Rigid continuation paths I. Quasilinear average complexity for solving polynomial systems

Pierre Lairez

RIGID CONTINUATION PATHS I. QUASILINEAR AVERAGE COMPLEXITY FOR SOLVING POLYNOMIAL SYSTEMS

PIERRE LAIREZ

Abstract. How many operations do we need on the average to compute an approximate root of a random Gaussian polynomial system? Beyond Smale's 17th problem that asked whether a polynomial bound is possible, we prove a quasi-optimal bound (input size)$^{1+o(1)}$. This improves upon the previously known (input size)$^{\frac{3}{2}+o(1)}$ bound.

The new algorithm relies on numerical continuation along *rigid continuation paths*. The central idea is to consider rigid motions of the equations rather than line segments in the linear space of all polynomial systems. This leads to a better average condition number and allows for bigger steps. We show that on the average, we can compute one approximate root of a random Gaussian polynomial system of n equations of degree at most D in $n+1$ homogeneous variables with $O(n^5D^2)$ continuation steps. This is a decisive improvement over previous bounds that prove no better than $\sqrt{2}^{\min(n,D)}$ continuation steps on the average.

Contents

1. Introduction

Following a line of research opened in the 20th century by Smale (1985, 1986), Renegar (1987, 1989), Demmel (1988), Shub (1993), Malajovich (1994), and Shub and Smale (1993a,b,c, 1994, 1996) and developped in the 21st century by Armentano et al. (2016, 2017), Bates et al. (2013), Beltrán (2011), Beltrán and Pardo (2008, 2009a,b, 2011), Beltrán and Shub (2009), Briquel et al. (2014), Bürgisser and Cucker (2011, 2013), Hauenstein and Liddell (2016), Hauenstein and Sottile (2012), Lairez (2017), and Malajovich (2016), to name a few, I am interested in the number of elementary operations that one need to compute one zero of a polynomial system in a numerical setting. On this topic, Smale's question is a landmark: "Can a zero

Date: November 9, 2017.
2000 *Mathematics Subject Classification.* Primary 68Q25; Secondary 65H10, 65H20, 65Y20.

of n complex polynomial equations in n unknowns be found approximately, on the average, in polynomial time with a uniform algorithm?" (Smale 1998, 17th problem). The wording is crafted to have a positive answer in spite of two major obstacles. The first one is the NP-completeness of many problems related to deciding the feasability of a polynomial system. Here, we consider well determined systems (as many equations as unknowns), over the complex numbers, in the average (*a fortiori* generic) case, so there will always be a zero. Second obstacle, the number of zeros is not polynomially bounded in terms of the input size (the number of coefficients that define the input system). Here, we ask for only one zero and numerical methods can take advantage of it.

Smale's question is now solved (Beltrán and Pardo 2009b; Bürgisser and Cucker 2011; Lairez 2017), it is an achievement but not an end. The most obvious question that pops up is to improve the degree hidden behind the words "polynomial time". This article presents an optimal answer, bringing down "polynomial time", that is $N^{O(1)}$, where N is the input size, to "quasilinear time", that is $N^{1+o(1)}$. Previous state of the art was $N^{\frac{3}{2}+o(1)}$.

1.1. **State of the art.** Let n and $d_1, \dots, d_n$ be positive integers, and let $\mathcal{H}$ be the vector space of tuples $(f_1, \dots, f_n)$ of complex homogeneous polynomials of degree $d_1, \dots, d_n$ respectively in the variables $x_0, \dots, x_n$. Let also D denote $\max(d_1, \dots, d_n)$.

We are interested in the average complexity of finding one zero of a polynomial system, given as an element of $\mathcal{H}$. The complexity is measured with respect to the *input size*, denoted N. This is the number of complex coefficients that describe a system, namely

$$N \doteq \dim_{\mathbb{C}} \mathcal{H} = \binom{d_1 + n}{n} + \cdots + \binom{d_n + n}{n}.$$

Note that $N \geqslant 2^{\min(n,D)}$. "Average complexity" means that we endow $\mathcal{H}$ with a probability measure (uniform on the unit sphere for some suitably chosen Hermitian norm) and that we analyze the behaviour of our algorithms *on the average*, assuming that the input is distributed according to this probability measure. We will make use of randomized algorithms, that draw random numbers during the computation. In this case, the average complexity is an average with respect to both the input's distribution and the randomness used internally by the algorithm.

1.1.1. *Classical theory.* In the Shub–Smale–Beltràn–Pardo–Bürgisser–Cucker way of doing things, we compute a zero of a homogeneous polynomial system $F \in \mathcal{H}$ by numerical continuation from a random system $G \in \mathcal{H}$ of which we happen to know a zero $\zeta \in \mathbb{P}(\mathbb{C}^{n+1})$. The continuation is performed along the deformation $F_t \doteq \frac{1}{\|tF+(1-t)G\|}(tF + (1-t)G)$. Starting from $t = 0$, we repeatedly increment the parameter t and track a zero of F_t with a projective Newton iteration applied to the previous approximation of the zero. If the increment is small enough then we can be sure not to loose the zero and to obtain, when t reaches 1, an approximate zero of the target system F. The total complexity of the algorithm depends on the

number of continuation steps that are performed, which in turn depends on the size of the increment. The key issue is to specify how small is "small enough". A sufficient condition was given by Shub and Smale (1993b) in terms of the *condition number* $\mu(F, z)$, a number which characterizes how much a zero z of a system F is affected by a small pertubation of F. After some refinements, Shub (2009) proved that $K(F, G, \zeta)$, the minimal number of steps to go from G to F while tracking the zero ζ, is bounded by

$$K(F, G, \zeta) \leqslant (\text{constant}) \int_0^1 \mu(F_t, \zeta_t) \sqrt{\|\dot{F}_t\| + \|\dot{\zeta}_t\|^2} \mathrm{d}t. \tag{1}$$

This is called the "μ estimate". Explicit algorithms that achieve this bound have been designed by Beltrán and Pardo (2011), Dedieu et al. (2013), and Hauenstein and Liddell (2016). A simpler but weaker form, called the "μ^2 estimate", reads

$$K(F, G, \zeta) \leqslant (\text{constant}) D^{\frac{3}{2}} d_{\mathbb{S}}(F, G) \int_0^1 \mu(F_t, \zeta_t)^2, \tag{2}$$

where $d_{\mathbb{S}}(F, G)$ is the distance in the unit sphere $\mathbb{S}(\mathcal{H})$ from F to G, that is the length of the continuation path. It is often used in practice because it is much easier to design algorithms that achieve this bound rather than the former. In one form or the other, this kind of integral estimate for the number of steps is the first mainstay of the method.

The second mainstay is a procedure discovered by Beltrán and Pardo (2011) and simplified by Bürgisser and Cucker (2011) to sample a Gaussian random system $G \in \mathcal{H}$ together with one of its zeros without the need for solving a polynomial system: (1) sample a random Gaussian linear map $L : \mathbb{C}^{n+1} \to \mathbb{C}^n$, (2) compute a nonzero vector $\zeta \in \mathbb{C}^{n+1}$ in the kernel of L and (3) sample a random Gaussian system in the linear subspace of $\mathcal{H}$ of all systems G such that $G(\zeta) = 0$ and $\mathrm{d}_\zeta G = L$. By construction, we obtain a system G and one of its zeros ζ. Less trivially, $G/\|G\|$ is uniformly distributed in the sphere $\mathbb{S}(\mathcal{H})$. We could think of a simpler procedure that (1) samples some $\zeta \in \mathbb{C}^{n+1}$ isotropically and (2) sample a random Gaussian system in the linear subspace of $\mathcal{H}$ of all systems G such that $G(\zeta) = 0$. This also gives a random system with one of its zeros, by construction, but the system is not uniformly distributed in the sphere after normalization.

These two mainstays together give a randomized algorithm to compute a zero of a polynomial system and a way to analyze its average complexity on a random input. On input $F \in \mathbb{S}(\mathcal{H})$, the algorithm is: (1) sample a random uniformly system $G \in \mathbb{S}(\mathcal{H})$ together with a zero ζ, (2) perform the numerical continuation from G to F tracking the zero ζ. If F itself is a uniformly distributed random variable, then for any $t \in [0, 1]$, F_t is also uniformly distributed, so (F_t, ζ_t) has the same distribution as (G, ζ). Therefore, the average number of steps performed by the algorithm is bounded by

$$\begin{aligned}
\mathbb{E}(K(F, G, \zeta)) &\leqslant (\text{constant})\, \mathbb{E}\left(D^{\frac{3}{2}} d_{\mathbb{S}}(F, G) \int_0^1 \mu(F_t, \zeta_t)^2 \mathrm{d}t \right) \\
&\leqslant (\text{constant}) D^{\frac{3}{2}} \int_0^1 \mathbb{E}\left(\mu(F_t, \zeta_t)^2\right) \mathrm{d}t
\end{aligned}$$

$$\leqslant (\text{constant}) D^{\frac{3}{2}} \mathbb{E}\left(\mu(G,\zeta)^2\right).$$

This leads us to the third mainstay: estimates for $\mathbb{E}\left(\mu(G,\zeta)^2\right)$. Beltrán and Pardo (2011, Theorem 23) proved that $\mathbb{E}\left(\mu(G,\zeta)^2\right) \leqslant nN$. Therefore, the average number of steps performed by the algorithm on a random input is

$$\mathbb{E}(K(F,G,\zeta)) \leqslant (\text{constant}) nN.$$

The cost of each continuation step (basically, the computation of μ and a Newton's iteration) can be done in $O(N)$ operations (when $D \geqslant 2$). All in all, the total average complexity of the algorithm is $O(nD^{\frac{3}{2}}N^2)$ when $N \to \infty$. When $\min(n, D) \to \infty$, then this is $N^{2+o(1)}$.

1.1.2. *Improvements.* How can we improve upon this complexity bound? We cannot do much about the $O(N)$ cost of a continuation step, it is already optimal. Concerning the number of steps, we can try to use the μ estimate instead of the μ^2 estimate. Bounding $\|\dot{\zeta}_t\|$ by $\mu(F_t,\zeta_t)\|\dot{F}_t\|$ (which turns the μ estimate into the μ^2 estimate) is optimal in the worst case, but on the average, when the direction $\dot{F}_t$ is random, this is pessimistic. Building upon this idea, Armentano et al. (2016) proved that $O(nD^{\frac{3}{2}}N^{\frac{1}{2}})$ continuation steps are enough on the average. This leads to a total average complexity of $N^{\frac{3}{2}+o(1)}$ operations.

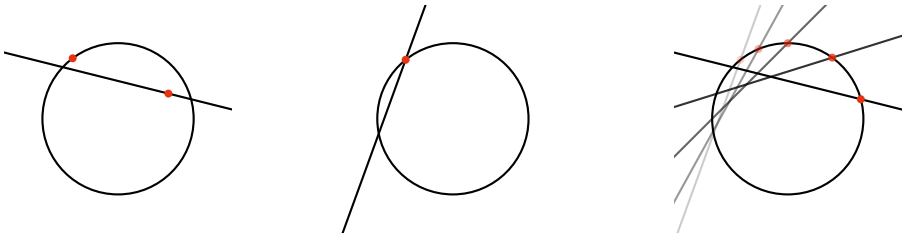
Beltrán and Shub (2009) proved that there exist continuation paths that makes the μ estimate polynomially bounded in terms of n and D. The construction is explicit but it requires the knowledge of a zero of a target system. This prevents it from being used algorithmically. Yet, it was the first time that the possibility of performing numerical continuation in very few steps (polynomially many with respect to n and D, not N) was supported.

Lastly, let us mention that the γ estimate of Hauenstein and Liddell (2016) may be used as a starting point to obtain very similar results to ours in a more traditional context. However, this direction is yet to be explored.

1.2. **Contribution.** I describe an algorithm that performs numerical continuation in $O(n^5D^2)$ steps on the average (and each step costs $O(nD^2N)$ operations). This leads to a total average complexity of $O(n^6D^4N)$ operations when $N \to \infty$ so find one approximate root of a random Gaussian system (Corollary 33). When $\min(n, D) \to \infty$, this is $N^{1+o(1)}$. The algorithm relies on analogues of the three mainstays of the classical theory: integral estimate for the number of steps, randomization of the start system and average analysis of some condition number. However, the basic tools are thoroughly renewed.

The starting point is the observation that a typical system in $\mathcal{H}$ is poorly conditioned. As mentionned above, the expected squared condition number of a random system at a random zero is bounded by nN and it turns out that this is rather sharp. In view of Smale's question, this is satisfying, much more than bounds involving the total number of zeros, but this N is the limit of the method.

To improve the average conditioning, an idea is to define the notion of conditioning with respect to a much lower dimensional parameter space, but still big enough to be able to develop an analogue of Beltràn and Pardo's algorithm. I propose here



(a) Compute one solution of each equation.

(b) Move the hypersurfaces to make the solution match.

(c) Continuously return to the original position while tracking the solution.

Figure 1. Resolution of a polynomial system with a rigid continuation path.

the *rigid* setting, where the parameter space is not the whole space of polynomial systems, but the group $\mathcal{U}$ made of n copies of the unitary group $U(n+1)$. The group $\mathcal{U}$, of real dimension n^3, acts by rigid motions on the n components of a fixed, well determined, polynomial system. Figure 1 illustrates a *rigid continuation path.*

Less parameters is less opportunities for a dramatic pertubation that will ruin the conditioning of a system. Beyond that, the continuation paths in the rigid setting preserve the geometry of the input equations. This opens a way for studying the average complexity of solving certain *structured* systems. Forthcoming work will address the topic.

A noteworthy contribution is the introduction of the *split gamma number* which tightly upperbounds Smale's gamma number (Theorem 11) and which allows for interesting average analysis, see §4.3.

Ackowledgment. It is my pleasure to thank Carlos Beltràn, Peter Bürgisser and Felipe Cucker for many helpful discussions and valuable comments.

1.3. **Notations and basic definitions.**

n	some positive integer (used as the *number of non homogeneous variables*).
$\mathbb{P}$	complex projective space of dimension n.
$[z]$	projective class of some nonzero $z \in \mathbb{C}^{n+1}$.
H_d	space of complex homogeneous polynomials of degree d in $x_0, \ldots, x_n$.
r	some positive integer (used as the *number of equations*).
$d_1, \ldots, d_r$	some positive integers (used as the *degree of the equations*).
D	the maximum of $d_1, \ldots, d_r$
$\mathcal{H}[r]$	space of homogeneous systems of r equations of degree $d_1, \ldots, d_r$, that is $H_{d_1} \times \cdots \times H_{d_r}$. Elements of $\mathcal{H}[r]$ are often considered as polynomial maps $\mathbb{C}^{n+1} \to \mathbb{C}^r$.
$U(k)$	group of unitary $k \times k$ matrices.
$\mathcal{U}$	the group of r-uples of unitary matrices, $U(n+1)^r$. Elements of $\mathcal{U}$ are denoted in boldface, like $\mathbf{u}$.
$\mathbf{1}_{\mathcal{U}}$	$(\mathrm{id}, \ldots, \mathrm{id})$, the neutral element in $\mathcal{U}$.
$\|-\|$	norm in a Hermitian space.

$\|-\|_W$	Weyl norm of a polynomial (see ??).
$\|\|\|-\|\|\|$	operator norm of a map between Hermitian spaces. For a multilinear map $\varphi : E^k \to V$, this is $\sup\{\|\varphi(e_1,\dots,e_k)\| \mid \|e_1\| = \cdots = \|e_k\| = 1\}$.
$\|-\|_{\mathrm{Frob}}$	Frobenius norm of a map between Hermitian spaces.
$\|-\|_u$	$1/\sqrt{2}$ times $\|-\|_{\mathrm{Frob}}$ (used as the Riemannian metric on the tangent spaces of $U(n+1)$).
$\varphi^\dagger$	Moore-Penrose pseudo-inverse of a surjective linear map $\varphi : E \to F$, it is the unique linear map $\psi : F \to E$ such that $\varphi\psi = \mathrm{id}_F$ and $\psi\varphi$ is the orthogonal projection onto the row space of φ (the orthogonal complement of the kernel).
$\mathrm{d}_z F$	the derivative of some polynomial map $F : \mathbb{C}^{n+1} \to \mathbb{C}^r$ at $z \in \mathbb{C}^{n+1}$. We will use the same notation with $z \in \mathbb{P}$, which means that we choose a representative $\bar{z} \in \mathbb{C}^{n+1}$ of z such that $\|\bar{z}\| = 1$.
$\mathrm{d}_z F^\dagger$	the pseudo-inverse of the derivative.
$\mathcal{N}_F$	projective Newton's operator associated to F
$\doteq$	"is defined as"
$A = O(B)$ as $C \to \infty$	"there are $C_0 \geqslant 0$ and $k \geqslant 0$ such that $C \geqslant C_0 \Rightarrow A \leqslant kB$".
standard normal variable	a Gaussian random variable of an Euclidean space with unit covariance matrix.

2. Rigid solution varieties

The classical solution variety is the subvariety of $\mathcal{H}[r] \times \mathbb{P}$ of all (F, ζ) such that ζ is a zero of F. We now introduce an analogue variety in the rigid setting.

