta\|^{1 - d_i} d_i^{-1/2}) Df(\zeta))^\dagger\|, \quad (f, \zeta) \in V \setminus \Sigma', \tag{2.1}$$

where $Df(\zeta))^\dagger = Df(\zeta)(Df(\zeta)Df(\zeta)^*)^{-1}$ denotes the Moore-Penrose pseudo-inverse of the full–rank matrix $Df(\zeta)$. Note that the affine representative chosen for ζ used to compute $\mu_{\text{norm}}(f, \zeta)$ is not relevant, namely μ_{norm} is well–defined as a mapping in $V \subseteq \mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1})$, as far as $Df(\zeta)$ is of maximal rank.

2.2. **Approximate zeros and the linear homotopy method.** An approximate zero $z_0 \in \mathbb{P}(\mathbb{C}^{n+1})$ of $f \in \mathcal{H}_{(d)}$ is a projective point, such that successive iterations of projective Newton's method of [Shu1], $z \mapsto z - (Df(\zeta) \mid_{\zeta^\perp})^{-1} f(z)$, converge quadratically to an actual zero $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$ of f (cf. [Shu1, Shu2] for background). Namely,

$$d_R(z_l, \zeta) \leq \frac{1}{2^{2^l - 1}} d_R(z_0, \zeta)$$

where d_R is the Riemannian distance in $\mathbb{P}(\mathbb{C}^{n+1})$ and z_l is the l–th iteration of projective Newton's method, with starting point z_0. Sometimes a distance different from d_R is used for this definition, but here we follow [Shu2] where the Riemannian distance d_R is used.

The linear homotopy method is a procedure designed to approximate solution paths in V. Given an input system $f \in \mathbb{S}$ and a pair $(g, \zeta_0) \in V$ such that $f \neq -g$, the linear homotopy generates a polygonal line that approximates the curve $\Gamma(f, g, \zeta_0) = \{(f_t, \zeta_t)\} \subseteq V$, where f_t parametrizes the short portion of the great circle joining g and f and ζ_t is defined by continuation. We can write the following formula for f_t,

$$t \to f_t = g \cos(t) + \frac{f - Re(\langle f, g \rangle) g}{\sqrt{1 - Re(\langle f, g \rangle)^2}} \sin(t), \quad t \in [0, d_R(f, g)],$$

where $d_R(f, g)$ is the Riemannian distance between f and g. This path is well-defined and smooth if some regularity conditions are satisfied, i.e. not intersecting the variety Σ' (cf. [ShSm1]).

Associated with $\Gamma(f,g,\zeta_0)$ we consider the quantity

$$\mathcal{C}(f,g,\zeta_0)=\int_{h\in L_{f,g}}\mu_{\mathrm{norm}}(h,\zeta_h)^2\,dL_{f,g}, \tag{2.2}$$

where ζ_h is the unique solution of h lying in $\Gamma(f,g,\zeta_0)$ and $L_{f,g}$ is the (shortest) portion of the great circle joining f and g.

In [Shu2, Corollary 1], it was proven that the number of steps of projective Newton method sufficient to approximate the path $\Gamma(f,g,\zeta_0)$, and thus produce an approximate zero of f, is at most $Cd^{3/2}\mathcal{C}(f,g,\zeta_0)$, where C is a constant and $d=\max\{d_i : 1\leq i\leq n\}$. The actual algorithm was not described in [Shu2], only its existence was proved. The specific description of the method has been done in [Bel]. An alternative has recently been presented in [BuCu]. In [BeLe, Ley], an implementation of the homotopy algorithm of [Bel] has been implemented in Macaulay2[3].

With the assumption of exact arithmetic computations of the Blum–Shub–Smale model of computation, the homotopy method of [Bel] is *guaranteed* to produce an approximate zero of f. The total number of projective Newton method steps is at most

$$Cd^{3/2}\mathcal{C}(f,g,\zeta_0),\ \text{with } C \text{ a constant.} \tag{2.3}$$

If this number is not finite, then the method shall never end.

Remark 4. *The number of projective Newton steps, both in theoretical studies (*[Shu2, Bel]*) and practical experiments (*[BeLe]*) can be bounded by the length of the path (f_t,ζ_t) in the so–called condition metric (i.e. the condition length of the path). This provides a better bound than the one of equation (2.2), but it is not know how to transfer it into algorithmic design. We do not discuss these details here, and we simply use the fact that the condition length of (f_t,ζ_t) is at most $\sqrt{2}\mathcal{C}(f,g,\zeta_0)$.*

Let $N+1$ be the complex dimension of $\mathcal{H}_{(d)}$. From [BaSt], a polynomial system f and all its partial derivatives can be evaluated in $O(N)$. Adding the cost of solving the linear system of Newton's operator, the number of arithmetic operations needed for Newton's method is thus $O(N+n^3)$. Hence, the total number of arithmetic operations needed to produce an approximate zero of f by approximating the path $\Gamma(f,g,\zeta_0)$, is at most

$$Cd^{3/2}(N+n^3)\mathcal{C}(f,g,\zeta_0),\ \text{C a constant.} \tag{2.4}$$

Thus, to study the complexity of this algorithm we must bound $\mathcal{C}(f,g,\zeta_0)$.

Our first result claims that the average value of $\mathcal{C}(f,g,\zeta_0)$ is surprisingly small for random choices of f,g and solution $\zeta_0\in V(g)$.

Theorem 5. *Let* E *denote expectation. Then,*

$$\mathrm{E}_{f\in\mathbb{S}}\left(\mathrm{E}_{g\in\mathbb{S}}\left(\frac{1}{\mathcal{D}}\sum_{\zeta_0\in V(g)}\mathcal{C}(f,g,\zeta_0)\right)\right)=\frac{\pi}{2}N\left(n\left(1+\frac{1}{n}\right)^{n+1}-2n-1\right)\leq\frac{\pi}{2}nN.$$

[3] Such an implementation is the first one that has certified output and at the same time attains the complexity bounds of [Shu2]. Comparisons with [LLT, BHSW, SoWa, Ver1, AlGe] are also discussed there.

The proof of this result is in Section 8. As a self–interesting intermediate result we obtain the exact moments of the condition number both in the linear and non–linear cases (see Theorem 23 and Section 7).

Theorem 5 already suggests a strategy for choosing an initial pair (g, ζ_0) for the path-following algorithm: First, choose at random a system g, then randomly choose a solution ζ_0 of g. According to Theorem 5, this procedure is expected to produce initial pairs with small average value of $\mathcal{C}(f, g, \zeta_0)$. However, this choosing procedure may look difficult, as it requires to solve a random system of equations, while the problem treated in this paper is precisely system solving. Our second result will prove an alternative to this process , which uses only simple procedures from linear algebra.

2.3. **How to randomly choose a root of a randomly chosen polynomial system.**

In short, the alternative method for randomly choosing an initial pair (g, ζ_0) works as follows: Choose at random a full rank $n \times (n+1)$ matrix M, and compute its solution ζ_0. Then, construct a polynomial system with solution ζ_0 whose "linear part" at ζ_0 is given by M, and add a random higher-degree term. Linear and non-linear parts must be correctly weighted.

We provide now the precise description of this process, which requires the introduction of some notation. Given a Hilbert space W, we denote by $B(W)$ and $\mathbb{S}(W)$ the unit ball and the unit sphere in W. For $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$ we consider the vector subspaces of $\mathcal{H}_{(d)}$,

$$R_\zeta = \{h \in \mathcal{H}_{(d)} : h(\zeta) = 0,\ Dh(\zeta) = 0\}, \quad L_\zeta = (R_\zeta)^\perp.$$

The structure of R_ζ and L_ζ are better understood if we first fix $\zeta = e_0 = (1, 0, \ldots, 0)^T$. Indeed, R_{e_0} is the set of polynomial systems $h = (h_1, \ldots, h_n) \in \mathcal{H}_{(d)}$ such that $h(e_0) = 0$ and $Dh(e_0) = 0$, namely

$$h_i(X) = X_0^{d_i - 2} p_{d_i - 2}(X_1, \ldots, X_n) + \cdots + X_0 p_1(X_1, \ldots, X_n) + p_0(X_1, \ldots, X_n),$$

for some polynomials $p_j, 0 \leq j \leq d_i - 2$. Thus, a polynomial system h is in R_{e_0} if all the coefficients of the monomials containing $X_0^{d_i}$ and $X_0^{d_i - 1}$ are zero. Reciprocally, a polynomial system h is in L_{e_0} if all the non zero monomials contain $X_0^{d_i}$ or $X_0^{d_i - 1}$. Note that for such a $h \in L_{e_0}$ we have that $h(1, X_1, \ldots, X_n)$ defines a linear function of $X_1, \ldots, X_n$. Thus, for any $h \in \mathcal{H}_{(d)}$ we can think on the orthogonal projection of h onto L_ζ as the "linear part" of h with respect to e_0.

Now, let $\zeta \in \mathbb{S}(\mathbb{C}^{n+1})$ and consider a $(n+1) \times (n+1)$ unitary matrix U such that $Ue_0 = \zeta$. Then, by the unitary invariance of the Bombieri–Weyl product in $\mathcal{H}_{(d)}$ we have

$$R_\zeta = \{h \circ U^* : h \in R_{e_0}\}, \quad L_\zeta = \{h \circ U^* : h \in L_{e_0}\}.$$

Choosing a random point in $B(R_\zeta)$ is now easy: just choose a random point in the more simple space $B(R_{e_0})$, find a unitary matrix U whose first column is ζ, and construct the system $h \circ U^*$.

Note that $\mathcal{H}_{(1)}$ is the set of $n \times (n+1)$ matrices, with the usual Frobenius norm. For $M \in \mathcal{H}_{(1)}$ and $\zeta \in V_{\mathbb{S}}(M) = \{\zeta \in \mathbb{S}(\mathbb{C}^{n+1}) : M\zeta = 0\}$ (the unit norm affine zeros of M), let $\varphi(M, \zeta) \in L_\zeta$ be the system of equations defined by

$$\varphi(M, \zeta)(z) = Diag(\langle z, \zeta \rangle^{d_i - 1} d_i^{1/2}) M z. \tag{2.5}$$

Note that $\|\varphi(M,\zeta)\| = \|M\|_F$ and $D(\varphi(M,\zeta))(\zeta) = Diag(d_i^{1/2})M$ (see the proof of Lemma 21 below.) Again, these formulas become clearer if we first fix $\zeta = e_0$. Then, $M = (0 \mid A)$ where A is a square matrix of size n. Let a_{ij}, $1 \leq i,j \leq n$ be the entries of A, and let $\varphi(M,\zeta) = (f_1,\ldots,f_n)$. Then,

$$f_i(X_0,\ldots,X_n) = d_i^{1/2} X_0^{d_i-1} \sum_{1\leq j\leq n} a_{ij}X_j.$$

Note that $\varphi(M,e_0)$ is in L_{e_0}, and similarly $\varphi(M,\zeta) \in L_\zeta$ for any $(M,\zeta) \in V_{(1)}$.

