Rigid continuation paths II. Structured polynomial systems

Peter Bürgisser, Felipe Cucker, Pierre Lairez

RIGID CONTINUATION PATHS II. STRUCTURED POLYNOMIAL SYSTEMS

PETER BÜRGISSER, FELIPE CUCKER, AND PIERRE LAIREZ

Abstract. We design a probabilistic algorithm that, on input $\varepsilon > 0$ and a polynomial system F given by black-box evaluation functions, outputs an approximate zero of F, in the sense of Smale, with probability at least $1-\varepsilon$. When applying this algorithm to $u \cdot F$, where u is uniformly random in the product of unitary groups, the algorithm performs $\mathrm{poly}(n,\delta) \cdot L(F) \cdot \big(\Gamma(F) \log \Gamma(F) + \log\log \varepsilon^{-1}\big)$ operations on average. Here n is the number of variables, δ the maximum degree, $L(F)$ denotes the evaluation cost of F, and $\Gamma(F)$ reflects an aspect of the numerical condition of F. Moreover, we prove that for inputs given by random Gaussian algebraic branching programs of size $\mathrm{poly}(n,\delta)$, the algorithm runs on average in time polynomial in n and δ. Our result may be interpreted as an affirmative answer to a refined version of Smale's 17th question, concerned with systems of *structured* polynomial equations.

Contents

1. Introduction

Can we solve polynomial systems in polynomial time? This question received different answers in different contexts. The NP-completeness of deciding the feasibility of a general polynomial system in both Turing and BSS models of computation is certainly an important difficulty but it does not preclude efficient algorithms for computing all the roots of a polynomial system or solving polynomial systems with as many equations as variables, for which the feasibility over algebraically closed fields is granted under genericity hypotheses. And indeed, there are several ways of computing all δ^n zeros of a generic polynomial system of n equations of degree $\delta > 1$ in n variables with $\mathrm{poly}(\delta^n)$ arithmetic operations (e.g. Renegar 1989; Lakshman 1991; Giusti, Lecerf, and Salvy 2001).

Date: October 20, 2020.
2000 *Mathematics Subject Classification.* Primary 68Q25; Secondary 65H10, 65H20, 65Y20.
Supported by the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 787840); a GRF grant from the Research Grants Council of the Hong Kong SAR (project number CityU 11300220); and the projet *De Rerum Natura* ANR-19-CE40-0018 of the French National Research Agency (ANR).

Smale's 17th problem (Smale 1998) is a clear-cut formulation of the problem in a numerical setting. It asks for an algorithm, with polynomial average complexity, for computing *one* approximate zero of a given polynomial system, where the complexity is to be measured with respect to the *dense input size* N, that is, the number of possible monomials in the input system. Smale's question was given recently a positive answer after seminal work by Shub and Smale (1993c,a,b, 1996, 1994), fundamental contributions by Beltrán and Pardo (2009, 2011) and Shub (2009), as well as our work (Bürgisser and Cucker 2011; Lairez 2017). The basic algorithmic idea underlying all these is *continuation along linear paths*. To find a zero of a system $F = (f_1, \dots, f_n)$ of n polynomials in n variables of degree at most δ, we first construct another system G with a built-in zero $\zeta_0 \in \mathbb{C}^n$ and consider the family $F_t \doteq tF + (1-t)G$ of polynomial systems. If G is generic enough, the zero ζ_0 of G extends as a continuous family (ζ_t) with $F_t(\zeta_t) = 0$, so that ζ_1 is a zero of F. It is possible to compute an approximation of ζ_1 by tracking ζ_t in finitely many steps. From the perspective of complexity analysis, the focal points are the choice of (G, ζ_0) and the estimation of the number of steps necessary for a correct approximation of ζ_1 (the cost of each step being not an issue as it is $O(N)$). The problem of choosing an initial pair (G, ζ_0) was for a long while a major obstable in complexity analysis. It was solved by Beltrán and Pardo (2009) who introduced an algorithm to sample a random polynomial system G together with a zero ζ_0 of it and provided a $\mathrm{poly}(n, \delta)N^2$ bound for the average number of steps in the numerical continuation starting from (G, ζ_0). This idea was followed in subsequent works with occasional cost improvements that decreased the exponent in N for the average number of steps. Note that for a system of n polynomial equations of degree δ in n variables, $N = n\binom{\delta+n}{n}$, and therefore $N \geqslant 2^{\min(\delta,n)}$. Regarding Smale's question, an $N^{O(1)}$ bound on this number is satisfactory but the question was posed, how much can the exponent in this bound be reduced?