Let $X_1, \dots, X_r$ be pure-dimensional subvarieties of $\mathbb{P}$, with $\sum_i \operatorname{codim} X_i \leqslant n$. Let $\mathcal{U}$ denote the group $U(n+1)^r$. It acts naturally on the product $(\mathbb{P})^r$ of r copies of the projective space. We denote its elements in boldface $\mathbf{u} = (u_1, \dots, u_r)$. Let $\mathcal{X}$ denote the product variety $X_1 \times \cdots \times X_r \subset (\mathbb{P})^r$. For $\mathbf{u} \in \mathcal{U}$, let $\mathbf{u}\mathcal{X}$ denote the image of $\mathcal{X}$ under the action of $\mathbf{u}$, that is $\prod_{i=1}^r u_i X_i$, and let $\cap\mathbf{u}\mathcal{X}$ denote the intersection $\cap_{i=1}^r u_i X_i \subseteq \mathbb{P}$. The *rigid solution variety* is defined as

$$\mathcal{V} \doteq \{(\mathbf{u}, x) \in \mathcal{U} \times \mathbb{P} \mid x \in \cap\mathbf{u}\mathcal{X}\}.$$

There is not a single solution variety, but rather any choice of subvarieties $X_1, \dots, X_r$ leads to a solution variety. In this section, we will study the geometry of $\mathcal{V}$ with $X_1, \dots, X_r$ fixed. Later on, we will assume that $X_1, \dots, X_r$ are hypersurfaces defined by random polynomials.

Let $\mathbb{G}(k)$ denote the Grassmannian of k-dimensional projective subspaces of $\mathbb{P}$, that is $k+1$-dimensional linear subspaces of $\mathbb{C}^{n+1}$. For a generic point $x \in X_i$, the projectivization of the tangent space of the cone over X_i at some representative $\bar{x} \in \mathbb{C}^{n+1}$ of x is an element of $\mathbb{G}(k_i)$ and is denoted $\mathbb{T}_x X_i$. If X_i is the zero set of some homogeneous polynomial system $F_i \in \mathcal{H}[m]$, then $\mathbb{T}_x(u_i X_i)$ is the projectivization of the kernel of $\mathrm{d}_x(F_i \circ u_i^*)$. Let $\mathcal{L} \doteq \mathbb{G}(\dim X_1) \times \cdots \times \mathbb{G}(\dim X_r)$, and for $\mathbf{h} = (h_1, \dots, h_r) \in \mathcal{L}$, let $\cap\mathbf{h}$ denote the intersection of the h_i in $\mathbb{P}$. To a generic

point $(\mathbf{u}, x)$ of $\mathcal{V}$, we associate the linearization

$$L(\mathbf{u}, x) \doteq \big(\mathbb{T}_x(u_1 X_1), \ldots, \mathbb{T}_x(u_r X_r)\big) \in \mathcal{L}.$$

This section aims at three goals: describe precisely the so-called *standard* distribution on $\mathcal{V}$ (Theorem 7), give an algorithm to sample from this distribution (Algorithm 1) and define the *split gamma number*, a variant of the gamma number well adapted to the rigid setting.

2.1. **Determinant of subspaces and incidence condition number.** Let $E_1, \ldots, E_r$ be linear subspaces of a Hermitian space V. Let π_i be the orthogonal projector on E_i. We define the *multiprojection* map $\operatorname{proj}(E_1, \ldots, E_r)$ by

$$\begin{aligned} \operatorname{proj}(E_1, \ldots, E_r) : \mathbb{C}^{n+1} &\longrightarrow E_1 \times \cdots \times E_r \\ v &\longmapsto (\pi_i v, \ldots, \pi_r v). \end{aligned}$$

We say that the family $E_1, \ldots, E_r$ is non-degenerate if $\sum_i \dim E_i = \dim\left(\sum_i E_i\right)$, or, equivalently, when the multiprojection map is not surjective.

We define the *determinant* of $E_1, \ldots, E_r$ as

$$\det(E_1, \ldots, E_r) \doteq \left|\det\left(\operatorname{proj}(E_1, \ldots, E_r)_{|E_1 + \cdots + E_r}\right)\right|,$$

in the non-degenerate case and $\det(E_1, \ldots, E_r) \doteq 0$ otherwise. Note that the determinant of a map between two Hermitian spaces is well defined up to multiplication by some $e^{i\theta}$, so that the modulus is well defined. We also define the *orthogonal determinant* of $E_1, \ldots, E_r$ as

$$\det{}^{\perp}(E_1, \ldots, E_r) \doteq \det(E_1^{\perp}, \ldots, E_r^{\perp}).$$

Lastly, we define the *incidence condition number* of $E_1, \ldots, E_r$ as

$$\kappa(E_1, \ldots, E_r) \doteq \left\|\left|\operatorname{proj}(E_1, \ldots, E_r)^{\dagger}\right|\right\|$$

when the multiprojection map is surjective, and $\kappa(E_1, \ldots, E_r) \doteq \infty$ otherwise. With the appropriate distance, the incidence condition number is the inverse of the distance of the tuple $(E_1, \ldots, E_r)$ to the closest $(F_1, \ldots, F_r)$ such that $\dim\left(\sum_i F_i\right) < \sum_i \dim F_i$ (Breiding and Vannieuwenhoven 2016, Theorems 1.1 and 1.3).

Lemma 1. *Let $P \doteq \pi_1 + \cdots + \pi_r$, it is a self-adjoint operator $\sum_i E_i \to \sum_i E_i$. In the non-degenerate case, $\det(E_1, \ldots, E_r)^2 = \det P$ and $\kappa(E_1, \ldots, E_r)^2 = \left\|\left|P^{-1}\right|\right\|$.*

Proof. The Hermitian transpose of $\operatorname{proj}(E_1, \ldots, E_r)$ is the map $\operatorname{sum}(E_1, \ldots, E_r) : \sum_i E_i \to \mathbb{C}^{n+1}$ defined by

$$\operatorname{sum}(E_1, \ldots, E_r) : (u_1, \ldots, u_r) \longmapsto u_1 + \cdots + u_r.$$

Clearly, $P = \operatorname{sum}(E_1, \ldots, E_r) \cdot \operatorname{proj}(E_1, \ldots, E_r)$ and the claims follow. □

Lemma 2. *For any subspaces $E_1, \ldots, E_r \subseteq \mathbb{C}^{n+1}$,*

$$\det{}^{\perp}(E_1, \ldots, E_r) = \det{}^{\perp}(E_1, E_2) \det{}^{\perp}(E_1 \cap E_2, E_3, \ldots, E_r).$$

Proof. This follows from the factorization

$$\operatorname{proj}(E_1^\perp, \dots, E_r^\perp)_{\sum_i E_i^\perp} = \left(\operatorname{proj}(E_1^\perp, E_2^\perp)_{|E_1^\perp + E_2^\perp} \times \operatorname{id}_{E_3^\perp} \times \cdots \times \operatorname{id}_{E_r^\perp}\right) \circ \operatorname{proj}(E_1^\perp + E_2^\perp, E_3 \dots, E_r^\perp)_{|\sum_i E_i^\perp}. \quad \square$$

2.2. **Reminders on Riemannian geometry.** We will work mainly with two Riemannian manifolds: $\mathbb{P}$, the n-dimensional complex projective space endowed with the Fubini–Study metric, and $U(n+1)$, the group of $(n+1) \times (n+1)$ unitary matrices. Concerning the latter, we endow $\mathbb{C}^{(n+1)\times(n+1)}$ with the norm

$$\|A\|_u \doteq \tfrac{1}{\sqrt{2}} \|A\|_{\mathrm{Frob}} \doteq \sqrt{\tfrac{1}{2} \operatorname{Tr}(AA^*)}, \quad A \in \mathbb{C}^{(n+1)\times(n+1)}, \tag{3}$$

and we choose on $U(n+1)$ the Riemannian metric induced from the embedding of $U(n+1)$ in $\mathbb{C}^{(n+1)\times(n+1)}$. This metric is invariant under left and right multiplication in $U(n+1)$.

Let X and Y be Riemannian manifolds and let $f : X \to Y$ be a continously differentiable map. For any $x \in X$, we define the *normal Jacobian* of f at x as

$$\mathrm{NJ}_x f \doteq \sqrt{\det\left(\mathrm{d}_x f \cdot \mathrm{d}_x f^*\right)}.$$

When $\mathrm{d}_x f$ is bijective, this is the absolute value of the usual Jacobian. A fundamental result is the *coarea formula*: for any integrable map $\Theta : X \to \mathbb{R}$,

$$\int_X \mathrm{d}x\ \Theta(x) = \int_Y \mathrm{d}y \int_{f^{-1}(y)} \mathrm{d}x\ \frac{\Theta(x)}{\mathrm{NJ}_x f} \tag{4}$$

Under the hypothesis that f is differentiable, this formula follows easily from parititions of unity and Fubini's theorem.

The special case of Riemannian submersions is important. We say that f is a *Riemannian submersion* if for any $x \in X$, the derivative $\mathrm{d}_x f$ induces an isometry from $(\ker \mathrm{d}_x f)^\perp$ to $T_{f(x)}Y$. In that case, we easily check two things: f is Lipschitz-continuous with constant 1 and $\mathrm{NJ}_x f = 1$ for all $x \in X$. Note also that for any submanifold Z of Y, if f is a Riemannian submersion then so is $f_{|f^{-1}(Z)}$. The scaling in the definition of $\| - \|_u$ is chosen to have the following result.

Lemma 3. *For any $p \in \mathbb{P}$, the map $\varphi : u \in U(n+1) \mapsto up \in \mathbb{P}$ is a Riemannian submersion. In particular, for any variety $X \subseteq \mathbb{P}$, and any integrable map $\Theta : \pi^{-1}X \to \mathbb{R}$,*

$$\int_{\pi^{-1}X} \mathrm{d}u\ \Theta(u) = \int_X \mathrm{d}x \int_{up=x} \mathrm{d}u\ \Theta(u),$$

where $\int_{up=x} \mathrm{d}u$ denotes the integration over the variety $\varphi^{-1}(x)$.

Proof. Thanks to the invariance of the Riemannian metric of $U(n+1)$ under right multiplication, it is enough to check that the defining property of Riemannian submersion holds at id, the identity matrix. With a suitable choice of coordinates, we may also assume that $p = [1 : 0 : \cdots : 0]$. The tangent space of $\mathbb{P}$ at p is canonically identified with $\{p\}^\perp$, that is $\{0\} \times \mathbb{C}^n$.

The tangent space $T_{\mathrm{id}}U(n+1)$ of $U(n+1)$ at id is the space of skew-Hermitian matrices, and for any $u \in T_{\mathrm{id}}U(n+1)$, $\mathrm{d}_{\mathrm{id}}\varphi(\dot{u}) = \dot{u}p$. Therefore,

$$(\ker \mathrm{d}_{\mathrm{id}}\varphi)^{\perp} = \left\{ \left(\begin{smallmatrix} 0 & v^* \\ v & \mathbf{0} \end{smallmatrix}\right) \mid v \in \mathbb{C}^n \right\},$$

and since $\left\| \left(\begin{smallmatrix} 0 & v^* \\ v & \mathbf{0} \end{smallmatrix}\right) \right\|_u = \|v\|$, the map $\dot{u} \in (\ker \mathrm{d}_{\mathrm{id}}\varphi)^{\perp} \to \dot{u}p$ is clearly an isometry.

The second claim follows from coarea formula (4), noting that the restriction of π to $\pi^{-1}X$ is again a Riemannian submersion. □

2.3. **Basic integral formulas.** For our problem, the manifold $\mathcal{V}$ has a natural distribution, the *standard distribution*, denoted ρ_{std}, defined as follows. Let $\mathbf{u} \in \mathcal{U}$ and $x \in \cap \mathbf{u}\mathcal{X}$ be uniformly distributed (both $\mathcal{U}$ and $\mathbf{u}\mathcal{X}$ are compact Riemannian manifolds, so "uniformly distributed" is well defined), the random variable $(\mathbf{u}, x)$ belong to $\mathcal{V}$ and ρ_{std} is defined to be its probability distribution.

For any $x \in \mathbb{P}$, $\mathbf{y} \in \mathcal{X}$ and $\mathbf{h} \in \mathcal{L}$, we define

$$\begin{aligned} \mathcal{L}_x &\doteq \{\mathbf{h} \in \mathcal{L} \mid x \in \cap \mathbf{h}\}, \\ \mathcal{U}_x &\doteq \{\mathbf{u} \in \mathcal{U} \mid x \in \cap \mathbf{u}\mathcal{X}\} \\ \mathcal{U}_{x,\mathbf{y}} &\doteq \{\mathbf{u} \in \mathcal{U}_x \mid \mathbf{u}\mathbf{y} = (x, \dots, x)\}, \\ \mathcal{U}_{x,\mathbf{y},\mathbf{h}} &\doteq \{\mathbf{u} \in \mathcal{U}_{x,\mathbf{y}} \mid L(\mathbf{u}, x) = \mathbf{h}\}. \end{aligned}$$

Lemma 4. *For any two submanifolds X and Y of $\mathbb{P}$, with $\operatorname{codim} X + \operatorname{codim} Y \leqslant n$, and for any integrable function $\Theta : \mathbb{P} \times U(n+1) \to \mathbb{R}$,*

$$\int_{U(n+1)} \mathrm{d}u \int_{X \cap uY} \mathrm{d}z\ \Theta(z,u) = \int_X \mathrm{d}x \int_Y \mathrm{d}y \int_{uy=x} \mathrm{d}u\ \Theta(x,u) \det{}^{\perp}(\mathbb{T}_x X, \mathbb{T}_x uY),$$

(where $\int_{uy=x} \mathrm{d}u$ denotes the integration over the variety of all $u \in U(n+1)$ such that $uy = x$, as in Lemma 3).

Proof. It is a corollary of the "basic integral formula" of Howard (1993, §2.7). Let $p \in \mathbb{P}$ be some point and let φ be the Riemannian submersion $u \in U(n+1) \mapsto up \in \mathbb{P}$ (Lemma 3). Let $M = \varphi^{-1}X$ and $N = \varphi^{-1}Y$. Howard's basic integral formula asserts that

$$\int_{U(n+1)} \mathrm{d}u \int_{M \cap uN} \mathrm{d}v\ \Theta(\varphi v, u) = \int_M \mathrm{d}v \int_N \mathrm{d}w\ \Theta(\varphi v, vw^{-1}) \det{}^{\perp}(v^{-1}T_v M, w^{-1}T_w N), \tag{5}$$

where $v^{-1}T_vM$ and $w^{-1}T_wN$ are subspace of $T_{\mathrm{id}}U(n+1)$. (The equality of our $\det^{\perp}$ with Howard's σ is given by Lemma 1.) On the one hand, by Lemma 3, we obtain the following expression for the left-hand side of (5):

$$\begin{aligned} \int_{U(n+1)} \mathrm{d}u \int_{M \cap uN} \mathrm{d}v\ \Theta(\varphi v, u) &= \int_{U(n+1)} \mathrm{d}u \int_{X \cap uY} \mathrm{d}z \int_{vp=z} \mathrm{d}v\ \Theta(z,u) \\ &= \operatorname{vol}(U(1) \times U(n)) \int_{U(n+1)} \mathrm{d}u \int_{X \cap uY} \mathrm{d}z\ \Theta(z,u), \end{aligned}$$

the last because $\{v \in U(n+1) \mid vp = z\}$ is isometric, by some left multiplication, to $U(1) \times U(n)$, the stabilizer of a point in $\mathbb{P}$. This is the left-hand side of the claimed equality. On the other hand, regarding the right-hand side of (5), we check easily

that

$$\begin{aligned}\det{}^{\perp}(v^{-1}T_vM, w^{-1}T_wN) &= \det{}^{\perp}(v^{-1}\mathbb{T}_{vp}X, w^{-1}\mathbb{T}_{wp}Y)\\ &= \det{}^{\perp}\left(\mathbb{T}_{vp}X, \mathbb{T}_{vp}(vw^{-1}Y)\right)\end{aligned}$$

and therefore (using Lemma 3 again)

$$\begin{aligned}&\int_M \mathrm{d}v \int_N \mathrm{d}w\ \Theta(\varphi v, vw^{-1}) \det{}^{\perp}(v^{-1}T_vM, w^{-1}T_wN)\\ &\qquad= \int_X \mathrm{d}x \int_{vp=x} \mathrm{d}v \int_Y \mathrm{d}y \int_{wp=y} \mathrm{d}w\ \Theta(x, vw^{-1}) \det{}^{\perp}(\mathbb{T}_xX, \mathbb{T}_xvw^{-1}Y)\\ &\qquad= \mathrm{vol}(U(1)\times U(n)) \int_X \mathrm{d}x \int_Y \mathrm{d}y \int_{uy=x} \mathrm{d}u\ \Theta(x,u) \det{}^{\perp}(\mathbb{T}_xX, \mathbb{T}_xuY),\end{aligned}$$

where the last equality is given by the change of variables $v = uw$. This gives the right-hand side of the claim. □

In our setting where we consider r varieties $X_1, \dots, X_r$, we can give the following "basic integral formula". It has been proved very similarly in the real case by Bürgisser and Lerario (2016, §7.5).