Then, consider the set $Y \subseteq B(\mathbb{C}^{N+1}) \times \mathbb{S}(\mathbb{C}^{n+1}) \times B(\mathcal{H}_{(d)})$,

$$Y = \{(M,l,\zeta,h) : \det(MM^*) \neq 0, M\zeta = 0, h \in R_\zeta\}.$$

Here, a point $(M,l) \in \mathbb{C}^{N+1}$ consists of a $n \times (n+1)$ matrix M and a vector l containing the rest of coordinates. Then, $\|(M,l)\|^2 = \|M\|_F^2 + \|l\|^2$ is the usual product norm. We consider Y endowed with the σ-algebra inherited from the ambient space, and with measure given by

$$Vol(A) = \int_{(M,l)\in B(\mathbb{C}^{N+1})} \int_{\zeta\in V_{\mathbb{S}}(M)} \int_{h\in B(R_\zeta)} \chi_A(M,l,\zeta,h)\, dR_\zeta\, dV_{\mathbb{S}}(M)\, d\mathbb{C}^{N+1},$$

where χ_A is the characteristic function of a measurable set $A \subseteq Y$. Let

$$\begin{array}{cccc} G_{(d)} : & Y & \longrightarrow & V_{(d)} \\ & (M,l,\zeta,h) & \mapsto & (g/\|g\|,\zeta), \end{array}$$

where

$$g(z) = \sqrt{1-\|M\|^2}h(z) + \varphi(M,\zeta)(z).$$

Remark 6. *Note that Y has been endowed with a product-like measure. Hence, randomly choosing an element $(M,l,\zeta,h) \in Y$ amounts to choosing (M,l), choosing an element ζ of norm 1 in the kernel of M, and then choosing an element in the vector space R_ζ. See the appendix for details.*

Consider the set $\mathcal{G}_{(d)} = G_{(d)}(Y)$ with the push-forward measure inherited from $G_{(d)}$ (i.e. to choose a random point in $\mathcal{G}_{(d)}$, we choose a random point $y \in Y$ and compute $G_{(d)}(y)$). Note that $\mathcal{G}_{(d)} = V_{(d)} \setminus \Sigma'$ as sets, although they have different measures.

The following remarkable result justifies the definition of $\mathcal{G}_{(d)}$: Choosing random pairs in $\mathcal{G}_{(d)}$ serves to emulate the probability distribution obtained when choosing a random solution of a random system. But the former process is much easier!

Theorem 7. *Let $\Theta : V_{(d)} \longrightarrow [0,\infty)$ be a measurable mapping such that $\Theta(f,\zeta)$ is invariant under (real) scaling of f, and let $\mathcal{G}_{(d)}$ be the set defined above. Then,*

$$\mathrm{E}_{(g,\zeta)\in\mathcal{G}_{(d)}} (\Theta(g,\zeta)) = \mathrm{E}_{g\in\mathbb{S}} \left(\frac{1}{\mathcal{D}} \sum_{\zeta_0\in V(g)} \Theta(g,\zeta_0) \right).$$

Namely, randomly choosing a pair $(g,\zeta) \in \mathcal{G}_{(d)}$ is equivalent to randomly choosing a solution ζ_0 of a randomly-chosen, polynomial system $g \in \mathbb{S}$.

The proof of Theorem 7 is in Section 9.

A reader interested in the practical construction of random points in $\mathcal{G}_{(d)}$ might find useful the following recipe: Choosing a point at random $(g,\zeta) \in \mathcal{G}_{(d)}$ is done by choosing a point at random in $y = (M,l,\zeta,h) \in Y$ and taking

$$G_{(d)}(y) = (g/\|g\|,\zeta) \in \mathcal{G}_{(d)}, \text{ where } g(z) = \sqrt{1-\|M\|^2}h(z) + \varphi(M,\zeta)(z)$$

Once that $y = (M, l, \zeta, h)$ is fixed, it is easy to compute $G_{(d)}(y)$. On the other hand, to choose a random $y \in Y$ we must follow the following process:

(1) Choose at random $(M, l) \in \mathbb{C}^{n^2+n} \times \mathbb{C}^{N+1-n^2-n} = \mathbb{C}^{N+1}$ with the uniform distribution in the unit ball for the usual Euclidean norm in $\mathbb{C}^{N+1}$. Note that M is a (n^2+n)–dimensional complex vector, that we consider as a $n \times (n+1)$ complex matrix. At this point we may discard l and just keep M. The reader may think that choosing (M, l) is not a good idea, as we will immediately discard l. As a matter of fact, it turns out that the probability distribution needed for M is exactly the one obtained by choosing (M, l) and then projecting on the M coordinate. Thus, choosing (M, l) in the unit ball and then discarding l is precisely what we need to do.

(2) With probability 1, we have produced a matrix M whose kernel has complex dimension 1. We let ζ be a unit norm element of $Ker(M)$. Note that ζ has to be chosen at random in such kernel, so we may compute it by any means, and then multiply it by a complex number of modulus one chosen at random with the uniform distribution in the unit complex circle.

(3) Now we choose any unitary matrix U such that $Ue_0 = \zeta$. For example, we can start with some maximal–rank matrix whose first column is ζ and then apply Househölder reflections to get U. The matrix U can again be chosen by any means.

(4) Choose a system $\tilde{h}$ at random in $B(R_{e_0})$, that is the unit ball (for the Bombieri–Weyl norm) of R_{e_0}, which we have seen is a well–defined space with a very simple description. Then, consider $h = \tilde{h} \circ U^*$. As we said above, this process is equivalent to choosing at random $h \in R_\zeta$. The advantage of the procedure we have just suggested is that it does not need to compute a orthogonal basis of R_ζ.

(5) Thus, we have produced $y = (M, l, \zeta, h)$ (where l can be discarded), which we then input in the mapping $G_{(d)}$ above to obtain the initial pair (g, ζ).

Although the process that we have just described may seem difficult and involves many notations, it is actually very simple for a computer to perform these operations. In [BeLe] the first author, with A. Leykin, implemented this process to generate random initial pairs.

2.4. **Good starting pairs.** Note that for fixed $(g, \zeta) \in V$, the number $\mathcal{A}(g, \zeta) = \mathrm{E}_{f \in \mathbb{S}} \left(\mathcal{C}(f, g, \zeta)\right)$ provides a bound for the average number of projective Newton steps needed for computing an approximate zero of $f \in \mathbb{S}$.

A pair $(g, \zeta) \in V$ is a *good starting pair* if $\mathcal{A}(g, \zeta)$ is polynomial in n, N, d. Formally, we look for pairs $(g_{(d)}, \zeta_{(d)})$ such that $\mathcal{A}(g_{(d)}, \zeta_{(d)}) \leq p(n, N, d)$ for any list of degrees (d), where $p : \mathbb{R}^3 \to \mathbb{R}$ is some fixed polynomial.

Corollary 8 (Existence and generation of good starting pairs)**.** *Let $(g, \zeta_0) \in \mathcal{G}_{(d)}$ be chosen randomly. Then, with probability at least $1/2$, (g, ζ_0) is a good starting pair, and*

$$\mathcal{A}(g, \zeta) \leq \pi n N.$$

Proof. From Theorems 5 and 7 and Fubini's Theorem, we conclude that

$$\mathrm{E}_{(g,\zeta) \in \mathcal{G}_{(d)}} \left(\mathrm{E}_{f \in \mathbb{S}} \left(\mathcal{C}(f, g, \zeta)\right)\right) \leq \frac{\pi}{2} n N. \tag{2.6}$$

The corollary follows from Markov's inequality. □

2.5. **Average Las Vegas Algorithm.** Note that for fixed $f \in \mathbb{S}$, the number $\mathrm{E}_{(g,\zeta)\in\mathcal{G}_{(d)}}\left(\mathcal{C}(f,g,\zeta)\right)$ provides a bound for the average complexity of approximating a homotopy path to solve f, where $(g,\zeta) \in \mathcal{G}_{(d)}$ are chosen randomly. This leads to the notion of *Average Las Vegas* algorithm.

Recall that a probabilistic algorithm is called *Las Vegas* if, for a given input x, it randomly generates an element y in some set, and performs a deterministic algorithm with input (x,y), in such a way that:

- If an answer is given by the algorithm, it is a correct answer.
- For every x, the running time $t(x)$ of the algorithm is polynomial in the size of the input. Note that the running time for a given choice of y may depend both in x and y, so we should denote it by $t(x,y)$. The running time $t(x)$ on input x is then defined as the *average* of the running times for random choices of y, i.e. $t(x) = \mathrm{E}_y(t(x,y))$.

The algorithm introduced in this paper is an Average Las Vegas algorithm, namely the same properties are satisfied but the second one is relaxed to

$$\mathrm{E}_{\{x:size(x)\le K\}}(t(x)) \le p(K)$$

for some polynomial p. That is, the average running time of the algorithm is polynomial in the size of the input. Our algorithm is

Adaptive Homotopy Method with random initial pair (ahmr)

Input: $f \in \mathbb{S}$.

- Choose randomly $(g,\zeta_0) \in \mathcal{G}_{(d)}$.
- Approximate the curve $\Gamma(f,g,\zeta_0)$ using the homotopy algorithm of [Bel].

Output: An approximate zero of f, with associated zero the unique zero of f lying on $\Gamma(f,g,\zeta_0)$.

Corollary 9. *The algorithm described above is Average Las Vegas, with average number of Newton steps at most $Cd^{3/2}nN$ (C a constant) and average running time (number of arithmetic ops.) $O(d^{3/2}nN(N+n^3))$.*

Proof. From Theorems 5 and 7, we conclude that

$$\mathrm{E}_{f\in\mathbb{S}}\left(\mathrm{E}_{(g,\zeta)\in\mathcal{G}_{(d)}}\left(\mathcal{C}(f,g,\zeta)\right)\right) \le \frac{\pi}{2}nN. \tag{2.7}$$

The corollary follows from the complexity bounds of equations (2.3), (2.4). □

The reader may compare the estimate of Corollary 9 with the one of [BePa3], $O\tilde{(}n^7N^3)$.

Corollary 9 proves that ahmr yields a solution to Smale's 17th problem. As many other well–known algorithms in Numerical Analysis and Computer Science, our algorithm is probabilistic. Namely, it starts by making some random choices and then performs some operations on the input and the random choices. It is however uniform as demanded by Smale's problem. The question of finding a deterministic, uniform algorithm for Smale's problem remains open.

2.6. **How to randomly choose a root of a given polynomial system.** As mentioned in the introduction, if $f \in \mathbb{S}$ is such that all of its zeros are regular, then the algorithm AHMR with input f produces an approximation to one of the zeros of f, all of them being equally probable. This is an interesting feature of our algorithm, and the reader may realize that it allows us to power up the random choice method of Section 2.3: in that section we have seen that it is possible to choose a random zero of a random system of equations. Using AHMR we are then able to approximate a random zero of any fixed system in quadratic time in the dense input length! (asking only that this system has no singular zeros.) This feature is not known for any other average polynomial time algorithm that solves systems of multivariate polynomial equations. This property may be used, for instance, to answer questions about the average value of a function defined in the zeros of some given input system f. It may also be applied to compute by homotopy methods solutions of input systems f that satisfy certain constraints provided that the probability distribution of the solutions with respect to the constraints is known (cf. as, for instance, being real as in [BoPa]).

2.7. **Relation to other works.** To the knowledge of the authors, the unique (proven) uniform average polynomial time algorithm to find a zero of a system of polynomial equations is the one of [BePa2] (which assumes a small probability of failure) and its Average Las Vegas version [BePa3]. Those two papers describe a probabilistic solution to Smale's 17th problem, as does this.

There is a huge bibliography on the complexity analysis of methods for solving polynomial equations that we do not intend to summarize here. See [BePa1] and references therein. The approach here (as well as that of [BePa2, BePa3]) was originally inspired by the works [ShSm1, ShSm2, ShSm5, ShSm4].

Several articles, some of which have already appeared, have followed this manuscript. In [BeSh] the variance of Algorithm AHMR is proven to be polynomial in the input size, and some higher moments are shown to be finite. In [BuCu] some of the results in this paper are combined with smooth analysis techniques to show a deterministic way to choose the initial pair (g, ζ_0) in such a way that the total complexity is close to polynomial, $O(N^{\log \log N})$. For that purpose, [BuCu] combines the homotopy algorithm with that of [Ren]. In [BeLe] an implementation in Macaulay2 of algorithm AHMR is presented.

3. The underlying geometry

In this section we summarize some notions and results which are essential for the understanding of our proofs. The results in this section are already implicitely present in the literature, so we just recall them for the sake of completeness.