Proposition 5. *For any measurable function $\Theta : \mathcal{V} \to [0,\infty)$*

$$\int_{\mathcal{U}} \mathrm{d}\mathbf{u} \int_{\cap \mathbf{u}\mathcal{X}} \mathrm{d}x\ \Theta(\mathbf{u}, x) = \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{U}_x} \mathrm{d}\mathbf{u}\ \Theta(\mathbf{u}, x) \det{}^{\perp}(L(\mathbf{u}, x)).$$

Proof. We proceed by induction on r. The case where $r = 0$ is trivial. Assume that the property holds for $r-1$ subvarieties $X_1, \dots, X_{r-1}$, for some $r \geqslant 1$, and let $\mathcal{U}'$, $\mathcal{X}'$, etc. denote the analogues of $\mathcal{U}$, $\mathcal{X}$, etc. for varieties $X_1, \dots, X_{r-1}$. From the decomposition $\mathcal{U} = \mathcal{U}' \times U(n+1)$, we obtain by Lemma 4

$$\begin{aligned}&\int_{\mathcal{U}} \mathrm{d}\mathbf{u} \int_{\cap \mathbf{u}\mathcal{X}} \mathrm{d}x\ \Theta(\mathbf{u}, x) = \int_{\mathcal{U}'} \mathrm{d}\mathbf{u}' \int_{U(n+1)} \mathrm{d}u_r \int_{(\cap \mathbf{u}'\mathcal{X}') \cap u_rX_r} \mathrm{d}x\ \Theta(\mathbf{u}', u_r, x)\\ &= \int_{\mathcal{U}'} \mathrm{d}\mathbf{u}' \int_{\cap \mathbf{u}'\mathcal{X}'} \mathrm{d}x \int_{X_r} \mathrm{d}y \int_{u_ry=x} \mathrm{d}u_r\ \Theta(\mathbf{u}', u_r, x) \det{}^{\perp}(\mathbb{T}_x(\cap \mathbf{u}'\mathcal{X}'), \mathbb{T}_x u_r X_r)\\ &= \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{U}'_x} \mathrm{d}\mathbf{u}' \int_{X_r} \mathrm{d}y \int_{u_ry=x} \mathrm{d}u_r\ \Theta(\mathbf{u}', u_r, x) \det{}^{\perp}(\mathbb{T}_x(\cap \mathbf{u}'\mathcal{X}'), \mathbb{T}_x u_r X_r) \det{}^{\perp}(L(\mathbf{u}', x))\end{aligned}$$

the last using the induction hypothesis. Lemma 2 shows that

$$\det{}^{\perp}(\mathbb{T}_x(\cap \mathbf{u}'\mathcal{X}'), \mathbb{T}_x u_r X_r) \det{}^{\perp}(L(\mathbf{u}', x)) = \det{}^{\perp}(L(\mathbf{u}, x)).$$

Moreover, the map $u_r \in U(n+1) \mapsto u_r^* x \in \mathbb{P}$ is a Riemannian submersion so

$$\int_{X_r} \mathrm{d}y \int_{u_ry=x} \mathrm{d}u_r h(u_r) = \int_{x \in u_rX_r} \mathrm{d}u_r h(u_r),$$

for any integrable function $h : U(n+1) \to \mathbb{R}$. This implies that

$$\int_{\mathcal{U}} \mathrm{d}\mathbf{u} \int_{\cap \mathbf{u}\mathcal{X}} \mathrm{d}x\ \Theta(\mathbf{u}, x) = \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{U}'_x} \mathrm{d}\mathbf{u}' \int_{x \in u_rX_r} \mathrm{d}u_r\ \Theta(\mathbf{u}', u_r, x) \det{}^{\perp}_x(L(\mathbf{u}, x)).$$

To conclude, we remark that $\mathcal{U}_x = \mathcal{U}'_x \times \{u_r \in U(n+1) \mid x \in u_rX_r\}$. □

If we apply the statement above to the case where the varieties X_i are projective subspaces, we obtain the following corollary.

Corollary 6. *For any measurable function* $\Theta : \mathcal{L} \to [0, \infty)$

$$\int_{\mathcal{L}} \mathrm{d}\mathbf{h} \int_{\cap \mathbf{h}} \mathrm{d}x\ \Theta(\mathbf{h}, x) = \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{L}_x} \mathrm{d}\mathbf{h}\ \Theta(\mathbf{h}, x) \det{}^{\perp}(\mathbf{h}).$$

Proof. Let's assume that each X_i is a projective subspace of $\mathbb{P}$. We define the map

$$\mathbf{u} \in \mathcal{U} \mapsto \mathbf{u}\mathcal{X} = (u_1 X_1, \ldots, u_r X_r) \in \mathcal{L},$$

which is a Riemannian submersion. In particular,

$$\int_{\mathcal{U}} \mathrm{d}\mathbf{u} \int_{\cap \mathbf{u}\mathcal{X}} \mathrm{d}x\ \Theta(\mathbf{u}\mathcal{X}, x) = \mathrm{vol}(\mathrm{Stab}_{\mathcal{U}}\, \mathcal{X}) \int_{\mathcal{L}} \mathrm{d}\mathbf{h} \int_{\cap \mathbf{h}} \mathrm{d}x\ \Theta(\mathbf{h}, x).$$

and

$$\int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{U}_x} \mathrm{d}\mathbf{u}\ \Theta(\mathbf{u}\mathcal{X}, x) \det{}^{\perp}(L(\mathbf{u}, x)) = \mathrm{vol}(\mathrm{Stab}_{\mathcal{U}}\, \mathcal{X}) \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{L}_x} \mathrm{d}\mathbf{h}\ \Theta(\mathbf{h}, x) \det{}^{\perp}(\mathbf{h}).$$

This reduces the claim to Proposition 5. □

We now have all we need to prove the main result of this section.

Theorem 7. *For any measurable function* $\Theta : \mathcal{V} \mapsto [0, +\infty)$*,*

$$\int_{\mathcal{U}} \mathrm{d}\mathbf{u} \int_{\cap \mathbf{u}\mathcal{X}} \mathrm{d}x\ \Theta(\mathbf{u}, x) = \int_{\mathcal{L}} \mathrm{d}\mathbf{h} \int_{\mathcal{X}} \mathrm{d}\mathbf{y} \int_{\cap \mathbf{h}} \mathrm{d}x \int_{\mathcal{U}_{x,\mathbf{y},\mathbf{h}}} \mathrm{d}\mathbf{u}\ \Theta(\mathbf{u}, x).$$

In other words, if $(\mathbf{u}, x) \in \mathcal{V}$ *is a* ρ_{std}*-distributed random variable distributed, then:*

- *(i) the r. v.* $L(\mathbf{u}, x)$ *is uniformly distributed in* $\mathcal{L}$*;*
- *(ii) the r. v.* $\mathbf{y} \doteq (u_1^* x, \ldots, u_r^* x)$ *is uniformly distributed in* $\mathcal{X}$ *and independent from* $L(\mathbf{u}, x)$*;*
- *(iii) conditionally to* $L(\mathbf{u}, x)$*, the r. v.* x *is uniformly distributed in* $\cap L(\mathbf{u}, x)$*;*
- *(iv) conditionally to* $L(\mathbf{u}, x)$*,* x *and* $\mathbf{y}$*, the r. v.* $\mathbf{u}$ *is uniformly distributed in* $\mathcal{U}_{x,\mathbf{y},L(\mathbf{u},x)}$*.*

Proof. By Corollary 6,

$$\begin{aligned}\int_{\mathcal{L}} \mathrm{d}\mathbf{h} \int_{\cap \mathbf{h}} \mathrm{d}x \int_{\mathcal{X}} \mathrm{d}\mathbf{y} \int_{\mathcal{U}_{x,\mathbf{y},\mathbf{h}}} \mathrm{d}v\ \Theta(\mathbf{u}, x) &= \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{L}_x} \mathrm{d}\mathbf{h} \det(\mathbf{h}) \int_{\mathcal{X}} \mathrm{d}\mathbf{y} \int_{\mathcal{U}_{x,\mathbf{y},\mathbf{h}}} \mathrm{d}v\ \Theta(\mathbf{u}, x) \\ &= \int_{\mathbb{P}} \mathrm{d}x \int_{\mathcal{X}} \mathrm{d}\mathbf{y} \int_{\mathcal{L}_x} \mathrm{d}\mathbf{h} \int_{\mathcal{U}_{x,\mathbf{y},\mathbf{h}}} \mathrm{d}v\ \Theta(\mathbf{u}, x) \det{}^{\perp}(L(\mathbf{u}, x)).\end{aligned}$$

Moreover, the map $\mathbf{u} \in \mathcal{U}_{x,\mathbf{y}} \mapsto L(\mathbf{u}, x) \in \mathcal{L}_x$ is a Riemannian submersion, thus

$$\int_{\mathcal{L}_x} \mathrm{d}\mathbf{h} \int_{\mathcal{U}_{x,\mathbf{y},\mathbf{h}}} \mathrm{d}v\ \Theta(\mathbf{u}, x) \det{}^{\perp}(L(\mathbf{u}, x)) = \int_{\mathcal{U}_{x,\mathbf{y}}} \mathrm{d}v\ \Theta(\mathbf{u}, x) \det{}^{\perp}(L(\mathbf{u}, x)).$$

The map $\mathbf{u} \in \mathcal{U}_x \mapsto (u_1^* x, \ldots, u_m^* x) \in \mathcal{X}$ is also a Riemannian submersion, thus

$$\int_{\mathcal{X}} \mathrm{d}\mathbf{y} \int_{\mathcal{U}_{x,\mathbf{y}}} \mathrm{d}v\ \Theta(\mathbf{u}, x) \det{}^{\perp}(L(\mathbf{u}, x)) = \int_{\mathcal{U}_x} \mathrm{d}v\ \Theta(\mathbf{u}, x) \det{}^{\perp}(L(\mathbf{u}, x)).$$

To conclude, we apply Proposition 5. □

2.4. **Sampling the solution variety.** Based on Theorem 7, Algorithm 1 can be used to sample a ρ_{std}-distributed random $(\mathbf{u}, \zeta) \in \mathcal{V}$. We explain briefly how to perform the four steps of the algorithm.

Algorithm 1. Sampling of a unitary solution variety
Input. $\varnothing$
Output. $(\mathbf{u}, \zeta) \in \mathcal{V}$.
Postcondition. $(\mathbf{u}, \zeta) \sim \rho_{\mathrm{std}}$ (Theorem 7)

function Sample
Sample $h_1 \in \mathbb{G}(\dim X_i), \dots, h_r \in \mathbb{G}(\dim X_r)$, uniformly and independently.
Sample $x \in h_1 \cap \cdots \cap h_r \subset \mathbb{P}$ uniformly.
Sample $y_1 \in X_1, \dots, y_r \in X_r$ uniformly and independently.
Sample $u_1, \dots, u_r \in U(n+1)$, such that $u_i y_i = x$ and $u_i(\mathbb{T}_{y_i} X_i) = h_i$, uniformly and independently.
return $(u_1, \dots, u_n) \in \mathcal{U}$ and $\zeta \in \mathbb{P}$.
end function

For each $1 \leqslant i \leqslant r$, we sample independently linear forms $\lambda_{i,1}, \dots, \lambda_{i,\operatorname{codim} X_i} \in (\mathbb{C}^{n+1})^*$ with a standard normal distribution. We define h_i as the zero locus of $\lambda_{i,1}, \dots, \lambda_{i,\operatorname{codim} X_i}$. Next, we compute a unitary basis of the linear subspace $h_1 \cap \cdots \cap h_n$ and use it to sample $x \in \mathbb{P}(h_1 \cap \cdots \cap h_n)$ with a uniform distribution.

To sample uniformly a point $y_i \in X_i$, we consider a random uniformly distributed subspace $L_i \in \mathbb{G}(\operatorname{codim} X_i)$. Almost surely, the intersection $X_i \cap L_i$ is finite and we sample uniformly a point y_i in it. This requires a polynomial system solving in $\operatorname{codim} X_i + 1$ homogeneous variables. The fact that y_i is uniformly distributed in X_i is a consequence of Theorem 7(ii) applied to two varieties, one is X_i and the other a fixed element of $\mathbb{G}(\operatorname{codim} X_i)$. Therefore, sampling y_i boils down to computing one zero of a homogeneous polynomial system in $\operatorname{codim} X_i + 1$ variables and $\operatorname{codim} X_i$ equations.

Once we get the y_i, we compute, for each $1 \leqslant i \leqslant r$, some $v_i \in U(n+1)$ which maps y_i to x and $\mathbb{T}_{y_i} X_i$ to h_i and we sample uniformly a w_i in the subgroup of all $w \in U(n+1)$ such that $w\zeta = \zeta$ and $wL_i = L_i$. This subgroup is isometrically isomorphic to $U(1) \times U(\dim X_i) \times U(\operatorname{codim} X_i)$. We can sample uniformly in a unitary group by considering the Q factor of the QR decomposition of a random Gaussian matrix. And then, we define $u_i \doteq w_i v_i$. When $X_1, \dots, X_r$ are all hypersurfaces, we have proved the following proposition.

Proposition 8. *If $X_1, \dots, X_r$ are all hypersurfaces, Algorithm 1 samples a ρ_{std}-distributed point in the solution variety $\mathcal{V}$ with*

- *r zero finding of bivariate homogeneous polynomials of degrees* $\deg X_1, \dots,$ $\deg X_r$ *respectively;*
- *$O(n^3)$ samplings of the standard normal distribution on $\mathbb{R}$;*
- *$O(n^4)$ arithmetic operations.*

2.5. **The split gamma number.** In the classical theory, the condition number μ plays two roles: First, by definition, it bounds the variation of a zero after a pertubation of the system. Second, it is an upper bound for the gamma with some regularity properties (the Lipschitz properties). Each role is reflected by a factor μ in the μ^2 estimate.

In the rigid setting, the two roles are played by different numbers: the variation of a zero is bounded by the incidence condition number κ and the upper bound for γ that we use is the *split gamma number* $\hat{\gamma}$. This will lead to a $\kappa\hat{\gamma}$ estimate for the complexity of numerical continuation in the rigid setting. In this section, we introduce the split gamma number and we start with some reminders about the gamma number.

Let $F = (f_1, \dots, f_r) \in \mathcal{H}[s]$ be a homogeneous polynomial system that we regard as a polynomial map $\mathbb{C}^{n+1} \to \mathbb{C}^s$. Let $d_i \doteq \deg f_i$ and $D \doteq \max_i d_i$.

When $s = n$, the polynomial system F has generically finitely many zeros and our primary goal is to compute them numerically and approximately. A fundamental tool is Newton's operator. We use here the projective version introduced by Shub (1989). For $z \in \mathbb{P}$, projective class of some $\bar{z} \in \mathbb{C}^{n+1}$, we define

$$\mathcal{N}_F(z) \doteq \left[\bar{z} - \mathrm{d}_{\bar{z}}F|_{z^\perp}^{-1}\left(F(\bar{z})\right)\right] \in \mathbb{P}, \tag{6}$$

where $z^\perp$ is the orthogonal complement of z in $\mathbb{C}^{n+1}$. The definition does not depend on the choice of $\bar{z}$. Given a zero $\zeta \in \mathbb{P}$ of F, we say that $z \in \mathbb{P}$ *approximates* ζ *as a zero of* F, or that z *is an approximate zero of* F *with associate zero* ζ, if for any $k \geqslant 0$,

$$d_{\mathbb{P}}\left(\mathcal{N}_F^k(z), \zeta\right) \leqslant 2^{1-2^k} d_{\mathbb{P}}(z, \zeta).$$

The main result of the gamma theory is a sufficient condition for a point to approximate a zero. For a polynomial system F, in the general case $r \leqslant n$, we will use the following definition for the gamma number of F at $z \in \mathbb{P}$:

$$\gamma(F, z) \doteq \begin{cases} \sup_{k \geqslant 2} \left\| \frac{1}{k!} \mathrm{d}_z F^\dagger \cdot \mathrm{d}_z^k F \right\|^{\frac{1}{k-1}} & \text{if } \mathrm{d}_z F \text{ is sujective,} \\ \infty & \text{otherwise.} \end{cases}$$

(The definition does not depend on the choice of a unit representative $\bar{z}$ of z.) We follow here the definition used by Shub and Smale (1996) and Dedieu (2006, §4). When $s = n$, the pseudo-inverse $\mathrm{d}_z F^\dagger$ is often replaced by $\mathrm{d}_z F|_{z^\perp}^{-1}$, as in Newton's iteration (*e.g.* Bürgisser and Cucker 2013; Shub and Smale 1994). If z is a zero of F, both definitions coincide, so the gamma theorem (Theorem 10) is equally true for both variants.

When $F = (f)$ is a single equation, that is $s = 1$, it is useful to remark that $\mathrm{d}_z f$ is a linear form and so $\mathrm{d}_z f^\dagger$ is simply $\|\mathrm{d}_z f\|^{-1}\pi$, where π is an isometric embedding of $\mathbb{C}$ in $\mathbb{C}^{n+1}$. This gives $\gamma(f, z)$ the following form:

$$\gamma(f, z) = \sup_{k \geqslant 2} \left(\tfrac{1}{k!} \|\mathrm{d}_z f\|^{-1} \|\mathrm{d}_z^k f\|\right)^{\frac{1}{k-1}}. \tag{7}$$

The following lower bound will be occasionally useful.

Lemma 9. *For any* $z \in \mathbb{P}$, $\gamma(F, z) \geqslant \frac{D-1}{2}$.

Proof. We may assume that $\mathrm{d}_z F$ is surjective, otherwise the bound is trivial. Let $d_i \doteq \deg f_i$. Using the homogeneity, for any $u \in \mathbb{C}^{n+1}$,

$$\mathrm{d}_z^2 F(z, u) = \begin{pmatrix} d_1 - 1 & & \\ & \ddots & \\ & & d_m - 1 \end{pmatrix} \mathrm{d}_z F(u). \tag{8}$$

We fix some $1 \leqslant i \leqslant r$ and consider $u \doteq \mathrm{d}_z F^\dagger(e_i)$, where $e_i \doteq (0, \dots, 1, \dots, 0) \in \mathbb{C}^r$ with a single one at the ith position. Becase $\mathrm{d}_z F$ is surjective, $\mathrm{d}_z F(\mathrm{d}_z F^\dagger(e_i)) = e_i$, and by (8), we have $\mathrm{d}_z^2 F(z,u) = (d_i - 1)e_i$ and then $\mathrm{d}_z F^\dagger(\mathrm{d}_z^2 F(z,u)) = (d_i - 1)u$. This implies that $\left\|\mathrm{d}_z F^\dagger \mathrm{d}_z^2 F\right\| \geqslant d_i - 1$, for any $1 \leqslant i \leqslant m$, and the claim follows. □

We now state the main result of the gamma theory, primarily due to Shub and Smale (1993b).

Theorem 10 (Shub, Smale)**.** *Let $F \in \mathcal{H}[n]$ be a homogeneous polynomial system. For any zero $\zeta \in \mathbb{P}$ of F and any $z \in \mathbb{P}$, if $d_{\mathbb{P}}(z,\zeta)\gamma(F,\zeta) \leqslant \frac{1}{6}$ then z approximates ζ as a zero of F.*

Proof. This is (Bürgisser and Cucker 2013, Theorem 16.38) with $r = 0.981$ (and we use that $\gamma(F,\zeta) \geqslant \frac{1}{2}$ when $D > 1$, Lemma 9). □

Let $F_1 \in \mathcal{H}[s_1], \dots, F_r \in \mathcal{H}[s_r]$ be homogeneous polynomial systems. Based on the incidence condition number of linear subspaces (§2.1), we define the *incidence condition number* of $F_1, \dots, F_r$ at a point $x \in \mathbb{P}$ by

$$\kappa(F_1, \dots, F_r; x) \doteq \kappa\left(\ker(\mathrm{d}_x F_1)^\perp, \dots, \ker(\mathrm{d}_x F_r)^\perp\right),$$

The *split gamma number* of $F_1, \dots, F_r$ at a point $x \in \mathbb{P}$ is defined by

$$\hat{\gamma}(F_1, \dots, F_r; x) \doteq \kappa(F_1, \dots, F_r; x)\left(\gamma(F_1,x)^2 + \cdots + \gamma(F_r,x)^2\right)^{\frac{1}{2}}.$$

The split gamma number separates the contribution of the γ number of each block of equations from the more geometric information contained in κ. Note that when $r = 1$, then $\hat{\gamma}(F,x) = \gamma(F,x)$.

Theorem 11. *Let $G \doteq (F_1, \dots, F_r) \in \mathcal{H}[s_1 + \cdots + s_r]$ denote the concatenation of the systems. For any $x \in \mathbb{P}$,*

$$\gamma(G,x) \leqslant \hat{\gamma}(F_1, \dots, F_r; x) \leqslant r\,\kappa(F_1, \dots, F_r; x)\gamma(G,x).$$

Proof. It is easy to see that $\hat{\gamma}(F_1, \dots, F_r; x)$ is finite if and only if $\gamma(G,x)$ is. Thus, we may assume that $\mathrm{d}G$ and the $\mathrm{d}F_i$ are surjective (we drop the index x in d_x). We begin with the first inequality. Let $K_i \doteq \ker \mathrm{d}_x F_i$ and $P \doteq \mathrm{proj}(K_1^\perp, \dots, K_r^\perp)$. We first prove that for any $k \geqslant 2$ and any $\mathbf{y} = (y_1, \dots, y_k) \in \left(\mathbb{C}^{n+1}\right)^k$,

$$\mathrm{d}G^\dagger \cdot \mathrm{d}^k G(\mathbf{y}) = P^\dagger\left(\mathrm{d}F_1^\dagger \cdot \mathrm{d}^k F_1(\mathbf{y}), \dots, \mathrm{d}F_r^\dagger \cdot \mathrm{d}^k F_r(\mathbf{y})\right). \tag{9}$$

It is clear that

$$\mathrm{d}^k G(\mathbf{y}) = \left(\mathrm{d}^k F_1(\mathbf{y}), \dots, \mathrm{d}^k F_r(\mathbf{y})\right) \in \mathbb{C}^{s_1} \times \cdots \times \mathbb{C}^{s_r}.$$

Let $v_i \doteq \mathrm{d}^k F_i(\mathbf{y})$ and $\mathbf{v} \doteq (v_1, \dots, v_r)$. Because $\mathrm{d}G$ is surjective, we have $\mathrm{d}G \cdot \mathrm{d}G^\dagger(\mathbf{v}) = \mathbf{v}$ and, equivalently, $\mathrm{d}F_i \cdot \mathrm{d}G^\dagger(\mathbf{v}) = v_i$. Therefore $\mathrm{d}F_i^\dagger \cdot \mathrm{d}F_i \cdot \mathrm{d}G^\dagger \mathbf{v} = \mathrm{d}F_i^\dagger v_i$. But $\mathrm{d}F_i^\dagger \cdot \mathrm{d}F_i$ is simply the orthogonal projection on $K_i^\perp$. This gives $P \cdot \mathrm{d}G^\dagger \cdot \mathrm{d}^k G(\mathbf{y}) = \left(\mathrm{d}F_1^\dagger v_1, \dots, \mathrm{d}F_r^\dagger v_r\right)$, and since the image of $\mathrm{d}G^\dagger$ is orthogonal to the kernel of P we have $P^\dagger \cdot P \cdot \mathrm{d}G^\dagger = \mathrm{d}G^\dagger$. This proves (9).

As a consequence, for any $k \geqslant 2$,

$$\begin{aligned}
\left\| \tfrac{1}{k!} \mathrm{d}G^\dagger \cdot \mathrm{d}^k G \right\|^{\frac{1}{k-1}} &\leqslant \left\|\pi^\dagger\right\|^{\frac{1}{k-1}} \left(\sum_i \left\| \tfrac{1}{k!} \mathrm{d}F_i^\dagger \cdot \mathrm{d}^k F_i \right\|^2 \right)^{\frac{1}{2k-2}} \\
&\leqslant \left\|\pi^\dagger\right\|^{\frac{1}{k-1}} \left(\sum_i \gamma(F_i, x)^{2k-2} \right)^{\frac{1}{2k-2}} \\
&\leqslant \left\|\pi^\dagger\right\| \left(\sum_i \gamma(F_i, x)^2 \right)^{\frac{1}{2}}
\end{aligned}$$

By definition, $\kappa(F_1, \ldots, F_r; x) = \left\|\pi^\dagger\right\|$, so we obtain the first inequality.

Concerning the second inequality, Equation (9) implies that

$$\begin{aligned}
\left\| \tfrac{1}{k!} \mathrm{d}F_i^\dagger \cdot \mathrm{d}^k F_i \right\| &\leqslant \|\pi\| \left\| \tfrac{1}{k!} \mathrm{d}G^\dagger \cdot \mathrm{d}^k G \right\| \\
&\leqslant \|\pi\| \, \gamma(G, x)^{k-1}.
\end{aligned}$$

We note that $\|\pi\| \leqslant r^{\frac{1}{2}}$ (as the direct sum of r orthogonal projections with orthogonal images) and therefore $\gamma(F_i, x) \leqslant r^{\frac{1}{2}} \gamma(G, x)$. Hence

$$\left(\sum_{i=1}^{r} \gamma(F_i, x)^2 \right)^{\frac{1}{2}} \leqslant \left(r \cdot r\gamma(G, x)^2 \right)^{\frac{1}{2}} \leqslant r\gamma(G, x),$$

and the second inequality follows. □

3. Solving polynomial systems

In this part, we consider a more specific setting than in the previous one. We are given a polynomial system $F = (f_1, \ldots, f_n) \in \mathcal{H}[n]$ and a $\mathbf{u} \in \mathcal{U} = U(n+1)^n$, and we look for a root of $\mathbf{u} \cdot F \doteq (f_1 \circ u_1^*, \ldots, f_n \circ u_n^*)$. We will perform the average analyses in the case where $\mathbf{u}$ is uniformly distributed and F is fixed. To this end, we consider the rigid solution variety $\mathcal{V}$ relative to the projective hypersurfaces $V(f_1), \ldots, V(f_n)$. In concrete terms, we have $r = n$ and

$$\mathcal{V} = \{(\mathbf{u}, x) \in \mathcal{U} \times \mathbb{P} \mid f_1(u_1^* x) = \cdots = f_n(u_n^* x) = 0\}.$$

The manifold $\mathcal{L}$ is just $(\mathbb{P})^n = \mathbb{P} \times \cdots \times \mathbb{P}$ and $L(\mathbf{u}, x)$ is the n-uple

$$L(\mathbf{u}, x) = ([\mathrm{d}_x(f_1 \circ u_1^*)], \ldots, [\mathrm{d}_x(f_n \circ u_n^*)]).$$

In particular, we can define $L(\mathbf{u}, x)$ on $\mathcal{U} \times \mathbb{P}$, not only on $\mathcal{V}$. We can identify $L(\mathbf{u}, x)$ with one of its preimage under the projection map $\mathbb{S}(\mathbb{C}^{n+1})^n \to (\mathbb{P})^n$, that is a $n \times (n+1)$ matrix with unit rows. Namely,

$$L(\mathbf{u}, x) = \operatorname{diag}\left(\|\mathrm{d}_x(f_1 \circ u_1^*)\|^{-1}, \ldots, \|\mathrm{d}_x(f_n \circ u_n^*\|^{-1} \right) \cdot \mathrm{d}_x\left(\mathbf{u} \cdot F\right).$$

Under this identification, we check that

$$\kappa(\mathbf{u} \cdot F, x) = \left\| L(\mathbf{u}, x)^\dagger \right\|. \tag{10}$$

3.1. **Condition number.** The rigid solution variety, considered as a manifold of pairs problem–solution has a natural condition number. We show that this is κ, defined in §2.5. This is what makes the split gamma number $\hat{\gamma}$ fit nicely in the setting of the rigid solution variety. The system F being fixed, we will denote $\kappa(\mathbf{u}\cdot F, x)$ and $\hat{\gamma}(\mathbf{u}\cdot F, x)$ simply as $\kappa(\mathbf{u}, x)$ and $\hat{\gamma}(\mathbf{u}, x)$.

Lemma 12. *For any $(\mathbf{u}, x)\in\mathcal{V}$ and any tangent vector $(\dot{\mathbf{u}}, \dot{x})\in T_{\mathbf{u},x}\mathcal{V}$,*

$$\|\dot{x}\| \leqslant \kappa(\mathbf{u}, x)\|\dot{\mathbf{u}}\|_u.$$

Moreover, for any $\dot{x}\in T_x\mathbb{P}$, there is a $\dot{\mathbf{u}}\in T_{\mathbf{u}}\mathcal{U}$ such that $(\dot{\mathbf{u}}, \dot{x})\in T_{\mathbf{u},x}\mathcal{V}$ and $\|\dot{x}\| = \kappa(\mathbf{u}, x)\|\dot{\mathbf{u}}\|_u$.

Proof. Without loss of generality, we may assume that $\mathbf{u} = \mathbf{1}_{\mathcal{U}}$ and $x = [1 : 0 : \cdots : 0]$, which will simplify notations. Differentiating the relations $f_i(u_i^* x) = 0$ that define $\mathcal{V}$ gives that

$$T_{\mathbf{u},x}\mathcal{V} = \{(\dot{\mathbf{u}}, \dot{x})\in T_{\mathbf{u}}\mathcal{U}\times T_x\mathbb{P} \mid \mathrm{d}_x f_i(\dot{x}) = \mathrm{d}_x f_i(\dot{u}_i x)\}\,. \tag{11}$$

Let $P \doteq \mathrm{proj}(K_1^\perp, \dots, K_n^\perp)$ where $K_i \doteq \ker \mathrm{d}_x f_i$. Note that the coordinate of $P(y)$ along $K_i^\perp$ is $\mathrm{d}_x f_i^\dagger \mathrm{d}_x f_i(y)$. Rewording (11), $(\dot{\mathbf{u}}, \dot{x})\in T_{\mathbf{u},x}\mathcal{V}$ if and only if

$$P(\dot{x}) = \left(\mathrm{d}_x f_i^\dagger \mathrm{d}_x f_i(\dot{u}_i x)\right)_{1\leqslant i\leqslant n}. \tag{12}$$

For all $1\leqslant i\leqslant n$, $\mathrm{d}_x f_i(x) = \deg(f_i) f_i(x) = 0$ (Euler's relation), therefore

$$\|\mathrm{d}_x f_i^\dagger \mathrm{d}_x f_i(\dot{u}_i x)\| = \|\mathrm{d}_x f_i^\dagger \mathrm{d}_x f_i(\pi_x(\dot{u}_i x))\| \leqslant \|\pi_x(\dot{u}_i x)\|,$$

where π_x is the orthogonal projection on $\{x\}^\perp$. We check that $\|\pi_x(\dot{u}_i x)\|\leqslant\|\dot{u}_i\|_u$, as in Lemma 3. If Equation (11) does hold, then $\|P(\dot{x})\|^2 \leqslant \sum_i \|\mathrm{d}_x f_i^\dagger \mathrm{d}_x f_i(\dot{u}_i x)\|^2 \leqslant \sum_i \|\dot{u}_i\|_u^2 = \|\dot{\mathbf{u}}\|_u^2$. Moreover $\dot{x} = P^\dagger(P(\dot{x}))$, because $\dot{x}$ is orthogonal to x and the kernel of P is $\mathbb{C}x$. Therefore $\|\dot{x}\|\leqslant |||P^\dagger|||\,\|\dot{\mathbf{u}}\|_u$. Since $\kappa(\mathbf{u}, x)\doteq |||P^\dagger|||$, this proves the first claim.

For the second claim, let $\dot{x}\in T_x\mathbb{P}$ such that $\|\dot{x}\| = 1$ and $\|P(\dot{x})\|$ minimal, that is equal to $|||P^\dagger|||^{-1}$. Looking at Equation (12) shows that there is a $\dot{\mathbf{u}}\in T_{\mathbf{u}}\mathcal{U}$ such that $(\dot{\mathbf{u}}, \dot{x})\in T_{\mathbf{u},x}\mathcal{V}$ and $\|\dot{u}_i\|_u = |P(\dot{x})_i|^2$. Therefore $\|\dot{\mathbf{u}}\|_u = |||P^\dagger|||^{-1}$. □

Proposition 13. *If $(\mathbf{u}, \zeta)\in\mathcal{V}$ is ρ_{std}-distributed then, $\mathbb{E}\left[\kappa(\mathbf{u}, \zeta)^2\right]\leqslant 6n^2$.*

Proof. Let M be a random $n\times(n+1)$ matrix whose lines are independent and uniformly distributed in $\mathbb{S}(\mathbb{C}^{n+1})$. It follows from (10) and Theorem 7(i) that $\kappa(\mathbf{u}, \zeta)$ has the same probability distribution as $|||M^\dagger|||$.

Let T be an $n\times n$ diagonal random matrix whose coefficients are independent chi distributed random variables with $2n+2$ degrees of freedom, so that TM is a random Gaussian matrix (the coefficients are independent standard normal complex numbers). Obviously, $M^\dagger = (TM)^\dagger\cdot T$ and therefore $|||M^\dagger|||\leqslant |||(TM)^\dagger|||\,|||T|||$. Hölder's inequality with conjugate exponents $n'\doteq 1+\frac{1}{n+1}$ and $n+2$, gives

$$\mathbb{E}\left[|||M^\dagger|||^2\right] \leqslant \mathbb{E}\left[|||(TM)^\dagger|||^{2n'}\right]^{\frac{1}{n'}}\mathbb{E}\left[|||T|||^{2n+4}\right]^{\frac{1}{n+2}}. \tag{13}$$

We now give upper bounds for both factors in the right-hand side. According to Beltrán and Pardo (2011, Theorem 20), whose result is derived from the work of

Edelman (1989),

$$\mathbb{E}\left[\|\|(TM)^\dagger\|\|^{2n'}\right] = 2^{-n'}\sum_{k=1}^{n}\binom{n+1}{k+1}\frac{\Gamma(k-n'+1)}{\Gamma(k)}n^{-k+n'-1}. \tag{14}$$

We proceed as in (Lairez 2017, Theorem 10) and deduce that

$$\mathbb{E}\left[\|\|(TM)^\dagger\|\|^{2n'}\right]^{\frac{1}{n'}} \leqslant \left(\frac{5}{4(2-n')}\right)^{\frac{1}{n'}}\frac{n}{2} \leqslant n. \tag{15}$$

Concerning the second factor, we remark that $\|\|T\|\|^2$ is the maximum of n chi-squared distributed random variables with $2n+2$ degrees of freedom, say $Z_1,\dots,Z_n$, hence

$$\mathbb{E}\left[\|\|T\|\|^{2n+4}\right] \leqslant \sum_{i=1}^{n}\mathbb{E}[Z_i^{n+2}] = 2^{n+2}\frac{(2n+2)!}{(n-1)!}.$$

Therefore,

$$\mathbb{E}\left[\|\|M^\dagger\|\|^2\right] \leqslant n\left(2^{n+2}\frac{(2n+2)!}{(n-1)!}\right)^{\frac{1}{n+2}} \leqslant 6n^2,$$

after a few numerical computations. □

3.2. **Lipschitz properties.** We aim at bounding the variation of the numbers κ and $\hat\gamma$ on the Riemannian manifold $\mathcal{U}\times\mathbb{P}$. In particular, we will prove that $1/\hat\gamma$ is Lipschitz-continuous. Traditionally, such results bound directly the value at some point x with respect to the value at some other point y and the distance from x to y. For example (Dedieu 2006, Lemme 131), for any $x,y\in\mathbb{P}$,

$$\gamma(F,y) \leqslant \gamma(F,x)\,q\left(d_{\mathbb{P}}(x,y)\gamma(F,x)\right), \tag{16}$$

where $q(u) \doteq \frac{1}{(1-u)(1-4u+2u^2)} = 1+5u+O(u^2)$, given that $d_{\mathbb{P}}(x,y)\gamma(F,x) \leqslant 1-1/\sqrt{2}$. As much as possible, I tried to express this kind of inequalities as a bound on the derivative of the function under consideration. At first order, this is equivalent.