Some of the main advances in [ShSm1, ShSm2, ShSm5, ShSm4] are due to the smart explotation of a geometric structure related to the polynomial system solving: the solution variety. This variety can be defined in several manners, depending on the space where the two components (f, ζ) are. Here, we have already defined it as

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

One can also consider the affine solution variety allowing the systems to be in the vector space instead of the sphere $\mathbb{S}$, and the zeros to be in the sphere instead of

$\mathbb{P}(\mathbb{C}^{n+1})$,

$$\hat{V} = \{(f,\zeta) \in \mathcal{H}_{(d)} \setminus \{0\} \times \mathbb{S}(\mathbb{C}^{n+1}) : f(\zeta) = 0\} \subseteq \mathcal{H}_{(d)} \times \mathbb{S}(\mathbb{C}^{n+1}).$$

The following result easily follows from the arguments in [BCSS, p. 193, proof of Prop. 1].

Proposition 10. *$\hat{V}$ is a smooth real submanifold of $\mathcal{H}_{(d)} \times \mathbb{S}(\mathbb{C}^{n+1})$ of dimension $2N+3$. Its tangent space is*

$$T_{(f,\zeta)}\hat{V} = \{(\dot{f},\dot{\zeta}) \in \mathcal{H}_{(d)} \times \mathbb{C}^{n+1} : \dot{f}(\zeta) + Df(\zeta)\dot{\zeta} = 0,\ \mathbb{R}e\langle \zeta, \dot{\zeta}\rangle = 0\}.$$

Moreover, V is a smooth submanifold of $\mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1})$ of dimension $2N+1$, that is equal to the dimension of $\mathbb{S}$.

Proof. Consider the set

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

Following [BCSS, p. 193, proof of Prop. 1], we have that $\tilde{V}$ is a smooth, complex submanifold of $\mathcal{H}_{(d)} \times \mathbb{C}^{n+1}$, of complex dimension $N+2$, and its tangent space is

$$T_{(f,\zeta)}\tilde{V} = \{(\dot{f},\dot{\zeta}) \in \mathcal{H}_{(d)} \times \mathbb{C}^{n+1} : \dot{f}(\zeta) + Df(\zeta)\dot{\zeta} = 0\}.$$

Note that polynomial systems in $\mathcal{H}_{(d)}$ contain homogeneous polynomials and hence a point $\zeta \in \mathbb{C}^{n+1} \setminus \{0\}$ is a zero of $f \in \mathcal{H}_{(d)}$ if and only if $\lambda\zeta$ is a zero of f, for any $\lambda \in \mathbb{C} \setminus \{0\}$. Thus, $\tilde{V}$ is transversal to $\mathcal{H}_{(d)} \times \mathbb{S}(\mathbb{C}^{n+1})$ and the set

$$\tilde{V} \cap (\mathcal{H}_{(d)} \times \mathbb{S}(\mathbb{C}^{n+1})) = \hat{V}$$

is a smooth real submanifold of $\mathcal{H}_{(d)} \times \mathbb{S}(\mathbb{C}^{n+1})$ of dimension $2N+3$, as claimed. The formula given for its tagent The tangent space of $\hat{V}$ is then

$$T_{(f,\zeta)}\hat{V} = \{(\dot{f},\dot{\zeta}) \in \mathcal{H}_{(d)} \times \mathbb{C}^{n+1} : \dot{f}(\zeta) + Df(\zeta)\dot{\zeta} = 0, \mathbb{R}e\langle \zeta, \dot{\zeta}\rangle = 0\}.$$

The same argument shows that the set

$$\{(f,\zeta) \in \mathbb{S} \times \mathbb{S}(\mathbb{C}^{n+1}) : f(\zeta) = 0\} \subseteq \mathbb{S} \times \mathbb{S}(\mathbb{C}^{n+1})$$

is a smooth submanifold of $\mathbb{S} \times \mathbb{S}(\mathbb{C}^{n+1})$ of dimension $2N+2$. Finally, our solution variety V is the quotient of this last set by the free action $(f,\zeta) \mapsto (f,\lambda\zeta)$ defined for $\zeta \in \mathbb{C}$, $|\zeta| = 1$. Hence, V is a smooth submanifold of $\mathbb{S} \times \mathbb{P}(\mathbb{C}^{n+1})$ and the dimension of V is $2N+1$. □

An important subset of the solution variety is the set

$$\Sigma' = \{(f,\zeta) \in V : rank(Df(\zeta)) < n\} \subseteq V,$$

that is the set of pairs (f,ζ) such that ζ is a singular zero of f. We define the discriminant variety

$$\Sigma = \{f \in \mathbb{S} : (f,\zeta) \in \Sigma' \text{ for some } \zeta \in V(f)\},$$

that is the set of systems which have some singular zero.

Proposition 11. *The set Σ is an algebraic subvariety of $\mathbb{S}$ of (real) codimension 2. There exists $r \geq 1$ and a decomposition*

$$\Sigma = K_1 \cup \cdots \cup K_r,$$

where each K_j is a smooth embedded submanifold of $\mathbb{S}$ of dimension at most $2N-1$.

Proof. We use a standard argument in Algebraic Geometry. First, note that the set

$$\{f \in \mathbb{P}(\mathcal{H}_{(d)}) : f \text{ has some singular solution}\}$$

is a proper, complex algebraic subvariety of $\mathbb{P}(\mathcal{H}_{(d)})$ (see for example [BCSS, p. 198]). Here, by proper we mean that the set is strictly contained in $\mathbb{P}(\mathcal{H}_{(d)})$. Thus, the set

$$\hat{\Sigma} = \{f \in \mathcal{H}_{(d)} : f \text{ has some singular solution}\}$$

is a proper, complex, homogeneous (i.e. containing the lines thorugh the origin) algebraic subvariety of $\mathcal{H}_{(d)}$. Thus, the real codimension of $\hat{\Sigma}$ in $\mathcal{H}_{(d)}$ is 2 and $\Sigma = \hat{\Sigma} \cap \mathbb{S}$ is an algebraic subvariety of $\mathbb{S}$ of real codimension 2.

On the other hand, from [Har, Th. 5.3] we have that $\hat{\Sigma} = \hat{K}_1 \cup H$ where $\hat{K}_1$ is a smooth submanifold of $\mathcal{H}_{(d)}$ (consisting on the simple points of $\hat{\Sigma}$) and H is a complex homogeneous algebraic subvariety of $\mathcal{H}_{(d)}$ of complex codimension at least 2. Inductively, we can write

$$\hat{\Sigma} = \hat{K}_1 \cup \cdots \cup \hat{K}_r,$$

where each $\hat{K_r}$ is a smooth homogeneous submanifold of $\mathcal{H}_{(d)}$ and the real dimension of each $\hat{K}_j$ is at most $\dim(\mathcal{H}_{(d)}) - 2 = 2N$. The decomposition in the proposition follows, taking $K_j = \hat{K}_j \cap \mathbb{S}$. □

Recall that for $f, g \in \mathbb{S}$, $f \neq \pm g$, $L_{f,g}$ is the (shortest) portion of the great circle from g to f. We prove now that for almost every choice of $(f, g) \in \mathbb{S} \times \mathbb{S}$ the arc $L_{f,g}$ does not intersect Σ.

Proposition 12. *The set*

$$\mathcal{S} = \{(f, g) \in \mathbb{S} \times \mathbb{S} : L_{f,g} \cap \Sigma \neq \emptyset\}$$

has zero measure in $\mathbb{S} \times \mathbb{S}$. Moeover, for $(f, g) \notin \mathcal{S}$ and for $h \in L_{f,g}$ the zeros of h can be continued from (and are thus in one to one correspondence with) the zeros of g, following the arc $L_{f,g}$.

Proof. Note that

$$\mathcal{S} \subseteq (\Sigma \times \mathbb{S}) \cup Image(\Phi),$$

where Φ is the following C^∞ map,

$$\begin{array}{cccc} \Phi : & (\mathbb{S} \setminus \Sigma) \times \Sigma \times (0, 2\pi) & \to & (\mathbb{S} \setminus \Sigma) \times \mathbb{S} \\ & (g, h, \theta) & \mapsto & (g, f) \end{array}$$

where $f \in \mathbb{S}$ is the system in the great circle containing g, h, at distance θ from g (in the direction from g to h). It is not difficult to give a precise formula for h but we do not need it. From Proposition 11, $Image(\Phi)$ is a finite union of sets, each of which is the C^∞ image of a smooth manifold of dimension at most $2N - 1 + 2N + 1 + 1 = 4N + 1$, thus a null set in $\mathbb{S} \times \mathbb{S}$ which has dimension $4N + 2$. The first assertion of the lemma follows.

The second claim is an easy consequence of the Implicit Function Theorem applied to the projection $\pi_1 : V \to \mathbb{S}$, which is locally invertible if $(f, \zeta) \notin \Sigma'$. See for example [ShSm5, Paragraphs 1–4, Sec. 2] or [BePa3, Prop. 3.1]. □

4. An integral geometry formula

The following result allows us to rewrite the quantity of Theorem 5 in terms of the expected value of the condition number in $\mathbb{S}$.

Theorem 13.

$$\mathrm{E}_{f\in\mathbb{S}}\left(\mathrm{E}_{g\in\mathbb{S}}\left(\sum_{\zeta_0\in V(g)}\mathcal{C}(f,g,\zeta_0)\right)\right)=\frac{\pi}{2}\mathrm{E}_{f\in\mathbb{S}}\left(\sum_{\zeta\in V(f)}\mu_{\mathrm{norm}}(f,\zeta)^2\right).$$

Proof. Let $\mathcal{S}$ be the set of Proposition 12. Then, for every $(f,g)\in(\mathbb{S}\times\mathbb{S})\setminus\mathcal{S}$ and for every $h\in L_{f,g}$, the mapping

$$\begin{array}{ccc} V(g) & \longrightarrow & V(h) \\ \zeta & \mapsto & \eta(h,\zeta)=V(h)\cap\Gamma(f,g,\zeta) \end{array} \tag{4.1}$$

is a bijection. Here, $\Gamma(f,g,\zeta)$ is the path defined in Section 2.2. Thus, we can write $V(h)=\cup_{\zeta_0\in V(g)}\eta(\zeta)$. In particular,

$$\sum_{\zeta_0\in V(g)}\mathcal{C}(f,g,\zeta_0)=\sum_{\zeta_0\in V(g)}\int_{h\in L_{f,g}}\mu_{\mathrm{norm}}(h,\eta(h,\zeta_0))^2\,dL_{f,g}=$$

$$\int_{h\in L_{f,g}}\sum_{\zeta\in V(h)}\mu_{\mathrm{norm}}(h,\zeta)^2\,dL_{f,g}.$$

Using Fubini's Theorem, we then have

$$\mathrm{E}_{f\in\mathbb{S}}\left(\mathrm{E}_{g\in\mathbb{S}}\left(\sum_{\zeta_0\in V(g)}\mathcal{C}(f,g,\zeta_0)\right)\right)=\frac{1}{Vol(\mathbb{S})^2}\int_{(f,g)\in\mathbb{S}\times\mathbb{S}}\int_{h\in L_{f,g}}\phi(h)\,dL_{f,g}\,d(\mathbb{S}\times\mathbb{S}), \tag{4.2}$$

where $\phi(h)=\sum_{\zeta\in V(h)}\mu_{\mathrm{norm}}(h,\zeta)^2$. We claim that for any measurable non–negative function $\hat{\phi}:\mathbb{S}\to\mathbb{R}$ the following holds,

$$\frac{1}{Vol(\mathbb{S})^2}\int_{(f,g)\in\mathbb{S}\times\mathbb{S}}\int_{h\in L_{f,g}}\hat{\phi}\,dL_{f,g}\,d(\mathbb{S}\times\mathbb{S})=\frac{\pi}{2Vol(\mathbb{S})}\int_{f\in\mathbb{S}}\hat{\phi}\,d\mathbb{S}, \tag{4.3}$$

which together with (4.2) readily implies the theorem. To prove (4.3), note that the unitary group $U(N+1)$ acting on $\mathbb{S}$ defines a transitive left action. Seing f as a vector in $\mathbb{C}^{N+1}$ whose components are the coefficients of the monomials of the f_i's, the action will send

$$(U,f)\to\Delta U\Delta^{-1}f,$$

where Δ is a diagonal matrix containing $\binom{d_i}{\alpha_0\cdots\alpha_n}$ in the position corresponding to the monomial $X_0^{\alpha_0}\cdots X_n^{\alpha_n}$ of f_i.