Proposition 14. *Let $F\in\mathcal{H}[r]$ and $\gamma_F : x\in\mathbb{P}\mapsto\gamma(F,x)$. For any $x\in\mathbb{P}$, we have $\|\mathrm{d}_x\gamma_F\| \leqslant 5\gamma_F^2$.*

Note that γ_F may not be differentiable everywhere, so the inequality $\|\mathrm{d}_x\gamma_F\| \leqslant 5\gamma_F(x)^2$ really means that

$$\limsup_{y\to x}\frac{|\gamma_F(y)-\gamma_F(x)|}{d_{\mathbb{P}}(y,x)} \leqslant 5\gamma_F(x)^2.$$

It is easy to check that Proposition 14 is equivalent to the Lipschitz continuity of the function $1/\gamma_F$, with Lipschitz constant at most 5. We give $\|\mathrm{d}\kappa\|$ and $\|\mathrm{d}\hat\gamma\|$ an analogue meaning.

Proof of Proposition 14. The most direct way to see this is by (16):

$$\frac{\gamma_F(y)-\gamma_F(x)}{d_{\mathbb{P}}(y,x)} \leqslant \gamma_F(x)\left(5u+O(u^2)\right) = 5\gamma_F(x)^2\left(1+O(d_{\mathbb{P}}(x,y))\right),$$

as $y\to x$, where $u \doteq d_{\mathbb{P}}(x,y)\gamma_F(x)$, and

$$\frac{\gamma_F(x)-\gamma_F(y)}{d_{\mathbb{P}}(y,x)} \leqslant \gamma_F(y)\left(5v+O(v^2)\right) = 5(\gamma_F(x)+o(1))^2\left(1+O(d_{\mathbb{P}}(x,y))\right),$$

where $v \doteq d_{\mathbb{P}}(x,y)\gamma_F(y)$. □

Lemma 15. *On $\mathcal{U} \times \mathbb{P}$, $\|\mathrm{d}\kappa\| \leqslant \kappa^2 + 3\kappa\hat{\gamma}$. Moreover, if $D > 1$ then $\|\mathrm{d}\kappa\| \leqslant 5\kappa\hat{\gamma}$.*

Proof. The second inequality follows from the first one: If $D > 1$ then at least one $\gamma(u_i \cdot f, x)$ is greater or equal to $\frac{D-1}{2}$, by Lemma 9. It follows that $\kappa \leqslant 2\hat{\gamma}$ and then $\kappa^2 + 3\kappa\hat{\gamma} \leqslant 5\kappa\hat{\gamma}$.

To prove the first inequality, we first remark that $1/\kappa(\mathbf{u}, x)$ is a Lipschitz-continuous function of $L(\mathbf{u}, x)$ with constant 1. Indeed, $1/\kappa(\mathbf{u}, x)$ is the least singular of $L(\mathbf{u}, x)$ as a matrix, see Equation (10), and Eckart–Young theorem express this number as the distance to the set of singular matrices, which is a Lipschitz continuous function with constant 1. Moreover $\mathrm{d}\kappa = -\kappa^2 \mathrm{d}\frac{1}{\kappa}$, so it is enough to prove that $\|\mathrm{d}L\|$ is bounded by $1 + 3\frac{\hat{\gamma}}{\kappa}$.

Let $L_i(u_i, z) \in \mathbb{P}$ be the ith component of $L(\mathbf{u}, z)$, that is the projective class of $\mathrm{d}_x(u_i \cdot f_i)$. The tangent space of $\mathbb{P}$ at $L_i(u_i, z)$ is isometrically identified with the quotient $\mathbb{C}^{n+1}/\mathbb{C} \cdot L_i(u_i, z)$. Denoting $h_i \doteq u_i \cdot f_i$, we check that, at a point $(\mathbf{u}, x)$,

$$\mathrm{d}L_i(0, \dot{x}) = \frac{1}{\|\mathrm{d}_x h_i\|} \mathrm{d}_x^2 h_i(\dot{x}) \mod L_i(u_i, z),$$

and in particular, $\|\mathrm{d}L_i(0, \dot{x})\| \leqslant 2\gamma(h_i, x)\|\dot{x}\|$ for any $\dot{x} \in T_x\mathbb{P}$. Besides, $L_i(u_i, u_i x) = L_i(\mathrm{id}, x) \circ u_i^*$, which is a 1-Lipschitz continuous function of u_i (Lemma 3 applied to the dual projective space). This proves that $\|\mathrm{d}L_i(\dot{u}_i, \dot{u}_i x)\| \leqslant \|\dot{u}_i\|_u$. Therefore,

$$\begin{aligned}
\|\mathrm{d}L_i(\dot{u}_i, \dot{x})\| &\leqslant \|\mathrm{d}L_i(\dot{u}_i, \dot{u}_i x)\| + \|\mathrm{d}L_i(0, \dot{x} - \dot{u}_i x)\| \\
&\leqslant \|\dot{u}_i\|_u + 2\gamma(h_i, x)\left(\|\dot{x}\| + \|\dot{u}_i\|_u\right) \\
&\leqslant \|\dot{u}_i\|_u + 2\sqrt{2}\gamma(h_i, x)\left(\|\dot{\mathbf{u}}\|_u^2 + \|\dot{x}\|^2\right)^{\frac{1}{2}},
\end{aligned}$$

and then

$$\begin{aligned}
\|\mathrm{d}L(\dot{\mathbf{u}}, \dot{x})\| &= \left(\sum_{i=1}^{n} \|\mathrm{d}L_i(\dot{u}_i, \dot{x})\|^2\right)^{\frac{1}{2}} \\
&\leqslant \|\dot{\mathbf{u}}\|_u + 2\sqrt{2}\left(\sum_i \gamma(h_i, x)^2\right)^{\frac{1}{2}} \left(\|\dot{\mathbf{u}}\|_u^2 + \|\dot{x}\|^2\right)^{\frac{1}{2}} \\
&\leqslant \left(1 + 3\frac{\hat{\gamma}(\mathbf{u}; x)}{\kappa(\mathbf{u}, x)}\right)\left(\|\dot{\mathbf{u}}\|_u^2 + \|\dot{x}\|^2\right)^{\frac{1}{2}},
\end{aligned}$$

which concludes the proof. □

We now derive a bound for $\|\mathrm{d}\hat{\gamma}\|$.

Lemma 16. *For any homonegeneous polynomial $f : \mathbb{C}^{n+1} \to \mathbb{C}$ the map*

$$(u, x) \in U(n+1) \times \mathbb{P} \longmapsto \frac{1}{\gamma(u \cdot f, x)}$$

is Lipschitz continuous with constant at most $5\sqrt{2}$.

Proof. The γ number is invariant under unitary transformations, that is $\gamma(u \cdot f, x) = \gamma(f, u^* x)$. Moreover, the map $(u, x) \in U(n+1) \times \mathbb{P} \mapsto u^* x \in \mathbb{P}$ is 1-Lipschitz continuous with respect to u (Lemma 3) and to x, thus it is $\sqrt{2}$-Lipschitz

continuous on $U(n+1)\times\mathbb{P}$. Since $1/\gamma(f,x)$ is 5-Lipschitz continous with respect to x (Proposition 14), the map $1/\gamma(f,ux)$ is $5\sqrt{2}$-Lipschitz continuous on $U(n+1)\times\mathbb{P}$. □

Lemma 17. *On $\mathcal{U}\times\mathbb{P}$, $\|\mathrm{d}\hat\gamma\| \leqslant 13\,\hat\gamma^2$. Equivalently, $1/\hat\gamma$ is 13-Lipschitz continuous.*

Proof. We may assume that $D > 1$, otherwise $\hat\gamma$ is identically 0. Let $\gamma_i(\mathbf{u},x) \doteq \gamma(u_i\cdot f_i, x)$ and $\eta \doteq \sum_i \gamma_i^2$, so that $\hat\gamma = \kappa\sqrt{\eta}$. By Proposition 16, $\|\mathrm{d}\gamma_i\| \leqslant 5\sqrt{2}\gamma_i^2$ and then

$$\|\mathrm{d}\eta\| \leqslant 2\sum_i \gamma_i\|\mathrm{d}\gamma_i\| \leqslant 10\sqrt{2}\sum_i \gamma_1^3 \leqslant 10\sqrt{2}\eta^{3/2}.$$

Therefore

$$\|\mathrm{d}\hat\gamma\| \leqslant \|\mathrm{d}\kappa\|\sqrt{\eta} + \tfrac{1}{2}\kappa\eta^{-\frac{1}{2}}\|\mathrm{d}\eta\| \leqslant 5\hat\gamma^2 + 5\sqrt{2}\hat\gamma\sqrt{\eta} \leqslant 13\hat\gamma^2,$$

where we used $\kappa \geqslant 1$. □

3.3. **Yet another continuation algorithm.** We describe a continuation algorithm in the rigid solution variety and bound its complexity in terms of the integral of $\kappa\hat\gamma$ along the continuation path. It is the analogue of the μ^2 estimate of the classical theory, see §1.1. The approach proposed here differ from the usual treatment only in a more systematic use of derivatives. We assume $D \geqslant 2$ as otherwise we only have a linear system of equations to solve.

As we will see, it is often valuable not to compute $\hat\gamma$ but rather an easier to compute upper bound. So we formulate the algorithm in terms of a function $g : \mathcal{U}\times\mathbb{P} \to (0,\infty]$ that we can choose freely, as long as:

(i) $\hat\gamma \leqslant g$, on $\mathcal{U}\times\mathbb{P}$;
(ii) $\frac{1}{g}$ is C-Lipschitz continuous, for some $C \geqslant 1$.

The following proposition describes one continuation step. Observe that the conclusion (a) concerning the triple $(\mathbf{v},\eta,z')$ is similar to the hypothesis (ii) concerning $(\mathbf{u},\zeta,z)$, so that we can chain the continuation steps.

Proposition 18. *Let $\mathbf{u},\mathbf{v}\in\mathcal{U}$, let ζ be a zero of $\mathbf{u}\cdot F$, let $z\in\mathbb{P}$ and let $z' \doteq \mathcal{N}_{\mathbf{v}\cdot F}(z)$. For any positive real number $A \leqslant \frac{1}{15C}$, if*

(a) $d_{\mathbb{P}}(z,\zeta)g(\mathbf{u},\zeta) \leqslant A$,
(b) $d_{\mathcal{U}}(\mathbf{u},\mathbf{v})\kappa(\mathbf{u},z)g(\mathbf{u},z) \leqslant \frac{1}{4}A$,

then there exists a unique zero η of $\mathbf{v}\cdot F$ such that

(i) z is an approximate zero of $\mathbf{v}\cdot F$ with associate zero η,
(ii) $d_{\mathbb{P}}(z',\eta)g(\mathbf{v},\eta) \leqslant A$.

Proof. It suffices to prove that $d_{\mathbb{P}}(z,\eta)g(\mathbf{v},\eta) \leqslant 2A$. Indeed, since $2A \leqslant \frac{1}{6}$ and $\gamma \leqslant \hat\gamma \leqslant g$, it would imply by Theorem 10 that z is an approximate zero of $\mathbf{v}\cdot F$ and thus that $d_{\mathbb{P}}(z',\eta) \leqslant \frac{1}{2}d_{\mathbb{P}}(z,\eta)$ and the claim would follows.

Let $T \doteq d_{\mathcal{U}}(\mathbf{u},\mathbf{v})$. Consider a geodesic path $t\in[0,\infty) \mapsto \mathbf{v}_t \in \mathcal{U}$ such that $\|\dot{\mathbf{v}}_t\| = 1$ and $\mathbf{v}_T = \mathbf{v}$. Let η_t be the zero of $\mathbf{v}_t\cdot F$ obtained by continuation of the zero ζ. Let $p_t \doteq (\mathbf{v}_t,\eta_t)$ and let $\delta_t \doteq d_{\mathbb{P}}(z,\zeta_t)$. Let moreover $g_t \doteq g(\mathbf{v}_t,\eta_t)$, $\beta_t \doteq g_t\delta_t$ and $\kappa_t \doteq \kappa(\mathbf{v}_t,\eta_t)$. For readability, we drop the index t and denote derivative with respect to t with a dot.

We first observe that $\dot\delta \leqslant \kappa$, by Lemma 12, and that $\|\dot p\| \leqslant (1+\dot\delta^2)^{\frac12} \leqslant 2\kappa$ (using $\kappa \geqslant 1$). By Lemma 15, $\dot\kappa \leqslant 5\kappa g\|\dot p\| \leqslant 10\kappa^2 g$ and by the Lipschitz hypothesis on g, $\dot g \leqslant Cg^2\|\dot p\| \leqslant 2C\kappa g^2$. This implies that $\frac{\mathrm d}{\mathrm dt}\kappa g \leqslant 12C(\kappa g)^2$ and it follows, after integration, that

$$\kappa g \leqslant \frac{\kappa_0 g_0}{1-12Ct\kappa_0 g_0},$$

for any $t \geqslant 0$ such that $1-12Ct\kappa_0 g_0 > 0$. Next, we compute similarly that $\dot\beta \leqslant 2C\beta\kappa g + \kappa g$, and therefore, after integration,

$$\log\left(\frac{1+2C\beta}{1+2C\beta_0}\right) \leqslant \log\left(\frac{1}{1-12Ct\kappa_0 g_0}\right).$$

Exponentiating both sides leads to

$$\beta_t \leqslant \frac{\beta_0 + t\kappa_0 g_0}{1-12Ct\kappa_0 g_0}. \tag{17}$$

We now bound $\kappa_0 g_0$. As a function on $\mathcal U \times \mathbb P$, we compute that

$$\|\mathrm d(\kappa g)\| \leqslant (5+C)\kappa g^2 \leqslant 6C\kappa g^2, \tag{18}$$

or, equivalently, $\|\mathrm d\log(\kappa g)\| \leqslant 6Cg$. With the same kind of argument as above, we show that for any $w \in \mathbb P$ on a shortest path between z and ζ,

$$g(\mathbf u, w) \leqslant \frac{g(\mathbf u,\zeta)}{1-Cd_{\mathbb P}(\zeta,z)g(\mathbf u,\zeta)}. \tag{19}$$

After integrating the relation $\mathrm d\log(\kappa g) \leqslant 6Cg$ on a path from $(\mathbf u,\zeta)$ to $(\mathbf u, z)$, and bouding the right-hand side with (19), we obtain

$$\kappa_0 g_0 \leqslant \kappa(\mathbf u,z)g(\mathbf u,z)\exp\left(\frac{6Cd_{\mathbb P}(z,\zeta)g(\mathbf u,\zeta)}{1-Cd_{\mathbb P}(z,\zeta)g(\mathbf u,\zeta)}\right). \tag{20}$$

We multiply by $d_{\mathcal U}(\mathbf u,\mathbf v)$ both sides, use the hypotheses (a) and (b), and obtain

$$d_{\mathcal U}(\mathbf u,\mathbf v)\kappa_0 g_0 \leqslant \tfrac14 A\exp\left(\frac{6CA}{1-CA}\right) < \tfrac{5}{13}A,$$

where the last inequality follows from the hypothesis $CA \leqslant \frac{1}{15}$. Inequality (17) for $t = d_{\mathcal U}(\mathbf u,\mathbf v)$ gives

$$d_{\mathbb P}(z,\eta)g(\mathbf v,\eta) \leqslant \frac{A+\frac{5}{13}A}{1-12C\frac{5}{13}A} \leqslant 2A,$$

using again that $CA \leqslant \frac{1}{15}$. This concludes the proof. □

This leads to the procedure NC (Algorithm 2) which computes an approximate zero of a system $\mathbf u\cdot F$ given a zero of another system $\mathbf v\cdot F$ using a numerical continuation along a path $(\mathbf w_t)_{0\leqslant t\leqslant T}$ from $\mathbf v$ to $\mathbf u$.

Theorem 19. *On input F, $\mathbf u$, $\mathbf v$ and z, assuming that z is a zero of $\mathbf v\cdot F$, procedure NC ouputs an approximate zero of $\mathbf u\cdot F$ or loops forever.*

Moreover, the number of Newton iterations performed by the algorithm is at most

$$100C\int_0^T \kappa(\mathbf w_t,\zeta_t)g(\mathbf w_t,\zeta_t)\mathrm dt,$$

Algorithm 2. Homotopy continuation
Input. $F \in \mathcal{H}[n]$, $\mathbf{u}, \mathbf{v} \in \mathcal{U}$ and $z \in \mathbb{P}$
Precondition. z is a zero of $\mathbf{v} \cdot F$.
Output. $w \in \mathbb{P}^n$.
Postcondition. w is an approximate zero of $\mathbf{u} \cdot F$.

```
function NC(F, u, v, z)
    (w_t)_{0≤t≤T} ← a 1-Lipschitz continuous path from v to u
    t ← 1/(60C κ(w_0, z)g(w_0, z))
    while t < T do
        z ← N_{w_t}(z)
        t ← t + 1/(60C κ(w_t, z)g(w_t, z))
    end while
    return z
end function
```

where $(\mathbf{w}_t)_{0\leqslant t\leqslant T}$ is the continuation path chosen by the algorithm and ζ_t is the zero of $\mathbf{w}_t \cdot F$ obtained by continuation of z. In particular, if this integral is finite then the procedure terminates.

Proof. Let $t_0 \doteq 0$, let t_k be the value of t at the beginning of the kth iteration, let $z_0 \doteq z$ and let z_k be the value of z at the end of the kth iteration; namely

$$t_{k+1} \doteq t_k + \frac{1}{60C\,\kappa(\mathbf{w}_{t_k}, z_k)g(\mathbf{w}_{t_k}, z_k)} \tag{21}$$

$$z_{k+1} \doteq \mathcal{N}_{\mathbf{w}_{t_{k+1}}}(z_k). \tag{22}$$

Let K be the largest integer such that $t_K \leqslant T$ (or $K \doteq \infty$ if there is not). The output of the algorithm, if any, is z_K. Thanks to the Lipschitz hypothesis for $\mathbf{w}$, we have, for any $k \geqslant 0$,

$$d_{\mathcal{U}}(\mathbf{w}_{t_k}, \mathbf{w}_{t_{k+1}}) \leqslant t_{k+1} - t_k = (60C\kappa(\mathbf{w}_{t_k}, z_k)g(\mathbf{w}_{t_k}, z_k))^{-1}.$$

We apply repeatedly Proposition 18 with $A \doteq \frac{1}{15C}$ and we obtain that

$$d_{\mathbb{P}}(z_k, \zeta_{t_k})g(\mathbf{w}_{t_k}, \zeta_{t_k}) \leqslant A,$$

for any $k \geqslant 0$ (conclusion (ii) of Proposition 18 is used for hypothesis (a) at the next step, initialization is trivial since z is a zero). By Proposition 18 again, for any $k \geqslant 0$ and any $t \in [t_k, t_{k+1}]$, z_k is an approximate zero of $\mathbf{w}_t \cdot F$. In particular, z_K is an approximate zero of $\mathbf{w}_T \cdot F$. This proves the correctness of the algorithm.

Concerning the bound on the number of iterations, we first note that

$$\begin{aligned}\int_0^T \kappa_t g_t \mathrm{d}t &\geqslant \sum_{k=0}^{K-1} (t_{k+1} - t_k) \min_{t_k \leqslant s < t_{k+1}} \kappa_s g_s \\ &= \sum_{k=0}^{K-1} \frac{\min_{t_k \leqslant s < t_{k+1}} \kappa_s g_s}{60C\,\kappa(\mathbf{w}_{t_k}, z_k)g(\mathbf{w}_{t_k}, z_k)},\end{aligned} \tag{23}$$

where we use the notations of the proof of Proposition 18 applied to the path $(\mathbf{w}_t)$. We prove with the same techniques that for any $t_k \leqslant s < t_{k+1}$

$$\kappa_s g_s \geqslant \frac{\kappa_{t_k} g_{t_k}}{1 + 12C(t_{k+1} - t_k)\kappa_{t_k} g_{t_k}}.$$

We also check, similarly to (20), that for any $t_{k-1} \leqslant s < t_k$,

$$\begin{aligned}\kappa_{t_k} g_{t_k} &\geqslant \kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k) \exp\left(-\frac{6Cd(z_k, \zeta_{t_k}) g(\mathbf{w}_{t_k}, \zeta_k)}{1 + Cd(z_k, \zeta_{t_k}) g(\mathbf{w}_{t_k}, \zeta_{t_k})}\right) \\ &\geqslant \kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k) \exp\left(-\frac{6CA}{1 + CA}\right) \\ &= \kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k) \exp(-\tfrac{3}{8}).\end{aligned}$$

Therefore, for any $t_k \leqslant s < t_{k+1}$,

$$\begin{aligned}\kappa_s g_s &\geqslant \frac{\kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k) \exp(-\frac{3}{8})}{1 + 12C(t_{k+1} - t_k)\kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k) \exp(-\frac{3}{8})} \\ &\geqslant \tfrac{3}{5} \kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k), \text{ using the value for } t_{k+1} - t_k \ (21),\end{aligned}$$

and then

$$\frac{\min_{t_k \leqslant s < t_{k+1}} \kappa_s g_s}{60C\, \kappa(\mathbf{w}_{t_k}, z_k) g(\mathbf{w}_{t_k}, z_k)} \geqslant \frac{1}{100C}.$$

Therefore, by (23), $\int_0^T \kappa_t g_t \mathrm{d}t \geqslant \frac{1}{100C} K$. □

We have some degrees of freedom but also some constraints in the choice of the path $(\mathbf{w}_t)_{0 \leqslant t \leqslant T}$ from $\mathbf{v}$ to $\mathbf{u}$. To be used by the numerical continuation algorithm, it must be 1-Lipschitz continuous. Moreover, in order to perform the average analysis, we require an extra hypothesis:

(∗) the path $(\mathbf{v}^{-1}\mathbf{w}_t)_{0 \leqslant t \leqslant T}$ (from $\mathbf{1}_{\mathcal{U}}$ to $\mathbf{v}^{-1}\mathbf{u}$) depends only on $\mathbf{v}^{-1}\mathbf{u}$.

An obvious choice of the path between $\mathbf{v}$ and $\mathbf{u}$ is the shortest one: we write $\mathbf{v}^{-1}\mathbf{u} = (\exp(A_1), \dots, \exp(A_n))$ for some skew-Hermitian matrices $A_1, \dots, A_n$ and define

$$\mathbf{w}_t \doteq \mathbf{v}\left(\exp(\tfrac{t}{T}A_1), \dots, \exp(\tfrac{t}{T}A_n)\right) \text{ for } 0 \leqslant t \leqslant T$$

where $T \doteq (\sum_i \|A_i\|_u^2)^{\frac{1}{2}}$. We can always choose the matrices A_i such that $T \leqslant (n+1)\frac{\pi}{\sqrt{2}}$. Naturally, it may not be convenient to compute matrix logarithms and exponentials, especially for the complexity analysis in the BSS model. In §4.2.3 we will see paths that are cheaper to compute.

3.4. **A randomized algorithm.** We now have in our hands everything we need to mimic Beltrán–Pardo algorithm in the rigid setting: a continuation algorithm with an integral estimate for its complexity and a recipe to sample points in the solution variety with the appropriate distribution. This leads to Algorithm 3 (Solve): for finding a zero of $\mathbf{u} \cdot F$, we first sample a random element $(\mathbf{v}, \eta)$ in $\mathcal{V}$, with Algorithm 1 (Sample), and then perform the numerical continuation along a path in $\mathcal{U}$ from $\mathbf{v}$ to $\mathbf{u}$.

Let T be the length of the path chosen in $\mathrm{NC}(\mathbf{u}, \mathbf{v}, \eta)$ and let K be the number of continuation steps performed in $\mathrm{NC}(\mathbf{u}, \mathbf{v}, \eta)$.

Theorem 20. *If $\mathbf{u} \in \mathcal{U}$ is uniformly distributed and $(\mathbf{v}, \eta) \sim \rho_{\mathrm{std}}$ then*

$$\mathbb{E}[K] \leqslant 100C \cdot \mathbb{E}[T] \cdot \mathbb{E}\left[\kappa(\mathbf{v} \cdot F, \eta) g(\mathbf{v} \cdot F, \eta)\right].$$

Algorithm 3. An analogue of Beltrán-Pardo algorithm in the rigid setting
Input. Homogeneous polynomial system $F \in \mathcal{H}[n]$ and $\mathbf{u} \in \mathcal{U}$
Output. $z \in \mathbb{P}$.
Postcondition. z is an approximate zero of $\mathbf{u} \cdot F$.

function Solve(F, $\mathbf{u}$)
  $(\mathbf{v}, \eta) \leftarrow$ Sample($\mathcal{V}_F$)
  return NC $(\mathbf{u}, \mathbf{v}, \eta)$
end function

Proof. Let $(\mathbf{v}, \eta) \sim \rho_{\mathrm{std}}$ be the pair computed by "Sample" in Algorithm 3. Let $(\mathbf{w}_t)_{0 \leqslant t \leqslant T}$ be the path from $\mathbf{v}$ to $\mathbf{u}$ chosen in NC. We assume that $\mathbf{v}^{-1}\mathbf{w}_t$ is a function of $\mathbf{v}^{-1}\mathbf{u}$. We first note that $\mathbf{v}$ and $\mathbf{v}^{-1}\mathbf{u}$ are independent and uniformly distributed in $\mathcal{U}$, because the Jacobian of the diffeomorphism

$$(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{U} \mapsto (\mathbf{v}, \mathbf{v}^{-1}\mathbf{u}) \in \mathcal{U} \times \mathcal{U}$$

is constant. Secondly, by hypothesis, for any $0 \leqslant s \leqslant 1$, the random variable $\mathbf{v}^{-1}\mathbf{w}_{Ts}$ depends only on $\mathbf{v}^{-1}\mathbf{u}$, so it is independent from $\mathbf{v}$. Therefore $\mathbf{w}_{Ts}$, which equals $\mathbf{v}(\mathbf{v}^{-1}\mathbf{w}_{Ts})$, is uniformly distributed and independent from $\mathbf{v}^{-1}\mathbf{u}$.

Let ζ_t be the zero of $\mathbf{w}_t \cdot F$ obtained by continuation of the zero η of $\mathbf{v} \cdot F$. Since ζ is uniformly distributed among the zeros of $\mathbf{v}$, it follows that ζ_t is uniformly distributed among the zeros of $\mathbf{w}_t$, because the numerical continuation induces a bijective correspondance between the two sets of zeros, almost surely. Therefore, for any $0 \leqslant s \leqslant 1$, $(\mathbf{w}_{Ts}, \zeta_{Ts})$ is a ρ_{std} distributed random variable independent from $\mathbf{v}^{-1}\mathbf{u}$, and in particular, independent from T.

Together with Theorem 19 (and the change of variable $t = Ts$), this implies

$$\begin{aligned}
\mathbb{E}[K] &\leqslant \mathbb{E}\left[100C \int_0^1 \kappa(\mathbf{w}_{Ts}, \zeta_{Ts}) g(\mathbf{w}_{Ts}, \zeta_{Ts}) T \mathrm{d}s\right] \\
&= 100C \int_0^1 \mathbb{E}\left[\kappa(\mathbf{w}_{Ts}, \zeta_{Ts}) g(\mathbf{w}_{Ts}, \zeta_{Ts}) T\right] \mathrm{d}s \\
&= 100C\, \mathbb{E}\left[\kappa(\mathbf{v}, \eta) g(\mathbf{v}, \eta)\right] \mathbb{E}[T],
\end{aligned}$$

which is the claim. □

4. Average complexity for dense polynomial systems

We now apply the previous results to the resolution of a random Gaussian system. The polynomial system $F \in \mathcal{H}[n]$ fixed in §3 is now a standard normal variable (a Gaussian random vector with mean 0 and covariance matrix Id). Since $\mathcal{U}$ acts isometrically on $\mathcal{H}[n]$, it follows that $\mathbf{u}^{-1} \cdot F$ has the same distribution as F. More precisely, if $\mathbf{u} \in \mathcal{U}$ is uniformly distributed and independent from F, then $\mathbf{u}^{-1} \cdot F$ is a standard normal variable that is independent from $\mathbf{u}$. Therefore, we will be able to use the previous results, and especially the average analysis (Theorem 20), to find a zero of $\mathbf{u} \cdot (\mathbf{u}^{-1} \cdot F)$, that is a zero of F.

In §4.1, we describe the function $\hat{\gamma}_{\mathrm{Frob}}$ that we will use for the numerical continuation (in the role of g). Next, in §4.2, we discuss the computational model and the

construction of paths in $\mathcal{U}$. And lastly, in §4.3, we perform the average analysis for Gaussian systems.

4.1. **An efficiently computable variant of γ.**

4.1.1. *Norms of a multilinear map.* Let E and F be Hermitian spaces and let $h : E^k \to F$ be a multilinear map, that is equivalent to the data of a linear map $E \to (E \to \cdots (E \to F))$. The *operator norm* of h is defined by

$$|||h||| \doteq \max\{\|h(x_1,\dots,x_k)\| \mid x_1,\dots,x_k \in \mathbb{S}(E)\}.$$

This is the operator norm on $E \to (E \to \cdots (E \to F))$ where each $E \to *$ is recursively endowed with the operator norm.

We also define the *Frobenius norm* of h as

$$\|h\|_{\mathrm{Frob}} \doteq \left(\sum_{1\leqslant i_1,\dots,i_k\leqslant n} \|a_{i_1,\dots,i_k}\|^2\right)^{\frac{1}{2}}, \tag{24}$$

where the $a_{i_1,\dots,i_k}$ are the coefficients of h in some unitary basis $e_1,\dots,e_{\dim E}$ of E, that is

$$a_{i_1,\dots,i_k} \doteq h(e_{i_1},\dots,e_{i_k}) \in F,$$

for $1 \leqslant i_1,\dots,i_k \leqslant n$. This norm may also be defined as the usual Frobenius on $E \to (E \to \cdots (E \to F))$ where each $E \to *$ is given the Frobenius norm.

Lemma 21. *For any multilinear map $h : E^k \to F$, $|||h||| \leqslant \|h\|_{\mathrm{Frob}} \leqslant (\dim E)^{\frac{k}{2}} \|h\|_M$.*

Proof. This is better seen if we consider h as a map $E \to (E \to \cdots (E \to F))$ as above. The claim follows from an induction on k and the usual comparison between the Frobenius and the operator norm. □

Lemma 22. *For any homogeneous polynomial $p(x_0,\dots,x_n)$ of degree k,*

$$\|p\|_W = \left\|\tfrac{1}{k!}\mathrm{d}_0^k p\right\|_{\mathrm{Frob}}.$$

Proof. For any $0 \leqslant i_1,\dots,i_k \leqslant n$,

$$\begin{aligned}\frac{1}{k!}\mathrm{d}_0^k p(e_{i_1},\dots,e_{i_k}) &= \frac{1}{k!}\frac{\partial^k p}{\partial x_{i_1}\cdots\partial x_{i_k}}\bigg|_{x=0}\\ &= \frac{j_0!\cdots j_n!}{k!}c_{j_0,\dots,j_n},\end{aligned}$$

where j_m is the number of indices i_* that are equal to m and where $c_{j_0,\dots,j_n}$ is the coefficient of $x_0^{j_0}\cdots x_n^{j_n}$ in p. There are exactly $\frac{k!}{j_0!\cdots j_n!}$ k-uples i_* that lead to a given $(n+1)$-uple j_*. Therefore,

$$\left\|\tfrac{1}{k!}\mathrm{d}_0^k p(e_{i_1},\dots,e_{i_k})\right\|_{\mathrm{Frob}}^2 = \sum_{j_0+\cdots+j_n=k} \frac{j_0!\cdots j_n!}{k!}|c_{j_0,\dots,j_n}|^2$$

and this is exactly $\|p\|_W^2$. □

4.1.2. *The γ number with Frobenius norms.* For a homogeneous polynomial $f : \mathbb{C}^{n+1} \to \mathbb{C}$, we define

$$\gamma_{\mathrm{Frob}}(f,z) \doteq \begin{cases} \|z\| \sup_{k\geqslant 2} \frac{1}{k!} \|\mathrm{d}_z f\|^{-1} \left\|\mathrm{d}^k_z f\right\|^{\frac{1}{k-1}}_{\mathrm{Frob}} & \text{if } \mathrm{d}_z f \text{ is non zero,} \\ \infty & \text{otherwise,} \end{cases}$$

and for a homogeneous polynomial system $F = (f_1, \dots, f_r) \in \mathcal{H}[r]$, we define

$$\hat{\gamma}_{\mathrm{Frob}}(F,z) \doteq \kappa(F,z) \left(\gamma_{\mathrm{Frob}}(f_1,z)^2 + \cdots + \gamma_{\mathrm{Frob}}(f_r,z)^2\right)^{\frac{1}{2}}.$$

Compare with the definion of γ and $\hat{\gamma}$, §2.5.

Lemma 23. *For any $F \in \mathcal{H}[r]$, $\hat{\gamma}(F,z) \leqslant \hat{\gamma}_{\mathrm{Frob}}(F,z) \leqslant (n+1)\hat{\gamma}(F,z)$.*

Proof. This follows from Lemma 21. □

In order to use Algorithm with $\hat{\gamma}_{\mathrm{Frob}}$ we must check a Lipschitz condition.

Lemma 24. *On $\mathcal{U} \times \mathbb{P}$, the function $(\mathbf{u}, z) \mapsto 1/\hat{\gamma}_{\mathrm{Frob}}(\mathbf{u}, z)$ is 13-Lipschitz continous.*

Proof. If is enough to show that for a polynomial $f \in \mathcal{H}[1]$, the function $x \in \mathbb{P} \mapsto \gamma_{\mathrm{Frob}}(f,x)$ is 5-Lipschitz continuous. Then the claim follows as in Lemma 17.