From the uniqueness of invariant measures (see for example [SeKu, Cor. 7.5.1]), the existence of this transitive group of isometries in $\mathbb{S}$ implies that

$$\int_{(f,g)\in\mathbb{S}\times\mathbb{S}}\int_{h\in L_{f,g}}\hat{\phi}\,dL_{f,g}\,d(\mathbb{S}\times\mathbb{S})=\lambda\int_{f\in\mathbb{S}}\hat{\phi}\,d\mathbb{S}, \tag{4.4}$$

for some constant $\lambda\in\mathbb{R}$. Take $\hat{\phi}\equiv 1$ in (4.4) to get

$$\int_{f\in\mathbb{S}}\int_{g\in\mathbb{S}}d_R(f,g)\,d\mathbb{S}d\mathbb{S}=\int_{(f,g)\in\mathbb{S}\times\mathbb{S}}d_R(f,g)\,d(\mathbb{S}\times\mathbb{S})=\lambda Vol(\mathbb{S}).$$

From Lemma 14 below we conclude that $\lambda=Vol(\mathbb{S})\pi/2$, and (4.3) then follows.

□

Lemma 14. *Let $k \geq 1$, and let $\mathbb{R}^{k+1}$ be endowed with some inner product $\langle \cdot, \cdot \rangle_1$. Let the unit sphere $S_1 = \{x \in \mathbb{R}^{k+1} : \|x\|_1 = 1\}$ have the Riemannian structure inherited from $(\mathbb{R}^{k+1}, \langle \cdot, \cdot \rangle_1)$. Then, for any $x \in S_1$ we have*

$$\int_{y \in S_1} d_R(x,y)\, dS_1 = Vol(S_1)\frac{\pi}{2},$$

where d_R is the Riemannian distance in S_1.

Proof. As in [BCSS, pp. 225,226], we have that $d_R(x,y) = \arccos\langle x, y\rangle_1$ where arccos is chosen in $[0,\pi]$. Thus,

$$\int_{y \in S_1} d_R(x,y)\, dS_1 = \int_{y \in S_1} \arccos\langle x, y\rangle_1\, dS_1.$$

Now, $g(y) = (\arccos\langle x, y\rangle_1 - \pi/2)$ is an odd function for $g(y) = -g(-y)$. Thus,

$$\int_{y \in S_1} \left(\arccos\langle x, y\rangle_1 - \frac{\pi}{2}\right)\, dS_1 = 0,$$

which implies

$$\int_{y \in S_1} \arccos\langle x, y\rangle_1\, dS_1 = \int_{y \in S_1} \frac{\pi}{2}\, dS_1 = Vol(S_1)\frac{\pi}{2},$$

as wanted. □

5. Jacobian and the solution variety: the Main Lemma

Recall from Proposition 10 that the affine solution variety $\hat{V}$ is a (real) smooth manifold of dimension $2N+3$. We equip $\hat{V}$ with the Riemannian metric inherited from that of $\mathcal{H}_{(d)} \times \mathbb{S}(\mathbb{C}^{n+1})$.

With the notations above, the set $\hat{V}_{(1)} = \{(M,\zeta) : M \in \mathcal{H}_{(1)}, \zeta \in \mathbb{S}(\mathbb{C}^{n+1}),\ M\zeta = 0\}$ is the linear solution variety. The metric in $\hat{V}_{(1)}$ is then given by,

$$\langle (M_1,\zeta_1),(M_2,\zeta_2)\rangle = \langle M_1, M_2\rangle_F + \langle \zeta_1, \zeta_2\rangle,$$

where $\langle M_1, M_2\rangle_F = trace(M_1 M_2^*)$ is the usual Frobenius product and for $i = 1,2$ we have

$$(M_i,\zeta_i) \in T_{(M,\zeta)}\hat{V}_{(1)} = \{(\dot{M},\dot{\zeta}) \in \mathcal{H}_{(1)} \times \mathbb{C}^{n+1} : \dot{M}\zeta + M\dot{\zeta} = 0, \mathbb{R}e\langle \zeta, \dot{\zeta}\rangle = 0\}.$$

The formula for $T_{(M,\zeta)}\hat{V}_{(1)}$ follows from Proposition 10.

Recall that for a mapping $\Phi : \mathcal{M} \to \mathcal{M}'$ where $\mathcal{M}, \mathcal{M}'$ are Riemannian manifolds, the Normal Jacobian of Φ at $x \in \mathcal{M}$ is

$$NJ(\Phi)(x) = \det\left(A\mid_{(Ker(A)^\perp)}\right), \quad \text{where } A = D\Phi(x).$$

Equivalently,

$$NJ(\Phi)(x) = \frac{Vol(Parallelepiped(Av_1,\ldots,Av_n))}{Vol(Parallelepiped(v_1,\ldots,v_n))}$$

where $v_1,\ldots,v_r$ form a basis of $Ker(D\Phi(x))^\perp$.

Consider the mapping

$$\begin{array}{rccc} \hat{\Pi}: & \hat{V} & \longrightarrow & \hat{V}_{(1)} \\ & (f,\zeta) & \mapsto & (Df(\zeta),\zeta) \end{array}$$

The following technical lemma is a key result for the main outcome of this manuscript.

Main Lemma. *Let $d \geq 2$. Then, the Normal Jacobian of $\hat{\Pi}$ is constant and equal to $\mathcal{D}^n$.*

Proof. For any unitary matrix U of size $(n+1)\times(n+1)$ we have the commutative diagram

$$\begin{tikzcd}
\hat{V} \arrow[r, "\hat{\Pi}"] \arrow[d] & \hat{V}_{(1)} \arrow[d] & (f,\zeta) \arrow[r, mapsto] \arrow[d] & (Df(\zeta),\zeta) \arrow[d] \\
\hat{V} \arrow[r, "\hat{\Pi}"] & \hat{V}_{(1)} & (f\circ U^*, U\zeta) \arrow[r, mapsto] & (Df(\zeta)U^*, U\zeta)
\end{tikzcd}$$

Thus, from [BCSS, Lemma 4, p. 244] we have $NJ(\hat{\Pi})(f,\zeta) = NJ(\hat{\Pi})(f\circ U^*, U\zeta)$. By choosing U such that $U\zeta = e_0 = (1,\ldots,0)^T$ it then suffices to check the result for $\zeta = e_0$. Note that

$$D\hat{\Pi}(f,e_0)(\dot{f},\dot{\zeta}) = (D\dot{f}(e_0) + D^{(2)}f(e_0)(\dot{\zeta},\cdot),\dot{\zeta}).$$

Here, $D^{(2)}f(e_0)(\dot{\zeta},\cdot)$ is seen as an element of $T\mathcal{H}_{(1)} = \mathcal{H}_{(1)}$, the only one satisfying

$$D^{(2)}f(e_0)(\dot{\zeta},\cdot)x = D^{(2)}f(e_0)(\dot{\zeta},x), \quad \forall\, x \in \mathbb{C}^{n+1}.$$

It follows that $Ker(D\hat{\Pi}(f,e_0)) = \{(\dot{f},0) \in T_{(f,e_0)}\hat{V} : D\dot{f}(e_0) = 0\} = R_{e_0}$, namely the set of pairs $(\dot{f},0)$ such that every monomial of the polynomials of f is at least quadratic in $X_1,\ldots,X_n$. We claim that

$$Ker(D\hat{\Pi}(f,e_0))^{\perp} = P_1 \oplus P_2 \oplus \{(0,\mathbf{i}\,e_0)\}, \tag{5.1}$$

where

$$P_1 = \{((\dot{f}_1,\ldots,\dot{f}_n),0) : \dot{f}_i(z) = \sum_{j=1}^{n} a_{ij} z_0^{d_i-1} z_j, a_{ij} \in \mathbb{C}\},$$

$$P_2 = \{((\dot{f}_1,\ldots,\dot{f}_n),\dot{\zeta}) : \dot{f}_i(z) = a_i z_0^{d_i}, a_i = -(Df(e_0)\dot{\zeta})_i\}.$$

For (5.1), note that every element in $P_1 \oplus P_2 \oplus \{(0,\mathbf{i}\,e_0)\}$ is contained in $T_{(f,e_0)}\hat{V}$ and is orthogonal to $Ker(D\hat{\Pi}(f,e_0))$ and then use the following dimension argument:

$$\dim_{\mathbb{R}}(P_1) + \dim_{\mathbb{R}}(P_2) + \dim_{\mathbb{R}}(\{(0,\mathbf{i}\,e_0)\}) = 2n^2 + 2n + 1.$$

$$\dim_{\mathbb{R}}\left(Ker(D\hat{\Pi}(f,e_0))^{\perp}\right) = 2N + 3 - \dim_{\mathbb{R}}\left(Ker(D\hat{\Pi}(f,e_0))\right) =$$

$$2N + 3 - (2N + 2 - 2n^2 - 2n) = 2n^2 + 2n + 1.$$

A complex orthogonal basis of the complex subspace P_1 is given by the set $\{Diag(z_0^{d_i-1})\delta_{ij}z_j : 1 \leq i,j \leq n\}$ where δ_{ij} is a matrix identically equal to zero with a 1 in the position (i,j). Note that the vector of the basis corresponding to δ_{ij} has a Bombiery-Weyl norm equal to $d_i^{-1/2}$. Also,

$$D\hat{\Pi}(f,e_0)(Diag(z_0^{d_i-1})\delta_{ij}z_j,0) = ((0\ \delta_{ij}),0). \tag{5.2}$$

On the other hand, for $(\dot{f},\dot{\zeta}) \in P_2$ we have

$$D\dot{f}(e_0)e_0 = Diag(-d_i(Df(e_0)\dot{\zeta})_i) = -Diag(d_i)Df(e_0)\dot{\zeta}.$$

Using that $D^{(2)}f(e_0)(\dot{\zeta},e_0) = Diag(d_i - 1)Df(e_0)\dot{\zeta}$ (see Lemma 15 below), for $(\dot{f},\dot{\zeta}) \in P_2$ we conclude

$$(D\dot{f}(e_0) + D^{(2)}f(e_0)(\dot{\zeta},\cdot))e_0 = D\dot{f}(e_0)e_0 + D^{(2)}f(e_0)(\dot{\zeta},e_0) =$$

$$-Diag(d_i)Df(e_0)\dot\zeta + Diag(d_i-1)Df(e_0)\dot\zeta = -Df(e_0)\dot\zeta.$$

Thus, for $(\dot f, \dot\zeta) \in P_2$ we have

$$D\hat\Pi(f,e_0)(\dot f,\dot\zeta) = ((-Df(e_0)\dot\zeta \ \mid \ M),\dot\zeta), \tag{5.3}$$

where M is some $n\times n$ matrix that we do not need to compute explicitely. Similarly,

$$D\hat\Pi(f,e_0)(0,\mathbf{i}\, e_0) = ((0 \ \mid \ M),\mathbf{i}\, e_0), \tag{5.4}$$

where M is again some $n\times n$ matrix. The volume of a parallelepiped in affine space does not vary if a multiple of the vector defining one edge is added to another vector. Hence, to compute the volume of the parallelepiped generated by the vectors obtained in equations (5.2), (5.3), (5.4), we can omit the matrices M in (5.3), (5.4). We conclude that

$$NJ_{(f,e_0)}\hat\Pi = \frac{Vol_{T_{\hat\Pi(f,e_0)}\hat V_{(1)}}(((-Df(e_0)(e_j) \ \mid \ 0),e_j) : j=0\ldots n)}{Vol_{P_2}((-Diag(z_0^{d_i})Df(e_0)(e_j),e_j) : j=0\ldots n)}\times$$

$$\frac{Vol_{T_{\hat\Pi(f,e_0)}\hat V_{(1)}}((\delta_{ij},0) : i,j=1\ldots n)}{Vol_{P_1}((Diag(z_0^{d_i-1})\delta_{ij}z_j,0) : i,j=1\ldots n)}. \tag{5.5}$$

Note that the first of these two quotients is equal to 1, for

$$\langle((-Df(e_0)(e_j) \ \mid \ 0),e_j),((-Df(e_0)(e_k) \ \mid \ 0),e_k)\rangle =$$
$$\langle(-Diag(z_0^{d_i})Df(e_0)(e_j),e_j),(-Diag(z_0^{d_i})Df(e_0)(e_k),e_k)\rangle, \ \ \forall\, j,k=1\ldots n.$$

Moreover, the second factor in (5.5) is easy to compute, as the vectors appearing are orthogonal: The numerator is equal to 1 and the denominator is equal to

$$\prod_{i,j=1}^n \|Diag(z_0^{d_i-1})\delta_{ij}z_j\|^2 = \prod_{i,j=1}^n d_i^{-1} = \mathcal{D}^{-n}.$$

The lemma follows. Note that in the last formula we have to consider $\|\cdot\|^2$ for the computation of the volume, because we are dealing with complex vectors. □

Lemma 15. *Let $f\in\mathcal{H}_{(d)}$ be such that $f(e_0)=0$ where $e_0=(1,0,\ldots,0)\in\mathbb{C}^{n+1}$. Then, for any $\dot\zeta\in\mathbb{C}^{n+1}$ we have:*

$$D^{(2)}f(e_0)(\dot\zeta,e_0) = Diag(d_i-1)Df(e_0)\dot\zeta.$$

Proof. Let $f=(f_1,\ldots,f_n)$. Fix $i\in\{1,\ldots,n\}$ and write

$$f_i(z_0,\ldots,z_n) = z_0^{d_i-1}\sum_{j=1}^n a_jz_j + z_0^{d_i-2}\sum_{j\le k} b_{jk}z_jz_k + z_0^{d_i-3}\left(\cdots\right).$$

A straight forward computation shows that the gradient of f_i at e_0 is $\nabla f_i(e_0) = (0,a_1,\ldots,a_n)$. Thus,

$$Df_i(e_0)\dot\zeta = (0,a_1,\ldots,a_n)\cdot\dot\zeta.$$

On the other hand, another straight forward computation yields

$$H = \mathrm{Hess} f_i(e_0) = \begin{pmatrix} 0 & (d_i-1)a_1 & \cdots & (d_i-1)a_n \\ (d_i-1)a_1 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ (d_i-1)a_n & * & \cdots & * \end{pmatrix}$$

Thus,

$$D^{(2)}f_i(e_0)(\dot\zeta, e_0) = e_0^T H\dot\zeta = (d_i-1)(0,a_1,\dots,a_n)\cdot\dot\zeta = (d_i-1)Df_i(e_0)\dot\zeta,$$

which holds for any $i\in\{1,\dots,n\}$. The lemma follows. □

6. Integration formulas and the solution variety

In this section, we describe a reduction method that allows us to integrate functions defined on $V_{(d)}$ by analyzing the behavior of these functions on the linear solution variety $V_{(1)}$. This method is similar to the one described on [ShSm2] or [BCSS, Chapter 12], but it has a shorter and more direct proof using the Main Lemma.

We will use the following result which closely follows [ShSm2, Sec. 2]. There is a subtle difference with that paper: here, the affine solution variety $\hat V$ is a subset of $\mathcal{H}_{(d)}\times\mathbb{S}$ while in [ShSm2, Sec. 2] the solution variety is a subset of $\mathbb{P}(\mathcal{H}_{(d)})\times\mathbb{P}(\mathbb{C}^{n+1})$. Due to this difference, we include a proof of the result.

Theorem 16. *Let $\hat\Theta:\hat V\to[0,\infty)$ be a measurable mapping. Then,*

$$\int_{f\in\mathcal{H}_{(d)}}\int_{\zeta\in V_{\mathbb{S}}(f)}\hat\Theta(f,\zeta)\,dV_{\mathbb{S}}(f)\,d\mathcal{H}_{(d)} = \int_{(f,\zeta)\in\hat V_{(d)}}\frac{\hat\Theta(f,\zeta)}{\det(I_n+Df(\zeta)^\dagger(Df(\zeta)^\dagger)^*)}\,d\hat V_{(d)}.$$

Proof. We denote by $\hat\pi_1:\hat V\longrightarrow\mathcal{H}_{(d)}$ the projection on the first coordinate. The Coarea formula [BCSS, pg. 241] then yields

(6.1)
$$\int_{f\in\mathcal{H}_{(d)}}\int_{\zeta\in V_{\mathbb{S}}(f)}\hat\Theta(f,\zeta)\,dV_{\mathbb{S}}(f)\,d\mathcal{H}_{(d)} = \int_{(f,\zeta)\in\hat V_{(d)}}\hat\Theta(f,\zeta)NJ(\hat\pi_1)(f,\zeta)\,d\hat V_{(d)}.$$

where $NJ(\hat\pi_1)(f,\zeta)$ is the Normal Jacobian of $\hat\pi_1$ at (f,ζ), that is

$$NJ(\hat\pi_1)(f,\zeta) = |\det(A\mid_{Ker(A)^\perp})|,\quad\text{where}\quad A=D\hat\pi_1(f,\zeta).$$

Now, note that

$$\begin{array}{rccc} A: & T_{(f,\zeta)}\hat V & \to & \mathcal{H}_{(d)}\\ & (\dot f,\dot\zeta) & \mapsto & \dot f\end{array}$$

and from Proposition 10 we have that

$$Ker(A)=\{(0,\dot\zeta): Df(\zeta)\dot\zeta=0,\ \mathbb{R}e\langle\zeta,\dot\zeta\rangle=0\}.$$

The kernel of $Df(\zeta)$ is the complex line defined by ζ. Thus, $Ker(A)^\perp=\zeta^\perp$ is the complex orthogonal complement of ζ. Then,

$$\begin{array}{rccc}\left(A\mid_{Ker(A)^\perp}\right)^{-1}: & \mathcal{H}_{(d)} & \to & \{(\dot f,\dot\zeta)\in\mathcal{H}_{(d)}\times\zeta^\perp:\dot f(\zeta)+Df(\zeta)\dot\zeta=0\}\\ & \dot f & \mapsto & (\dot f,-Df(\zeta)^\dagger\dot f(\zeta))\end{array}$$

We are in the conditions of [ShSm2, Lemma 1, p. 274 and Remark, p. 275] which yields

$$\left|\det\left(\left(A\mid_{Ker(A)^\perp}\right)^{-1}\right)\right| = \det(I_{N+1}+Q^*Q)^{-1},$$

where Q is the linear mapping $Q(\dot f)=-Df(\zeta)^\dagger\dot f(\zeta)$. If we first fix $\zeta=e_0=(1,0,\dots,0)\in\mathbb{C}^{n+1}$ then following [ShSm2, p. 276] we have

$$QQ^*\dot\zeta = Df(e_0)^\dagger(Df(e_0)^\dagger)^*\dot\zeta,\quad\text{for }\zeta\in e_0^\perp\equiv\mathbb{C}^n.$$

Thus, in that case we have

$$\det(I_{N+1}+Q^*Q)=\det(I_n+Q^*Q)=\det(I_n+Df(e_0)^\dagger(Df(e_0)^\dagger)^*),$$

and

$$NJ(\hat\pi_1)(f,e_0)=|\det(A\mid_{Ker(A)^\perp})|=$$

$$\left|\det\left(\left(A\mid_{Ker(A)^\perp}\right)^{-1}\right)\right|^{-1}=\det(I_{N+1}+Q^*Q)^{-1}=\det(I_n+Df(e_0)^\dagger(Df(e_0)^\dagger)^*)^{-1}.$$

By unitary invariance as in [ShSm2, p. 276] we then have

$$NJ(\hat\pi_1)(f,\zeta)=\det(I_n+Df(\zeta)^\dagger(Df(\zeta)^\dagger)^*)^{-1},\quad\forall\,(f,\zeta)\in\hat V.$$

The proposition follows from this last formula and (6.1) □

Corollary 17. *Let $\hat\Theta:\hat V_{(d)}\longrightarrow[0,\infty)$ be a measurable mapping. Then,*

$$\int_{f\in\mathcal{H}_{(d)}}\int_{\zeta\in V_{\mathbb{S}}(f)}\hat\Theta(f,\zeta)\,dV_{\mathbb{S}}(f)\,d\mathcal{H}_{(d)}=$$

$$\mathcal{D}\int_{M\in\mathcal{H}_{(1)}}\int_{\zeta\in V_{\mathbb{S}}(M)}\int_{h\in R_\zeta}\hat\Theta(h+\varphi(M,\zeta),\zeta)\,dR_\zeta\,dV_{\mathbb{S}}(M)\,d\mathcal{H}_{(1)}.$$

where $\varphi(M,\zeta)$ is defined by equation (2.5).

Proof. From Theorem 16,

$$\int_{f\in\mathcal{H}_{(d)}}\int_{\zeta\in V_{\mathbb{S}}(f)}\hat\Theta(f,\zeta)\,dV_{\mathbb{S}}(f)\,d\mathcal{H}_{(d)}=\int_{(f,\zeta)\in\hat V_{(d)}}\frac{\hat\Theta(f,\zeta)}{\det(I_n+Df(\zeta)^\dagger(Df(\zeta)^\dagger)^*)}\,d\hat V_{(d)}.$$

From the Main Lemma and the Coarea Formula, this last equals

$$\mathcal{D}^{-n}\int_{(M,\zeta)\in\hat V_{(1)}}\frac{1}{\det(I+M^\dagger(M^\dagger)^*)}\int_{(f,\zeta)\in\hat\Pi^{-1}(M,\zeta)}\hat\Theta(f,\zeta)\,d\hat\Pi^{-1}(M,\zeta)\,d\hat V_{(1)}.$$

Again from Theorem 16, this time applied to $V_{(1)}$, the last formula equals

$$\mathcal{D}^{-n}\int_{M\in\mathcal{H}_{(1)}}\int_{\zeta\in V_{\mathbb{S}}(M)}\int_{(f,\zeta)\in\hat\Pi^{-1}(M,\zeta)}\hat\Theta(f,\zeta)\,d\hat\Pi^{-1}(M,\zeta)\,dV_{\mathbb{S}}(M)\,d\mathcal{H}_{(1)}.$$

Finally, the change of variables formula applied to the mapping

$$\begin{array}{ccc}\mathcal{H}_{(1)} & \longrightarrow & \mathcal{H}_{(1)}\\ M & \mapsto & Diag(d_i^{-1/2})M\end{array}$$

whose Jacobian is $\mathcal{D}^{-(n+1)}$ yields the corollary. Note that identifying

$$\hat\Pi^{-1}(Diag(d_i^{1/2})M,\zeta)\equiv\varphi(M,\zeta)+R_\zeta,$$

we can substitute the inner integral

$$\int_{(f,\zeta)\in\hat\Pi^{-1}(Diag(d_i^{1/2})M,\zeta)}\hat\Theta(f,\zeta)\,d\hat\Pi^{-1}(Diag(d_i^{1/2})M)=\int_{h\in R_\zeta}\hat\Theta(h+\varphi(M,\zeta),\zeta)\,dR_\zeta.$$