Let $(x_t)_{0\leqslant t\leqslant 1}$ be a differentiable path in $\mathbb{C}^{n+1}$. Let $a_t \doteq \|\mathrm{d}_{x_t} f\|^{-1}$ and $(B_k)_t \doteq \frac{1}{k!} a_t d^k_{x_t} f$. From now on, we drop the index t and denote the derivative with respect to t with a dot or with $\frac{\mathrm{d}}{\mathrm{d}t}$. We compute that $\dot{a} = -a\langle a\mathrm{d}^2 f(\dot{x}), a\mathrm{d}f\rangle$, which implies that

$$|\dot{a}| \leqslant 2a\gamma(f,x)\|\dot{x}\|.$$

We note that $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{d}^k_x f = \mathrm{d}^{k+1}_x f(\dot{x})$, therefore,

$$\dot{B}_k = -a\langle a\mathrm{d}^2 f(\dot{x}), a\mathrm{d}f\rangle B_k + (k+1)B_{k+1}(\dot{x}),$$

and it follows, since for any l, $\|B_l\|_{\mathrm{Frob}} \leqslant \gamma_{\mathrm{Frob}}(H,x)^{l-1}$, that

$$\begin{aligned} \left|\tfrac{\mathrm{d}}{\mathrm{d}t}\|B_k\|_{\mathrm{Frob}}\right| &\leqslant \|\dot{B}_k\|_{\mathrm{Frob}} \\ &\leqslant 2\gamma(f,x)\|B_k\|_{\mathrm{Frob}}\|\dot{x}\| + (k+1)\|B_{k+1}\|_{\mathrm{Frob}}\|\dot{x}\| \\ &\leqslant (3+k)\gamma_{\mathrm{Frob}}(h,x)^k\|\dot{x}\|. \end{aligned}$$

It follows that

$$\left|\tfrac{\mathrm{d}}{\mathrm{d}t}\|B_k\|^{\frac{1}{k-1}}_{\mathrm{Frob}}\right| = \frac{1}{k-1}\|B_k\|^{\frac{1}{k-1}-1}_{\mathrm{Frob}} \left|\tfrac{\mathrm{d}}{\mathrm{d}t}\|B_k\|_{\mathrm{Frob}}\right| \leqslant \frac{3+k}{k-1}\frac{\gamma^{k+1}_{\mathrm{Frob}}}{\|B_k\|_{\mathrm{Frob}}}\|\dot{x}\|.$$

By definition, γ_{Frob} is the supremum of all $\|B_k\|^{1/(k-1)}_{\mathrm{Frob}}$ $(k \geqslant 2)$, finitely many of which are non zero, so at a given time t, there is some k such that $\gamma_{\mathrm{Frob}} = \|B_k\|^{1/(k-1)}_{\mathrm{Frob}}$ in a (one sided) neighborhood of t. Therefore $\dot{\gamma}_{\mathrm{Frob}} = \frac{\mathrm{d}}{\mathrm{d}t}\|B_k\|^{1/(k-1)}_{\mathrm{Frob}}$ and the computation above shows that

$$|\dot{\gamma}_{\mathrm{Frob}}| \leqslant \frac{3+k}{k-1}\gamma^2_{\mathrm{Frob}}\|\dot{x}\| \leqslant 5\gamma^2_{\mathrm{Frob}}.$$

Consequently, $\left|\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\gamma_{\mathrm{Frob}}}\right| \leqslant 5$ and the claim follows. □

4.2. **Implementation details, complexity.**

4.2.1. *Computational model.* We use the Blum–Shub–Smale model (Blum et al. 1989) extended with a "6th type of node", as did Shub and Smale (1996). Unlike Shub and Smale, we will apply it to univariate (or rather homogeneous bivariate) polynomials only. A node of this type has the following behavior. If it is given as input a homogeneous polynomial $f \in \mathbb{C}[x, y]$ and an approximate zero $z \in \mathbb{P}^1$ of f, with associated zero ζ, it outputs ζ. In any other case, it fails. There is no need to specify how it fails because we will make sure that this will not happen.

From the practical point of view, given a point z which approximates a zero ζ of a homogeneous polynomial $f \in \mathbb{C}[x, y]$, one can refine the approximation to obtain $d_{\mathbb{P}}(z, \zeta) \leqslant \varepsilon$ in $\log_2 \log_2 \frac{\pi}{\varepsilon}$ Newton's iterations. For most practical purpose, this looks like infinite precision. In that sense, the 6th type of node does not add much power.

Do we really need a 6th type of node? The continuation method proposed here uses a start system defined in terms of the zeros of some homogeneous bivariate polynomials. Naturally, the algorithm would also work with approximate zeros only. However, if we do it this way, then the distribution of the start system is not easily described, it is only close to a nice distribution. I showed (Lairez 2017) how to deal with the complexity analysis in an analogue situation but it is too technical an argument for the little value it adds.

Interval arithmetic gives another way to remove the need for this extra type of node: wherever an exact zero is expected, we use bounding boxes instead and perform the subsequent operations with interval arithmetic. If the precision happens to be insufficient, we refine the bounding boxes with Newton's iteration and start over the computation. The convergence of Newton's iteration is so fast that even with naive estimations of the numerical stability, the number of start over will be moderate. However, this is no less technical to formalize.

For convenience, we will also assume the ability to compute fractional powers of a positive real number at unit cost. This will allow us to compute Hermitian norms and the numbers γ_{Frob} and $\hat{\gamma}_{\mathrm{Frob}}$ exactly.

4.2.2. *Computation of $\hat{\gamma}_{\mathrm{Frob}}$, cost of a continuation step.* The reason for introducing $\hat{\gamma}_{\mathrm{Frob}}$ is that we can compute it with low complexity. By contrast, computing $\hat{\gamma}$ is NP-hard because it involves the computation of the spectral norm of symmetric multilinear maps, and there is not either a polynomial-time approximation scheme, unless P = NP (Hillar and Lim 2013, Theorem 10.2). Beware though, a too naive algorithm for computing γ_{Frob} requires $\Omega(N^2)$ operations.

Proposition 25. *Given a homogeneous polynomial $f \in H_d$ and $z \in \mathbb{P}$, one can compute $\gamma_{\mathrm{Frob}}(f, z)$ with $O(nd^2 \dim H_d)$ operations, as $\dim H_d \to \infty$.*

Proof. The main task is to compute $\left\|\mathrm{d}_z^k f\right\|_{\mathrm{Frob}}$ for all $2 \leqslant k \leqslant d$. Let g be the shifted polynomial $g : x \mapsto f(z + x)$. Obviously, $\mathrm{d}_0^k g = \mathrm{d}_z^k f$. According to Lemma 22, computing $\|\mathrm{d}_0^k g\|_{\mathrm{Frob}}$ boils down to computing the homogeneous component of degree k of g. Let S denote the *size* of g, that is the number of coefficients in a dense non-homogeneous polynomial of degree d in $n + 1$ variables. He have $S =$

$\binom{n+d+1}{d} \leqslant d \dim H_d$. Once we have computed g in the monomial basis, we may obtain $\gamma_{\mathrm{Frob}}(f, z)$ in $O(S)$ operations.

To compute g, we shift the variables one after the other. To compute $f(x_0 + z_0, x_1, \dots, x_n)$, we write

$$f = \sum_{i_1+\cdots+i_n \leqslant d} p_{i_1,\dots,i_n}(x_0) x_1^{i_1} \cdots x_n^{i_n},$$

where the $p_{i_1,\dots,i_n}(x_0)$ are polynomials of degree at most d; there are at most S of them and computing them requires no arithmetic operation. One can compute $p_{i_1,\dots,i_n}(x_0 + z_0)$ with $O(d^2)$ operations with a naive algorithm. Note that we can do this with only $d^{1+o(1)}$ operations using fast evaluation and interpolation algorithms (Bostan et al. 2017; von zur Gathen and Gerhard 1999). We recover $f(x_0 + z_0, x_1, \dots, x_n)$ in $O(S)$ operations. We repeat this shift operation for each one of the $n + 1$ variables and this gives the claim. □

Corollary 26. *In Algorithm 2 with $g = \hat{\gamma}_{\mathrm{Frob}}$, one continuation step can be performed in $O(nD^2N)$ operations.*

Proof. A step boils down to one evaluation of $\hat{\gamma}_{\mathrm{Frob}}$ and κ and one Newton's iteration. We can compute κ within a factor of 2 (which is good enough for our purpose) in $O(n^3)$ operations, using a tridiagonalization with Householder's reflections and a result by Kahan (1966). By Proposition 25, we can compute $\hat{\gamma}_{\mathrm{Frob}}$ with $O(nD^2N)$ operations (within a factor 2). Note that $N = \sum_i \dim H_{d_i}$. And Newton's iteration costs $O(n^3 + N) = O(nN)$ operations. □

4.2.3. *Continuation paths in the unitary group.* While geodesics in $\mathcal{U}$ are a natural choice for continuation paths, see §3.4, they are not easy to compute in the BSS model. We can describe more elementary continuation paths using Householder's reflections. For $l \in \mathbb{P}$ and $\theta \in \mathbb{R}$, let $R(l, \theta) \in U(n + 1)$ be the unique map such that $R(l, \theta)|_l = e^{i\theta} \mathrm{id}_l$ and $R(l, \theta)|_{l^\perp} = \mathrm{id}_{l^\perp}$.

Given a unitary matrix $v \in U(n + 1)$, the procedure of Householder (1958) (with the necessary changes in the complex case) decomposes u as

$$v = e^{i\alpha} R(l_1, \pi) \cdots R(l_n, \pi).$$

And so one can define the 1-Lipshitz continuous path

$$w_t \doteq e^{i\frac{t}{\tau}\alpha} R(l_1, \tfrac{t}{\tau}\pi) \cdots R(l_n, \tfrac{t}{\tau}\pi),$$

where $\tau^2 \doteq \frac{1}{2}\left(\alpha^2 + n\pi^2\right) \leqslant \frac{n+1}{2}\pi^2$. Given $\mathbf{u}, \mathbf{v} \in \mathcal{U}$, we define a 1-Lipshitz continuous path $(\mathbf{w}_t)_{t \geqslant 0}$ from $\mathbf{1}_{\mathcal{U}}$ to $\mathbf{v}^{-1}\mathbf{u}$, component-wise with the method above. It reaches $\mathbf{v}^{-1}\mathbf{u}$ at $t = \sqrt{\frac{n(n+1)}{2}}\pi < 4n$, the diameter of $\mathcal{U}$. To compute w_t on a BSS machine, we can replace the trigonometric functions with any other functions parametrizing the circle.

For a given t, the cost of computating of $\mathbf{w}_t$ is dominated by the multiplication of n matrices, that is $O(n^4)$ operations.

4.3. **Gaussian random systems.** Let $F \in \mathcal{H}[r]$ be a standard normal variable, let $\zeta \in \mathbb{P}$ be a random uniformly distributed zero of F and let $\bar{\zeta} \in \mathbb{C}^{n+1}$ be a random uniformly distributed vector such that $\|\bar{\zeta}\| = 1$ and $[\bar{\zeta}] = \zeta$.

Let $R_\zeta \subset \mathcal{H}[m]$ be the orthogonal complement of the subspace of all $G \in \mathcal{H}[r]$ such that $G(\zeta) = 0$ and $\mathrm{d}_\zeta G = 0$. Let $r_\zeta : \mathcal{H}[r] \to R_\zeta$ denote the orthogonal projection.

Theorem 27 (Beltrán and Pardo 2011, Theorem 7)**.** *With the notations above, let*

$$M \doteq \operatorname{diag}(\sqrt{d_1}, \dots, \sqrt{d_r})^{-1} \mathrm{d}_{\bar{\zeta}} F.$$

We have:

(i) $M\zeta = 0$*;*
(ii) $M \doteq \operatorname{diag}(\sqrt{d_1}, \dots, \sqrt{d_r})^{-1} \mathrm{d}_\zeta F$ *is a random Gaussian matrix;*
(iii) given $\bar{\zeta}$*,* $r_\zeta(F)$ *is a standard normal variable in* R_ζ *that is independent from* M*.*

Note that this result is proved by Beltrán and Pardo only when $r = n$ (see also Bürgisser and Cucker 2013, Chap. 17). The proof in the underdetermined case $r < n$ follows exactly the same lines. Alternatively, one can deduce the underdetermined case from the well-determined case $r = n$ by appending to the system F a number of $n - r$ random independent Gaussian linear equations. In particular, when $r = 1$, we have the following corollary.

Corollary 28. *Let* $f \in H_d$ *be a standard normal variable and let* ζ *be a random uniformly distributed point in* $\{z \in \mathbb{P} \mid f(z) = 0\}$ *and let* $\bar{\zeta}$ *be a random uniformly distributed vector such that* $\|\bar{\zeta}\| = 1$ *and* $[\bar{\zeta}] = \zeta$*. Then*

(i) $\frac{1}{\sqrt{d}} \mathrm{d}_\zeta f$ *is a standard normal variable in* $(\mathbb{C}^{n+1})^*$*;*
(ii) given ζ*,* $r_\zeta(f)$ *is a standard normal variable in* R_ζ *that is independent from* $\mathrm{d}_\zeta f$*.*

Lemma 29. *Let* $f \in H_d$ *and* $\zeta = [1 : 0 : \cdots : 0]$*. We write*

$$f = \sum_{i=0}^{d} x_0^{d-i} g_i(x_1, \dots, x_n),$$

for some uniquely determined homogeneous polynomials $g_0, \dots, g_d$ *of degree* $0, \dots, d$ *respectively. For any* $k \geqslant 2$*,*

$$\left\| \tfrac{1}{k!} \mathrm{d}_\zeta^k f \right\|_{\mathrm{Frob}}^2 \leqslant \binom{d}{k} \sum_{l=0}^{k} \binom{d-l}{k-l} \left\| x_0^{d-l} g_l \right\|_W^2 .$$

Proof. Let $\tilde{f} \doteq f(x_0 + 1, x_1, \dots, x_n)$ and let $\tilde{f}_{(k)}$ be the homogeneous component of degree k of $\tilde{f}$. We compute that

$$\tilde{f}_{(k)} = \sum_{l=0}^{d} \left[(x_0 + 1)^{d-l} g_l \right]_{(k)} = \sum_{l=0}^{k} \binom{d-l}{k-l} x_0^{k-l} g_l.$$

The terms of the sum are orthogonal for Weyl's inner product, and moreover $\left\|\frac{1}{k!}\mathrm{d}_\zeta^k f\right\|_{\mathrm{Frob}} = \left\|\tilde{f}_{(k)}\right\|_W$ by Lemma 22, therefore

$$\left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f\right\|_{\mathrm{Frob}}^2 = \sum_{l=0}^{k} \binom{d-l}{k-l}^2 \left\|x_0^{k-l} g_l\right\|_W^2 .$$

Looking closely at the definition of Weyl's inner product reveals that

$$\left\|x_0^{k-l} g_l\right\|_W^2 = \frac{\binom{d}{d-l}}{\binom{k}{k-l}} \left\|x_0^{d-l} g_l\right\|_W^2 = \frac{\binom{d}{k}}{\binom{d-l}{k-l}} \left\|x_0^{d-l} g_l\right\|_W^2 ,$$

and the claim follows. □

Lemma 30. *Let $f \in H_d$ be a standard normal variable and $\zeta \in \mathbb{P}$ be a uniformly distributed zero of f. For any $k \geqslant 2$,*

$$\mathbb{E}\left[\|\mathrm{d}_\zeta f\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f_i\right\|_{\mathrm{Frob}}^2\right] \leqslant \frac{1}{nd}\binom{d}{k}\binom{d+n}{k} \leqslant \left(\tfrac{1}{4}d^2(d+n)\right)^{k-1} .$$

Proof. We can choose (random) coordinates such that $\zeta = [1 : 0 : \cdots : 0]$. Let $g_0, \dots, g_d$ be the polynomials defined in Lemma 29. To begin with, we describe the distribution of the random polynomials $g_0, \dots, g_d$. The polynomial g_0 (of degree 0) is simply zero because ζ is a zero of f. The second one g_1 has degree 1, and as a linear form, it is equal to $\mathrm{d}_\zeta f$. According to Theorem 27(ii), it is a standard normal variable after multiplication by $d^{-\frac{1}{2}}$. In particular $d\|g_1\|_W^{-2}$ is the inverse of a chi-squared variable with $2n+2$ degree of freedom, therefore

$$\mathbb{E}\left[\|g_1\|_W^{-2}\right] = \frac{1}{2nd}.$$

For $l \geqslant 2$, the polynomial $x_0^{d-l} g_l$ is the orthogonal projection of $R_\zeta(f)$ on the subspace of all polynomials of the form $x_0^{d-l} p(x_1, \dots, x_n)$. By Theorem 27(iii), $R_\zeta(f)$ is a standard normal variable (given ζ), thus $x_0^{d-l} g_l$ is also a standard normal variable in the appropriate subspace. Moreover $g_1, \dots, g_d$ are independent. Therefore, for any $l \geqslant 2$,

$$\mathbb{E}\left[\left\|x_0^{d-l} g_l\right\|_W^2\right] = \dim_{\mathbb{R}}\left\{x_0^{d-l} p(x_1, \dots, x_n) \in H_d\right\} = 2\binom{n-1+l}{l}.$$

Note also that $\|g_1\|_W^{-2}\left\|x_0^{d-l} g_1\right\|^2 = \frac{1}{d}$. It follows, by Lemma 29, that

$$\begin{aligned}
\mathbb{E}\left[\|\mathrm{d}_\zeta f\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f\right\|_{\mathrm{Frob}}^2\right] &\leqslant \sum_{l=0}^{k} \mathbb{E}\left[\binom{d}{k}\sum_{l=0}^{k}\binom{d-l}{k-l}\|g_1\|_W^{-2}\left\|x_0^{d-l} g_l\right\|_W^2\right] \\
&\leqslant \binom{d}{k}\left(\binom{d-1}{k-1}\frac{1}{d} + \sum_{l=2}^{k}\binom{d-l}{k-l}\mathbb{E}\left[\|g_1\|_W^{-2}\right]\mathbb{E}\left[\|x_0^{d-l} g_l\|_W^2\right]\right) \\
&\leqslant \binom{d}{k}\left(\binom{d-1}{k-1}\frac{1}{d} + \frac{1}{nd}\sum_{l=2}^{k}\binom{d-l}{k-l}\binom{n-1+l}{l}\right) \\
&\leqslant \frac{1}{nd}\binom{d}{k}\sum_{l=0}^{k}\binom{d-l}{k-l}\binom{n-1+l}{l} \\
&= \frac{1}{nd}\binom{d}{k}\binom{d+n}{k}.
\end{aligned}$$

To check the binomial identity $\sum_{l=0}^{k}\binom{d-l}{k-l}\binom{n-1+l}{l}=\binom{d+n}{k}$, we remark that $\binom{d-l}{k-l}$ counts the number of monomials of degree $k-l$ in $d-k+1$ variables while $\binom{n+l}{l}$ counts the number of monomials of degree l in $n+1$ variables. Therefore, the sum over l counts the number of monomials of degree k in $(d-k+1)+(n+1)$ variables, that is $\binom{d+n}{k}$.