□

Corollary 18. *Let $\Theta : V_{(d)} \longrightarrow [0,\infty)$ be a measurable mapping. Let $\phi : [0,1] \to \mathbb{R}$ be a measurable non–negative function and consider the mapping*

$$\begin{array}{cccc} \hat{\Theta} : & \{(f,\zeta) \in \hat{V} : \|f\| \le 1\} & \to & [0,\infty) \\ & \hat{\Theta}(f,\zeta) & \mapsto & \phi(||f||)\Theta(f/||f||,\zeta). \end{array}$$

Then,

$$\int_{f\in\mathbb{S}} \sum_{\zeta\in V(f)} \Theta(f,\zeta)\, d\mathbb{S} = \frac{\mathcal{D}}{2\pi\int_0^1 \phi(t)t^{2N+1}\,dt}\times$$

$$\int_{M\in B(\mathcal{H}_{(1)})} (1-\|M\|^2)^p \int_{\zeta\in V_{\mathbb{S}}(M)} I(M,\zeta)\, dV_{\mathbb{S}}(M)\, d\mathcal{H}_{(1)},$$

where $p = N - n^2 - n + 1$ is the complex dimension of the vector space R_ζ and

$$I(M,\zeta) = \int_{h\in B(R_\zeta)} \hat{\Theta}(\sqrt{1-\|M\|_F^2}\, h + \varphi(M,\zeta),\zeta)\, dR_\zeta.$$

Proof. Using polar coordinates,

$$\frac{1}{2\pi}\int_{f\in B(\mathcal{H}_{(d)})}\int_{\zeta\in V_{\mathbb{S}}(f)} \hat{\Theta}(f,\zeta)\, dV_{\mathbb{S}}(f)\, d\mathcal{H}_{(d)} =$$

$$\left(\int_0^1 \phi(t)t^{2N+1}\,dt\right)\times\left(\int_{f\in\mathbb{S}}\sum_{\zeta\in V(f)}\Theta(f,\zeta)\, d\mathbb{S}\right).$$

On the other hand, from Corollary 17,

$$\int_{f\in B(\mathcal{H}_{(d)})}\int_{\zeta\in V_{\mathbb{S}}(f)} \hat{\Theta}(f,\zeta)\, dV_{\mathbb{S}}(f)\, d\mathcal{H}_{(d)} =$$

$$\mathcal{D}\int_{M\in B(\mathcal{H}_{(1)})}\int_{\zeta\in V_{\mathbb{S}}(M)}\int_{h\in R_\zeta, ||h||\le\sqrt{1-||M||_F^2}} \hat{\Theta}(h+\varphi(M,\zeta),\zeta)\, dR_\zeta\, dV_{\mathbb{S}}(M)\, d\mathcal{H}_{(1)}.$$

Using the change of variables $h \mapsto h(1-\|M\|^2)^{-1/2}$, this last equals

$$\mathcal{D}\int_{M\in B(\mathcal{H}_{(1)})}(1-\|M\|_F^2)^p\int_{\zeta\in V_{\mathbb{S}}(M)} I(M,\zeta)\, dV_{\mathbb{S}}(M)\, d\mathcal{H}_{(1)},$$

and the corollary follows. □

7. The moments of the condition number, linear case

In this section we compute these moments in the linear case, namely assuming that all the degrees are equal to 1. Thus, we consider $\mathcal{H}_{(1)}$ the vector space of $n\times(n+1)$ matrices and $\mathbb{S}_{(1)}$ the unit sphere (for the Frobenius norm) in $\mathcal{H}_{(1)}$.

There are several estimates for the probability distribution of the condition number of $n\times(n+1)$ matrices. Note that the case with $n=1$ is trivial as $\mu_{\text{norm}}\equiv 1$ is constant. The sharpest published bounds for $n\ge 2$ are those of [ChDo] and [EdSu], which deal with general $n\times m$ matrices. In previous versions of this paper, we used the estimates of [ChDo] which (after normalization) produce the bound

$$\mathrm{E}_{M\in\mathbb{S}_{(1)}}(||M^\dagger||^\alpha) \le \frac{4}{4-\alpha}(cn^6)^{\alpha/4}, \quad 0<\alpha<4,$$

where $c \le 16$. But this bound is not optimal. During the preparation of the revised version of this paper, it was pointed out to us by Alan Edelman that the $n\times(n+1)$

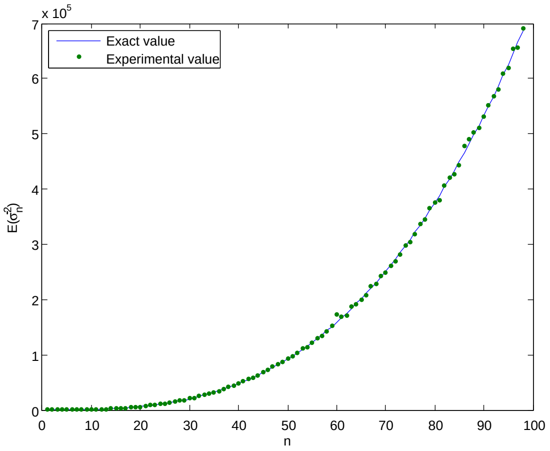

FIGURE 1. Comparison of exact and experimental values of $\|M^\dagger\|^2 = \sigma_n^{-2}(M)$. The experimental value has been obtained with Matlab, computing the average from 30000 random matrices in $\mathbb{S}_{(1)}$ for each $n = 1, \ldots, 98$.

case might admit an exact computation of the moments, which is what we do in this section. We will prove the following result .

Theorem 19. *Let $0 < \alpha < 4$ be a real number. Then,*

$$\mathrm{E}_{\mathbb{S}_{(1)}}(\|M^\dagger\|^\alpha) = \frac{\Gamma(n^2+n)}{\Gamma(n^2+n-\alpha/2)} \sum_{k=0}^{n-1} \frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)}$$

In particular,

$$\mathrm{E}_{\mathbb{S}_{(1)}}(\|M^\dagger\|^2) = (n^2+n-1)\left(n\left(1+\frac{1}{n}\right)^{n+1} - 2n - 1\right) \leq n(n^2+n-1).$$

In Figure 1 we compare experimental and theoretical values of $\mathrm{E}_{\mathbb{S}_{(1)}}(\|M^\dagger\|^2)$.

We start by computing the moments of $\|M^\dagger\|$ for a gaussian matrix M.

Theorem 20. *Let $M \in \mathcal{H}_{(1)}$ be chosen randomly in $\mathcal{H}_{(1)}$ with the Complex Gaussian distribution (that is, each element of M is chosen as $m_{ij} = a_{ij} + \sqrt{-1}b_{ij}$ where a_{ij} and b_{ij} are real Gaussian with mean 0 and variance 1). Then,*

$$\mathrm{E}_{\mathcal{H}_{(1)}}(\|M^\dagger\|^\alpha) = 2^{-\alpha/2} \sum_{k=0}^{n-1} \frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)}.$$

In particular,

$$\mathrm{E}_{\mathcal{H}_{(1)}}(\|M^\dagger\|^2) = \frac{n}{2}\left(1+\frac{1}{n}\right)^{n+1} - n - \frac{1}{2}.$$

Proof. For a $n \times (n+1)$ matrix M, let $\sigma_1(M), \dots, \sigma_n(M)$ be the singular values of M in decreasing order. Note that $\|M^\dagger\| = \sigma_n^{-1}$. According to [Ede, Formula (5.5)], the probability density function of $\sigma_n^{-2}(M)$ is

$$\Xi(\lambda) = \frac{\tilde{K}}{(n-1)!}\lambda e^{-\lambda n/2}\int_{(0,\infty)^{n-1}} F(\lambda, x_1, \dots, x_{n-1})\, dx_1 \cdots dx_{n-1},$$

where

$$\tilde{K}^{-1} = 2^{n^2+n}\prod_{i=1}^{n}\Gamma(n-i+2)\Gamma(n-i+1),$$

$$F(\lambda, x_1, \dots, x_{n-1}) = x_1^2 \cdots x_{n-1}^2 e^{-x_1/2} \cdots e^{-x_{n-1}/2}\prod_{i=1}^{n-1}(x_i+\lambda)\Delta,$$

$$\Delta = \prod_{1 \le i < j \le n-1}(x_i - x_j)^2.$$

By expanding the term $\prod_{i=1}^{n-1}(x_i+\lambda)$, we have that

$$\Xi(\lambda) = \frac{\tilde{K}}{(n-1)!}\lambda e^{-\lambda n/2}P(\lambda),$$

where $P(\lambda) = \sum_{k=0}^{n-1} C_{n-1-k}\lambda^{n-1-k}$ is a polynomial of degree $n-1$, whose λ^{n-1-k} coefficient is

$$C_{n-1-k} = \sum_{j_1,\dots,j_k}\int_{(0,\infty)^{n-1}} x_{j_1}\cdots x_{j_k}x_1^2\cdots x_{n-1}^2\Delta e^{-x_1/2}\cdots e^{-x_{n-1}/2}\, dx_1 \cdots dx_{n-1},$$

where the sum is carried out over every choice of different k numbers $1 \le j_1, \cdots, j_k \le n-1$. By symmetry, we have

$$C_{n-1-k} = \binom{n-1}{k}\int_{(0,\infty)^{n-1}} x_1\cdots x_k x_1^2\cdots x_{n-1}^2\Delta e^{-x_1/2}\cdots e^{-x_{n-1}/2}\, dx_1 \cdots dx_{n-1}.$$

The change of variables $x_i = 2y_i$ gives

$$C_{n-1-k} = \binom{n-1}{k}2^{n^2+k-1}\int_{(0,\infty)^{n-1}} y_1\cdots y_k y_1^2\cdots y_{n-1}^2\hat{\Delta} e^{-y_1}\cdots e^{-y_{n-1}}\, dy_1 \cdots dy_{n-1},$$

where $\hat{\Delta} = \prod_{1 \le i < j \le n-1}(y_i - y_j)^2$. This last integral is a particular case of one of the forms of Selberg's Integral, and its value is known, see [AAR, Cor. 8.2.2]:

$$\prod_{j=1}^{k}(n+2-j)\prod_{j=1}^{n-1}\frac{\Gamma(j+2)\Gamma(j+1)}{\Gamma(2)} = k!\binom{n+1}{k}\prod_{i=1}^{n}\Gamma(n-i+2)\Gamma(n-i+1).$$

Thus, we have

$$\frac{\tilde{K}}{(n-1)!}C_{n-1-k} = \frac{2^k\binom{n+1}{k}}{2^{n+1}\Gamma(n-k)}.$$

We conclude that

$$\Xi(\lambda) = \sum_{k=0}^{n-1}\frac{2^k\binom{n+1}{k}}{2^{n+1}\Gamma(n-k)}\lambda^{n-k}e^{-\lambda n/2}, \tag{7.1}$$

The expected value of $\sigma_n^{-\alpha}(M)$ ($\alpha \in (-\infty, 4)$) is then

$$\mathrm{E}_{\mathcal{H}_{(1)}}(\sigma_n^\alpha) = \int_0^\infty \lambda^{-\alpha/2}\Xi(\lambda)\, d\lambda = \sum_{k=0}^{n-1}\frac{2^k\binom{n+1}{k}}{2^{n+1}\Gamma(n-k)}\int_0^\infty \lambda^{n-k-\alpha/2}e^{-\lambda n/2}\, d\lambda.$$

The integral inside the sum is equal to $2^{n-k+1-\alpha/2}\Gamma(n-k+1-\alpha/2)n^{-(n-k+1-\alpha/2)}$. Thus, we get

$$\mathrm{E}_{\mathcal{H}_{(1)}}(\sigma_n^{-\alpha}) = 2^{-\alpha/2}\sum_{k=0}^{n-1}\frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)},$$

as wanted. In the particular case that $\alpha=2$ we can simplify this formula as follows.

$$\mathrm{E}_{\mathcal{H}_{(1)}}(\sigma_n^{-2}) = \frac{1}{2}\sum_{k=0}^{n-1}\frac{\binom{n+1}{k}}{n^{n-k}} = \frac{1}{2}\left(-2n-1+n\sum_{k=0}^{n+1}\frac{\binom{n+1}{k}}{n^{n-k+1}}\right) = \frac{n}{2}\left(1+\frac{1}{n}\right)^{n+1} - n - \frac{1}{2}.$$

□

7.1. **Proof of Theorem 19.** Integration in polar coordinates now gives the following:

$$2^{-\alpha/2}\sum_{k=0}^{n-1}\frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)} = \mathrm{E}_{\mathcal{H}_{(1)}}(\sigma_n^{-\alpha}) =$$

$$\frac{1}{2^{n^2+n}\pi^{n^2+n}}\int_{M\in\mathcal{H}_{(1)}}\sigma_n(M)^{-\alpha}e^{-\|M\|_F^2/2}\,dM =$$

$$\frac{1}{2^{n^2+n}\pi^{n^2+n}}\int_0^\infty e^{-r^2/2}\int_{M\in\mathcal{H}_{(1)},\|M\|=r}\sigma_n(M)^{-\alpha}\,dM\,dr =$$

$$\frac{1}{2^{n^2+n}\pi^{n^2+n}}\int_0^\infty r^{2n^2+2n-1-\alpha}e^{-r^2/2}\,dr\int_{M\in\mathbb{S}_{(1)}}\sigma_n(M)^{-\alpha}\,dM =$$

$$\frac{2^{-1-\alpha/2}\Gamma(n^2+n-\alpha/2)}{\pi^{n^2+n}}\int_{M\in\mathbb{S}_{(1)}}\sigma_n(M)^{-\alpha}\,dM.$$

Hence,

$$\int_{M\in\mathbb{S}_{(1)}}\sigma_n(M)^{-\alpha}\,dM = \frac{2\pi^{n^2+n}}{\Gamma(n^2+n-\alpha/2)}\sum_{k=0}^{n-1}\frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)},$$

and we conclude that

$$\mathrm{E}_{\mathbb{S}_{(1)}}(\|M^\dagger\|^\alpha) = \mathrm{E}_{\mathbb{S}_{(1)}}(\sigma_n^{-\alpha}) = \frac{1}{Vol(\mathbb{S}_{(1)})}\int_{M\in\mathbb{S}_{(1)}}\sigma_n^{-\alpha}(M)\,dM =$$

$$\frac{\Gamma(n^2+n)}{2\pi^{n^2+n}}\int_{M\in\mathbb{S}_{(1)}}\sigma_n^{-\alpha}(M)\,dM =$$

$$\frac{\Gamma(n^2+n)}{\Gamma(n^2+n-\alpha/2)}\sum_{k=0}^{n-1}\frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)},$$

as wanted. The particular case $\alpha=2$ admits, as in Theorem 20, a shorter formula.

8. Proof of Theorem 5

We will use Corollary 18. Let $-\infty < \alpha < 4$ and let $\Theta(f,\zeta) = \mu_{\text{norm}}(f,\zeta)^\alpha$, $\phi(t) = t^{-\alpha}$.

Lemma 21. *Let $(M,\zeta) \in \hat{V}_{(1)}$, and let $f = \sqrt{1-\|M\|^2}\ h + \varphi(M,\zeta)$. In the notations of Corollary 18,*

$$\hat{\Theta}(f,\zeta) = ||M^\dagger||^\alpha.$$

Proof. Note that $h \in R_\zeta$ and hence $Dh(\zeta) = 0$ as a matrix. Thus, we have

$$Df(\zeta) = D(\varphi(M,\zeta))(\zeta).$$

Now, for $\dot{\zeta} \in \mathbb{C}^{n+1}$, using that $(M,\zeta) \in \hat{V}_{(1)}$ and hence $M\zeta = 0$ we have

$$D(\varphi(M,\zeta))(\zeta)(\dot{\zeta}) = \frac{d}{dt}\mid_{t=0}\left(\varphi(M,\zeta)(\zeta + t\dot{\zeta})\right) \underset{(2.5)}{=} Diag(d_i^{1/2})M\dot{\zeta}.$$

We conclude that

$$Df(\zeta) = Diag(d_i^{1/2})M.$$

Thus,

$$\hat{\Theta}(f,\zeta) = \frac{1}{\|f\|^\alpha}\mu_{\text{norm}}\left(\frac{f}{\|f\|},\zeta\right)^\alpha \underset{(2.1)}{=}$$

$$\frac{1}{\|f\|^\alpha}\left\|\left(Diag(d_i^{-1/2})D\left(\frac{f}{\|f\|}\right)(\zeta)\right)^\dagger\right\|^\alpha = \|M^\dagger\|^\alpha,$$

as wanted. □

Proposition 22.

$$\mathrm{E}_{f\in\mathbb{S}}\left(\sum_{\zeta\in V(f)}\mu_{\text{norm}}(f,\zeta)^\alpha\right) = \frac{\mathcal{D}\Gamma(N+1)\Gamma(n^2+n-\alpha/2)}{\Gamma(N+1-\alpha/2)\Gamma(n^2+n)}\mathrm{E}_{M\in\mathbb{S}_{(1)}}(||M^\dagger||^\alpha).$$

Proof. From Corollary 18 and Lemma 21,

$$\int_{f\in\mathbb{S}}\sum_{\zeta\in V(f)}\mu_{\text{norm}}(f,\zeta)^\alpha\, d\mathbb{S} = \frac{\mathcal{D}Vol(B(\mathbb{C}^p))}{\int_0^1 t^{2N+1-\alpha}\,dt}\int_{M\in B(\mathcal{H}_{(1)})}(1-\|M\|^2)^p||M^\dagger||^\alpha\, d\mathcal{H}_{(1)},$$

where $p = N - n^2 - n + 1$. Taking polar coordinates for M, this equals

$$\frac{\mathcal{D}Vol(B(\mathbb{C}^p))}{\int_0^1 t^{2N+1-\alpha}\,dt}\int_0^1(1-s^2)^p s^{2n^2+2n-1-\alpha}\,ds\int_{M\in\mathbb{S}_{(1)}}||M^\dagger||^\alpha\, d\mathcal{H}_{(1)}.$$

The proposition follows from

$$Vol(B(\mathbb{C}^k)) = \frac{\pi^k}{\Gamma(k+1)},\quad Vol(\mathbb{S}(\mathbb{C}^k)) = 2\frac{\pi^k}{\Gamma(k)},\quad Vol(\mathbb{S}) = 2\frac{\pi^{N+1}}{\Gamma(N+1)},$$

$$\int_0^1(1-s^2)^p s^{2n^2+2n-1-\alpha}\,ds = \frac{\Gamma(p+1)\Gamma(n^2+n-\alpha/2)}{2\Gamma(N+2-\alpha/2)}.$$

□

Finally, from Proposition 22 and Theorem 19 we conclude the following result.

Theorem 23. *Let $0 < \alpha < 4$ be a real number. Then,*

$$\mathrm{E}_{f\in\mathbb{S}}\left(\sum_{\zeta\in V(f)}\mu_{\mathrm{norm}}(f,\zeta)^{\alpha}\right)=\frac{\mathcal{D}\Gamma(N+1)}{\Gamma(N+1-\alpha/2)}\sum_{k=0}^{n-1}\frac{\binom{n+1}{k}\Gamma(n-k+1-\alpha/2)}{n^{n-k+1-\alpha/2}\Gamma(n-k)}.$$

In particular,

$$\mathrm{E}_{f\in\mathbb{S}}\left(\sum_{\zeta\in V(f)}\mu_{\mathrm{norm}}(f,\zeta)^{2}\right)=\mathcal{D}N\left(n\left(1+\frac{1}{n}\right)^{n+1}-2n-1\right)\leq nN\mathcal{D}.$$

Theorem 5 is immediate from Theorem 13 and Theorem 23.

9. Proof of Theorem 7

Following the $\Theta,\hat{\Theta}$ notation of Corollary 18 with $\phi\equiv 1$, we denote

$$I(M,\zeta)=\int_{h\in B(R_\zeta)}\hat{\Theta}(\sqrt{1-\|M\|_F^2}\,h+\varphi(M,\zeta),\zeta)\,dR_\zeta.$$

Note that

$$\int_{(g,\zeta)\in\mathcal{G}_{(d)}}\Theta(g,\zeta)\,d\mathcal{G}_{(d)}=\int_{(M,l,\zeta,h)\in Y}\Theta(G_{(d)}(M,l,\zeta,h))\,dY=$$

$$\int_{(M,l)\in B(\mathbb{C}^{N+1})}\int_{\zeta\in V_{\mathbb{S}}(M)}I(M,\zeta)\,dV_{\mathbb{S}}(M)\,d\mathbb{C}^{N+1}.$$

Projecting elements $(M,l)\in\mathbb{C}^{N+1}$ onto M, this last equals

$$Vol(B(\mathbb{C}^p))\int_{M\in B(\mathcal{H}_{(1)})}(1-\|M\|^2)^p\int_{\zeta\in V_{\mathbb{S}}(M)}I(M,\zeta)\,dV_{\mathbb{S}}(M)\,d\mathcal{H}_{(1)},$$

where $p=N-n^2-n+1$. From Corollary 18, we conclude that

$$\int_{(g,\zeta)\in\mathcal{G}_{(d)}}\Theta(g,\zeta)\,d\mathcal{G}_{(d)}=\frac{2\pi Vol(B(\mathbb{C}^p))}{(2N+2)\mathcal{D}}\int_{f\in\mathbb{S}}\sum_{\zeta\in V(f)}\Theta(f,\zeta)\,d\mathbb{S}.$$

Thus,

$$\mathrm{E}_{(g,\zeta)\in\mathcal{G}_{(d)}}(\Theta(g,\zeta))=\frac{\pi Vol(\mathbb{S})Vol(B(\mathbb{C}^p))}{(N+1)Vol(\mathcal{G}_{(d)})\mathcal{D}}\mathrm{E}_{f\in\mathbb{S}}\left(\sum_{\zeta\in V(f)}\Theta(f,\zeta)\right).$$

The theorem follows substituting

$$\frac{\pi Vol(\mathbb{S})Vol(B(\mathbb{C}^p))}{(N+1)Vol(\mathcal{G}_{(d)})}=\frac{\pi Vol(\mathbb{S})Vol(B(\mathbb{C}^p))}{2\pi(N+1)Vol(B(\mathbb{C}^{N+1}))Vol(B(\mathbb{C}^p))}=1.$$

10. Uniform equidistribution in the output set

In a general setting, it is interesting to analyze the probability distribution of the output of a probabilistic algorithm. Assume for example that we have a deterministic algorithm for finding one zero of systems of equations, such that for every fixed system, it produces always the same zero. In some sense, the "amount of information" that such an algorithm gives us is small. On the other hand, if the algorithm involves some random choices and all the roots are equidistributed, then the amount of information that the algorithm provides is big, as there is no "hidden" solution that will scape from our algorithm easily. We want to bring to the attention of the reader a concept from Information Theory and Random Number Generator Theory, which seems to be a useful way to measure the intermediate states between those two extremes:

Definition 24. *Let* ALG *be an algorithm which may involve some random choices and which may produce different outputs $x_1, \ldots, x_r$. Let p_i be the probability that output x_i is produced. The Shannon Entropy of* ALG *is*

$$H(\text{ALG}) = -\sum_{i=1}^{r} p_i \log(p_i).$$

It is a simple exercise to show that $H(\text{ALG}) \leq \log r$, and equality holds if and only if $p_i = 1/r$ for all i. Thus, H being maximal is equivalent to equidistribution of outputs, and the closer it is to maximal, the closer we are to equidistribution.