Concerning the second inequality, we remark that the maximum value, of $\left(\frac{1}{nd}\binom{d}{k}\binom{d+n}{k}\right)^{\frac{1}{k-1}}$ with $k\geqslant 2$, is reached for $k=2$. That is, for any $k\geqslant 2$,

$$\begin{aligned}\frac{1}{nd}\binom{d}{k}\binom{d+n}{k}&\leqslant\left(\frac{1}{nd}\binom{d}{2}\binom{d+n}{2}\right)^{k-1}.\\&\leqslant\left(\tfrac{1}{4}(d-1)(d+n)\left(\frac{d-1}{n}+1\right)\right)^{\frac{1}{k-1}},\end{aligned}$$

which leads to the claim. □

Lemma 31. *With the same notations as Lemma 30,*

$$\mathbb{E}\left[\gamma_{\mathrm{Frob}}(f,\zeta)^2\right]\leqslant\tfrac{1}{4}d^3(d+n).$$

Proof. We bound

$$\begin{aligned}\mathbb{E}\left[\gamma_{\mathrm{Frob}}(f,\zeta)^2\right]&\leqslant\sum_{k=2}^{d}\mathbb{E}\left[\left(\|\mathrm{d}_\zeta f\|^{-1}\left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f_i\right\|_{\mathrm{Frob}}\right)^{\frac{2}{k-1}}\right]\\&\leqslant\sum_{k=2}^{d}\mathbb{E}\left[\left(\|\mathrm{d}_\zeta f\|^{-2}\left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f_i\right\|_{\mathrm{Frob}}^2\right)\right]^{\frac{1}{k-1}}&&\text{by Jensen's inequality,}\\&\leqslant\sum_{k=2}^{d}\tfrac{1}{4}d^2(d+n)&&\text{by Lemma 30,}\\&\leqslant\tfrac{1}{4}d^3(d+n),\end{aligned}$$

and this is the claim. □

Theorem 32. *If $F\in\mathcal{H}[n]$ is a standard normal variable and $\zeta\in\mathbb{P}$ is a random uniformly distributed zero of F, then*

$$\mathbb{E}\left[\kappa(F,\zeta)\hat{\gamma}_{\mathrm{Frob}}(F,\zeta)\right]\leqslant 3n^4D^2.$$

Proof. Let $\mathbf{u}\in\mathcal{U}$ be a random uniformly distributed variable. We first remark that $\mathbf{u}\cdot F$ and F have the same distribution, because of the unitary invariance of the Gaussian distribution. So it suffices to bound $\mathbb{E}\left[\kappa(\mathbf{u}\cdot F,\eta)\hat{\gamma}_{\mathrm{Frob}}(\mathbf{u}\cdot F,\eta)\right]$, where η is a uniformly distributed zero of $\mathbf{u}\cdot F$.

From the definition of $\hat{\gamma}_{\mathrm{Frob}}(\mathbf{u}\cdot F,\eta)$ and the unitary invariance of γ_{Frob}, we have

$$\begin{aligned}&\kappa(\mathbf{u}\cdot F,\eta)\hat{\gamma}_{\mathrm{Frob}}(\mathbf{u}\cdot F,\eta)\\&\qquad=\kappa(\mathbf{u}\cdot F,\eta)^2\left(\gamma_{\mathrm{Frob}}(f_1,u_1^*\eta)^2+\cdots+\gamma_{\mathrm{Frob}}(f_r,u_r^*\eta)^2\right)^{\frac{1}{2}}.\end{aligned}$$

By Theorem 7, $u_1^*\eta,\dots,u_n^*\eta$ are independent and uniformly distributed in $V(f_1),\dots,V(f_n)$ and $\kappa(\mathbf{u}\cdot F,\eta)$, which depends only on $L(\mathbf{u}\cdot F,\eta)$, is independent with them.

Therefore

$$\begin{aligned}\mathbb{E}\left[\kappa(\mathbf{u}\cdot F,\zeta)\hat{\gamma}(\mathbf{u}\cdot F,\zeta)\right] &\leqslant \mathbb{E}\left[\kappa(\mathbf{u}\cdot F,\zeta)^2\right]\left(\sum_{i=1}^{n}\mathbb{E}\left[\gamma_{\mathrm{Frob}}(f_i,u_i^*\eta)^2\right]\right)^{\frac{1}{2}}\\ &\leqslant 6n^2\left(\sum_{i=1}^{n}\tfrac{1}{4}d_i^3(d_i+n)\right)^{\frac{1}{2}}\\ &\leqslant 6n^3\left(\tfrac{1}{4}n^2D^4\right)^{\frac{1}{2}}=3n^4D^2.\end{aligned}$$

□

Corollary 33 (Main result)**.** *On input* $\mathbf{u}^{-1}\cdot F$ *and* $\mathbf{u}$*, Algorithm "Solve" outputs an approximate zero of* F *with* $O(n^5D^2)$ *continuation steps on the average and* $O(n^6D^4N)$ *operations on the average, when* $N\to\infty$*. When* $\min(n,D)\to\infty$*, this is* $N^{1+o(1)}$*.*

Proof. The cost of Algorithm "Solve" splits in two parts: the cost of the sampling and the cost of the numerical continuation. Note that the pair $(\mathbf{u}^{-1}\cdot F,\mathbf{u})$ has the same distribution as $(F,\mathbf{u})$, so without changing the expectation, we may study the cost of "Solve" on input $(F,\mathbf{u})$.

Concerning the sampling, we proved (Proposition 8) that it can be performed with sampling $O(n^3)$ times the standard normal distribution on $\mathbb{R}$ and computing of a zero of n homogeneous bivariate equations of degree at most d (and $O(n^3)$ extra operations). The equations that we solve for the sampling are restrictions of $f_1,\dots,f_n$ on random lines. Since the f_i are Gaussian, their restrictions are also Gaussian. We can use, for example, the continuation algorithm of Beltrán and Pardo (2011) to do this with average cost $O(D^4)$. So the total average cost of the sampling is $O(n^3+nD^4)$.

Concerning the numerical continuation. We saw in §4.2.2 that the cost of a continuation step is $O(nD^2N)$. In §4.2.3, we saw that we can choose interpolation paths of length at most $4n$. Therefore, by Theorem 19, the expectation of the number of continuation steps is

$$\mathbb{E}[K]\leqslant\mathbb{E}\left[\mathbb{E}\left[K|F\right]\right]\leqslant 100\cdot 13\cdot 4n\cdot\mathbb{E}\left[\kappa(\mathbf{v}\cdot F,\eta)\hat{\gamma}_{\mathrm{Frob}}(\mathbf{v}\cdot F,\eta)\right],$$

where $\mathbf{v}\in\mathcal{U}$ is uniformly distributed and η is a random root of $\mathbf{v}\cdot F$. By Theorem 32, we need $O(n^5D^2)$ continuation steps on the average, as $N\to\infty$, that is $O(n^6D^4N)$ elementary operations.

When $\min(n,D)\to\infty$, then both n and D are $\binom{n+D}{n}^{o(1)}=N^{o(1)}$. □

References

D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, and M. Shub (2016). "Condition Length and Complexity for the Solution of Polynomial Systems". In: *Foundations of Computational Mathematics*. DOI: 10.1007/s10208-016-9309-9.

– (2017). "A Stable, Polynomial-Time Algorithm for the Eigenpair Problem". In: *Journal of the European Mathematical Society*. To appear. arXiv: 1505.03290.

D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler (2013). *Numerically Solving Polynomial Systems with Bertini*. Vol. 25. Software, Environments, and Tools. SIAM, Philadelphia, PA.

C. Beltrán (2011). "A Continuation Method to Solve Polynomial Systems and Its Complexity". In: *Numerische Mathematik* 117.1, pp. 89–113. DOI: 10.1007/s00211-010-0334-3.

C. Beltrán and L. M. Pardo (2008). "On Smale's 17th Problem: A Probabilistic Positive Solution". In: *Foundations of Computational Mathematics* 8.1, pp. 1–43. DOI: 10.1007/s10208-005-0211-0.

– (2009a). "Efficient Polynomial System-Solving by Numerical Methods". In: *Journal of Fixed Point Theory and Applications* 6.1, pp. 63–85. DOI: 10.1007/s11784-009-0113-x.

– (2009b). "Smale's 17th Problem: Average Polynomial Time to Compute Affine and Projective Solutions". In: *Journal of the American Mathematical Society* 22.2, pp. 363–385. DOI: 10.1090/S0894-0347-08-00630-9.

– (2011). "Fast Linear Homotopy to Find Approximate Zeros of Polynomial Systems". In: *Foundations of Computational Mathematics* 11.1, pp. 95–129. DOI: 10.1007/s10208-010-9078-9.

C. Beltrán and M. Shub (2009). "Complexity of Bezout's Theorem. VII. Distance Estimates in the Condition Metric". In: *Foundations of Computational Mathematics* 9.2, pp. 179–195. DOI: 10.1007/s10208-007-9018-5.

L. Blum, M. Shub, and S. Smale (1989). "On a Theory of Computation and Complexity over the Real Numbers: NP-Completeness, Recursive Functions and Universal Machines". In: *Bulletin of the American Mathematical Society*. N.S. 21.1, pp. 1–46. DOI: 10.1090/S0273-0979-1989-15750-9.

A. Bostan et al. (2017). *Algorithmes Efficaces En Calcul Formel*. 1st ed. Palaiseau: Frédéric Chyzak (self-pub.)

P. Breiding and N. Vannieuwenhoven (2016). *The Condition Number of Join Decompositions*. arXiv: 1611.08117.

I. Briquel, F. Cucker, J. Peña, and V. Roshchina (2014). "Fast Computation of Zeros of Polynomial Systems with Bounded Degree under Finite-Precision". In: *Mathematics of Computation* 83.287, pp. 1279–1317. DOI: 10.1090/S0025-5718-2013-02765-2.

P. Bürgisser and F. Cucker (2011). "On a Problem Posed by Steve Smale". In: *Annals of Mathematics. Second Series* 174.3, pp. 1785–1836. DOI: 10.4007/annals.2011.174.3.8.

– (2013). *Condition: The Geometry of Numerical Algorithms*. Vol. 349. Grundlehren der Mathematischen Wissenschaften. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-38896-5.

P. Bürgisser and A. Lerario (2016). *Probabilistic Schubert Calculus*. arXiv: 1612.06893.

J.-P. Dedieu (2006). *Points Fixes, Zéros et La Méthode de Newton*. Vol. 54. Mathématiques & Applications. Springer. DOI: 10.1007/3-540-37660-7.

J.-P. Dedieu, G. Malajovich, and M. Shub (2013). "Adaptive Step-Size Selection for Homotopy Methods to Solve Polynomial Equations". In: *IMA Journal of Numerical Analysis* 33.1, pp. 1–29. DOI: 10.1093/imanum/drs007.

J. W. Demmel (1988). "The Probability That a Numerical Analysis Problem Is Difficult". In: *Mathematics of Computation* 50.182, pp. 449–480. DOI: 10.1090/S0025-5718-1988-0929546-7.

A. Edelman (1989). "Eigenvalues and Condition Numbers of Random Matrices". USA: Massachusetts Institute of Technology.

J. D. Hauenstein and A. C. Liddell (2016). "Certified Predictor–corrector Tracking for Newton Homotopies". In: *Journal of Symbolic Computation* 74, pp. 239–254. DOI: 10.1016/j.jsc.2015.07.001.

J. D. Hauenstein and F. Sottile (2012). "Algorithm 921: alphaCertified: Certifying Solutions to Polynomial Systems". In: *ACM Transactions on Mathematical Software* 38.4, pp. 1–20. DOI: 10.1145/2331130.2331136.

C. J. Hillar and L.-H. Lim (2013). "Most Tensor Problems Are NP-Hard". In: *Journal of the ACM* 60.6, pp. 1–39. DOI: 10.1145/2512329.

A. S. Householder (1958). "Unitary Triangularization of a Nonsymmetric Matrix". In: *Journal of the ACM* 5.4, pp. 339–342. DOI: 10.1145/320941.320947.

R. Howard (1993). "The Kinematic Formula in Riemannian Homogeneous Spaces". In: *Memoirs of the American Mathematical Society* 106.509. DOI: 10.1090/memo/0509.

W. Kahan (1966). *Accurate Eigenvalues of a Symmetric Tri-Diagonal Matrix*. CS41. Stanford University.

P. Lairez (2017). "A Deterministic Algorithm to Compute Approximate Roots of Polynomial Systems in Polynomial Average Time". In: *Foundations of Computational Mathematics*. DOI: 10.1007/s10208-016-9319-7.

G. Malajovich (1994). "On Generalized Newton Algorithms: Quadratic Convergence, Path-Following and Error Analysis". In: *Theoretical Computer Science* 133.1, pp. 65–84. DOI: 10.1016/0304-3975(94)00065-4.

– (2016). *Complexity of Sparse Polynomial Solving: Homotopy on Toric Varieties and the Condition Metric*. arXiv: 1606.03410.

J. Renegar (1987). "On the Efficiency of Newton's Method in Approximating All Zeros of a System of Complex Polynomials". In: *Mathematics of Operations Research* 12.1, pp. 121–148. DOI: 10.2307/3689676.

– (1989). "On the Worst-Case Arithmetic Complexity of Approximating Zeros of Systems of Polynomials". In: *SIAM Journal on Computing* 18.2, pp. 350–370. DOI: 10.1137/0218024.

M. Shub (1989). "On the Distance to the Zero Set of a Homogeneous Polynomial". In: *Journal of Complexity* 5.3, pp. 303–305. DOI: 10.1016/0885-064X(89)90027-7.

– (1993). "Some Remarks on Bezout's Theorem and Complexity Theory". In: *From Topology to Computation: Proceedings of the Smalefest*. Springer, New York, pp. 443–455.

M. Shub (2009). "Complexity of Bezout's Theorem. VI. Geodesics in the Condition (Number) Metric". In: *Foundations of Computational Mathematics* 9.2, pp. 171–178. DOI: 10.1007/s10208-007-9017-6.

M. Shub and S. Smale (1993a). "Complexity of Bézout's Theorem. I. Geometric Aspects". In: *Journal of the American Mathematical Society* 6.2, pp. 459–501. DOI: 10.2307/2152805.

– (1993b). "Complexity of Bezout's Theorem. II. Volumes and Probabilities". In: *Computational Algebraic Geometry (Nice, 1992)*. Vol. 109. Progr. Math. Boston: Birkhäuser, pp. 267–285.

– (1993c). "Complexity of Bezout's Theorem. III. Condition Number and Packing". In: *Journal of Complexity* 9.1, pp. 4–14. DOI: 10.1006/jcom.1993.1002.

– (1994). "Complexity of Bezout's Theorem. V. Polynomial Time". In: *Theoretical Computer Science* 133.1. Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993), pp. 141–164. DOI: 10.1016/0304-3975(94)90122-8.

– (1996). "Complexity of Bezout's Theorem. IV. Probability of Success; Extensions". In: *SIAM Journal on Numerical Analysis* 33.1, pp. 128–148. DOI: 10.1137/0733008.

S. Smale (1985). "On the Efficiency of Algorithms of Analysis". In: *Bulletin of The American Mathematical Society, New Series* 13.2, pp. 87–121. DOI: 10.1090/S0273-0979-1985-15391-1.

– (1986). "Newton's Method Estimates from Data at One Point". In: *The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics (Laramie, Wyo., 1985)*. Springer, New York, pp. 185–196.

– (1998). "Mathematical Problems for the next Century". In: *The Mathematical Intelligencer* 20.2, pp. 7–15. DOI: 10.1007/BF03025291.

J. von zur Gathen and J. Gerhard (1999). *Modern Computer Algebra*. New York: Cambridge University Press.

Inria Saclay Île-de-France, 91120 Palaiseau, France