As claimed in Theorem 2, for algorithm AHMR we have equidistribution of the output. Of course, we do not need to express this fact using Shannon's Entropy. However, we want to insist that it might be a very useful concept in other settings. For example, it is used in [BeLe] to analyze data and then conjecture the equidistribution of the roots using another randomized algorithm (for which no equidistribution or complexity results have been proved so far). We have found no previous work where this concept is used in the context of analysis of algorithms. Hence, although we will not use this concept for other purposes on this paper, we want to insist on its potential importance.

We write now a more detailed version of our Theorem 2

Theorem 25. *Let f be such that $\zeta \in V(f)$ implies $(f, \zeta) \notin \Sigma'$. Hence, f has exactly $\mathcal{D}$ solutions $\zeta_1, \ldots, \zeta_{\mathcal{D}}$. Run* AHMR *on input f. Then,*

- *With probability* 1*, the algorithm produces an output z, which is an approximate zero of some exact zero $\zeta \in V(f)$.*
- *Every exact zero of f is equally probable as the one associated to the output of* AHMR*, namely* $\text{Prob}(\zeta = \zeta_i) = 1/\mathcal{D}$ *for* $1 \leq i \leq \mathcal{D}$.

10.1. **Proof of Theorem 25.** As in the proof of Theorem 13, we let $\Sigma = \{h \in \mathbb{S} : (f, \zeta) \in \Sigma' \text{ for some } \zeta \in V(h)\}$. Our hypotheses is that $f \notin \Sigma$. Moreover, Σ is a (real) algebraic variety of $\mathbb{S}$ of real codimension 2, and thus for every $g \in \mathbb{S} \setminus Z$ (Z a null set), the arc $L_{f,g}$ does not intersect Σ (this argument can be formalized as in the proof of Proposition 12). Let $g \in \mathbb{S} \setminus Z$. Assume that some solution ζ of f is fixed. Let

$$\begin{array}{cccl} \Theta : & V & \to & \{0,1\} \\ & (g, \zeta_0) & \mapsto & \begin{cases} 1 & (f,\zeta) \in \Gamma(f, g, \zeta_0) \\ 0 & \textit{otherwise} \end{cases} \end{array}$$

The mapping (4.1) is a bijection between the solutions of g and those of f. Thus, for fixed $g \in \mathbb{S} \setminus Z$ we have $\Theta(\zeta_0) = 1$ exactly for one solution of g. Hence,

$$\text{E}_{g \in \mathbb{S} \setminus Z} \left(\frac{1}{\mathcal{D}} \sum_{\zeta_0 \in V(g)} \Theta(g, \zeta_0) \right) = \frac{1}{\mathcal{D}}.$$

Being Z a null set and using Theorem 7 we conclude,

$$\text{E}_{(g,\zeta_0) \in \mathcal{G}_{(d)}} \left(\Theta(g, \zeta_0) \right) = \frac{1}{\mathcal{D}},$$

namely the probability that a randomly chosen pair $(g, \zeta_0) \in \mathcal{G}_{(d)}$ satisfies $\zeta \in \Gamma(f, g, \zeta_0)$ is exactly equal to $1/\mathcal{D}$. This finishes the proof.

10.2. **Finding few, some, most or all solutions.** In this section we analyze the probability of getting some or all solutions of a system by repeatedly using algorithm AHMR. Namely, we will use AMHR a number of times on the same intput $f \in \mathbb{S}$. This is equivalent to considering the algorithm

AHMR^s

Input: $f \in \mathbb{S}$.

- Choose randomly s pairs $(g^1, \zeta_0^1), \ldots, (g^s, \zeta_0^s) \in \mathcal{G}_{(d)}$.
- Approximate the curves $\Gamma(f, g^s, \zeta_0^s)$ using the homotopy algorithm of [Bel].

Output: s approximate zeros of f, with associated zeros the ones lying on $\Gamma(f, g^j, \zeta_0^j)$, $1 \leq j \leq s$.

The average running time of AHMR^s is s times the average running time of AHMR, thus at most $O(sd^{3/2}nN(N+n^3))$ from Corollary 9. The following result follows from Theorem 25 and elementary computations of Enumerative Probability.

Corollary 26. *Fix $1 \leq s$, $1 \leq k \leq \mathcal{D}$ and let $f \in \mathbb{S} \setminus \Sigma$. The probability that* AHMR^s *produces approximations to k or more different solutions of f, is at least*

$$1 - \binom{\mathcal{D}}{k-1} \frac{(k-1)^s}{\mathcal{D}^s}$$

Proof. The probability (in the space $\mathcal{G}_{(d)}^s$) that AHMR^s produces $k-1$ or less different solutions of f is equal to the probability that a randomly chosen $Z := (z_1, \ldots, z_s) \in \{1, \ldots, \mathcal{D}\}^s$ (w.r.t. the uniform distribution) satisfies

$$\sharp\{z_1, \ldots, z_s\} \leq k - 1.$$

According to the Inclusion-Exclusion Principle of Probability Theory, this last is at most

$$\binom{\mathcal{D}}{k-1} \frac{(k-1)^s}{\mathcal{D}^s}.$$

The corollary follows. □

This result allows us to obtain some upper bounds on the probability that some or all the solutions of f are reached within s tries of AHMR. For example,

Corollary 27. *Fix $l \geq 1$ and let $f \in \mathbb{S} \setminus \Sigma$. Let $s = \lceil 2l\mathcal{D} \log \mathcal{D} \rceil$. Running s times algorithm* AHMR *on input f produces approximations to all the solutions of $\mathcal{D}$, with probability greater than or equal to $1 - \mathcal{D}^{-l}$.*

Proof. From Corollary 26, the probability we are trying to compute is at least

$$1 - \binom{\mathcal{D}}{\mathcal{D}-1} \frac{(\mathcal{D}-1)^{2l\mathcal{D}\log\mathcal{D}}}{\mathcal{D}^{2l\mathcal{D}\log\mathcal{D}}} = 1 - \frac{(\mathcal{D}-1)^{2l\mathcal{D}\log\mathcal{D}}\mathcal{D}}{\mathcal{D}^{2l\mathcal{D}\log\mathcal{D}}}.$$

Applying logarithms, we check that this last is greater than or equal to $1 - \mathcal{D}^{-l}$ as claimed. □

Corollary 27 proves for the first time that, accepting a very small probability of failure, a complete description of the solution set of a system $f \in \mathbb{S} \setminus \Sigma$ can be done with average running time $O(d^{3/2}nN(N+n^3)\mathcal{D}\log\mathcal{D})$, that is linear in the Bézout number $\mathcal{D}$ and almost quadratic in the input length.

It is clear that any description of an approximate zero of every solution of f requires running time at least as big as $\mathcal{D}$, for $\mathcal{D}$ projective points are needed. Moreover, from [CGH$^+$03] we know that any other "natural" encoding of the complete solution set of f, no matter how compressed, cannot be obtained in running time less than $\mathcal{D}$. Thus, our algorithm AHMR is essentially optimal as a universal solver, at least with respect to its average complexity.

References

[AlGe] E. L. Allgower and K. Georg, *Numerical continuation methods*, Springer Series in Computational Mathematics, vol. 13, Springer-Verlag, Berlin, 1990, An introduction.

[AAR] G.E. Andrews, R. Askey, and R. Roy, *Special functions*, Encyclopedia of Mathematics and its Applications, vol. 71. Cambridge University Press 1999.

[BuCu] P. Bürgisser and F. Cucker, *On a problem posted by Smale*, To appear.

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

[Bel] C. Beltrán, *A continuation method to solve polynomial systems, and its complexity*, To appear, `http://sites.google.com/site/beltranc/preprints`.

[BHSW] D.J. Bates, J.D. Hauenstein, A.J. Sommese, and C. Wampler, *Bertini: software for numerical algebraic geometry*, Available at `http://www.nd.edu/∼sommese/bertini`.

[BaSt] W. Baur. and V. Strassen, *The complexity of partial derivatives*, Theoret. Comput. Sci. **22** (1983), no 3, 317–330.

[BeLe] C. Beltrán and A. Leykin, *Certified numerical homotopy tracking*, To appear, arXiv:0912.0920, 2009.

[BePa1] C. Beltrán and L.M. Pardo, *On the complexity of non–universal polynomial equation solving: old and new results.*, Foundations of Computational Mathematics: Santander 2005. L. Pardo, A. Pinkus, E. Süli, M. Todd editors., Cambridge University Press, 2006, pp. 1–35.

[BePa2] ———, *On Smale's 17th problem: a probabilistic positive solution*, Found. Comput. Math. **8** (2008), no. 1, 1–43.

[BePa3] ———, *Smale's 17th problem: Average polynomial time to compute affine and projective solutions*, J. Amer. Math. Soc. **22** (2009), 363–385.

[BeSh] C. Beltrán and M. Shub, *A note on the finite variance of the averaging function for polynomial system solving*, Found. Comput. Math. **10**, no. 1, 115–125.

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

[BoPa] C.E. Borges and L.M. Pardo, *On the probability distribution of data at points in real complete intersection varieties*, J. Complexity **24** (2008), 492–523.

[ChDo] Z. Chen and J. J. Dongarra, *Condition numbers of Gaussian random matrices*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 3, 603–620.

[CGH$^+$03] D. Castro, M. Giusti, J. Heintz, G. Matera, and L. M. Pardo, *The hardness of polynomial equation solving*, Found. Comput. Math. **3** (2003), no. 4, 347–420.

[Ede] A. Edelman, *Eigenvalues and condition numbers of random matrices*, Ph. D. Thesis, Math. Dept. MIT.

[EdSu] A. Edelman, B.D Sutton, *Tails of Condition Number Distributions*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 2, 547–560.

[Har] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, New York, 1977.

[LLT] T.L Lee, T.Y. Li, and C.H. Tsai, *Hom4ps-2.0: A software package for solving polynomial systems by the polyhedral homotopy continuation method*, Available at http://hom4ps.math.msu.edu/HOM4PS_soft.htm.

[Ley] A. Leykin. *Numerical algebraic geometry for Macaulay2.* To appear, arXiv:0911.1783, 2009.

[Ren] J. Renegar, *On the worst-case arithmetic complexity of approximating zeros of polynomials*, J. Complexity **3** (1987), no. 2, 90–113.

[SeKu] I.E. Segal, R.A. Kunze, *Integrals and operators* Second Edition. Springer-Verlag, Berlin, 1978.

[Shu1] M. Shub, *Some remarks on Bezout's theorem and complexity theory*, From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990) (New York), Springer, 1993, pp. 443–455.

[Shu2] ———, *Complexity of Bézout's theorem. VI: Geodesics in the condition (number) metric*, Found. Comput. Math. **9** (2009), no. 2, 171–178.

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

[ShSm1] M. Shub and S. Smale, *Complexity of Bézout's theorem. I. Geometric aspects*, J. Amer. Math. Soc. **6** (1993), no. 2, 459–501.

[ShSm2] ———, *Complexity of Bezout's theorem. II. Volumes and probabilities*, Computational algebraic geometry (Nice, 1992), Progr. Math., vol. 109, Birkhäuser Boston, Boston, MA, 1993, pp. 267–285.

[ShSm4] ———, *Complexity of Bezout's theorem. IV. Probability of success; extensions*, SIAM J. Numer. Anal. **33** (1996), no. 1, 128–148.

[ShSm5] ———, *Complexity of Bezout's theorem. V. Polynomial time*, Theoret. Comput. Sci. **133** (1994), no. 1, 141–164, Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993).

[SoWa] A. J. Sommese and C. W. Wampler, II, *The numerical solution of systems of polynomials*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005, Arising in engineering and science.

[Ver1] J. Verschelde, *Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation.*, ACM Trans. Math. Softw. **25** (1999), no. 2, 251–276, Available at `http://www.math.uic.edu/∼jan`.

[Ver2] J. Verschelde, *Math Review of "On Smale's 17th problem: a probabilistic positive solution" by C. Beltrán and L.M. Pardo.* MR2403529 (2009h:65082) in Mathscinet.

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

E-mail address: `beltranc@unican.es,luis.pardo@unican.es`