1.1. **Rigid continuation paths.** The first part of this work (Lairez 2020)[1] gave an answer. It introduced continuation along rigid paths: the systems F_t have the form $F_t \doteq (f_1 \circ u_1(t), \dots, f_n \circ u_n(t))$ where the $u_i(t) \in U(n+1)$ are unitary matrices that depend continuously on the parameter t, while $f_1, \dots, f_n$ are fixed homogeneous polynomials. Compared to the previous setting, the natural parameter space for the continuation is not anymore the full space of all polynomial systems of a given degree, but rather the group $U(n+1)^n$, denoted $\mathcal{U}$. We developed analogues of Beltrán and Pardo's results for rigid paths. Building on this, we could prove a $\mathrm{poly}(n, \delta)$ bound on the average number of continuation steps required to compute one zero of a Kostlan random polynomial system,[2] yielding a $N^{1+o(1)}$ total complexity bound. This is the culmination of several results in this direction which improved the average analysis number of continuation steps (see Table 1) for solving random dense polynomial systems.

1.2. **Refinement of Smale's question.** What is at stake beyond Smale's question, is the understanding of numerical continuation as it happens in practice with a heuristic computation of the step lengths.[3] Experiments have shown that certified

[1] Hereafter refered to as "Part I".

[2] A Kostlan random polynomial system is a dense polynomial system where all coefficients are independent Gaussian complex random variables with an appropriate scaling, see §I.4.

[3] To heuristically determine a step length that is as large as possible, the principle is to try some step length and check if Newton's iteration seems to converge. Upon failure, the step length is reduced, and it is increased otherwise. Of course, this may go wrong in many ways.

	distribution	$\mathbb{E}$[#steps]	$\mathbb{E}$[total cost]
Shub and Smale (1996)	Kostlan	essentially poly(δ^n)	*not effective*
Shub and Smale (1994)	Kostlan	poly$(n,\delta)N^3$	*not effective*
Beltrán and Pardo (2009)	Kostlan	poly$(n,\delta)N^2$	poly$(n,\delta)N^3$
Beltrán and Shub (2009)	Kostlan	poly(n,δ)	*not effective*
Beltrán and Pardo (2011)	Kostlan	poly$(n,\delta)N$	poly$(n,\delta)N^2$
Bürgisser and Cucker (2011)	noncentered	poly$(n,\delta)N/\sigma$	poly$(n,\delta)N^2/\sigma$
Armentano et al. (2016)	Kostlan	poly$(n,\delta)N^{\frac{1}{2}}$	poly$(n,\delta)N^{\frac{3}{2}}$
Lairez (2020)	Kostlan	poly(n,δ)	poly$(n,\delta)N$

Table 1. Comparison of previous complexity analysis of numerical continuation algorithms for solving systems of n polynomial equations of degree δ in n variables. The parameter $N = n\binom{n+\delta}{n}$ is the dense input size. The parameter σ is the standard deviation for a noncentered distribution, in the context of smoothed analysis. Some results are not effective in that they do not lead to a complete algorithm to solve polynomial systems.

algorithms in the Shub-Smale line perform much smaller steps—and consequently many more steps—than heuristic methods for numerical continuation (Beltrán and Leykin 2012, 2013). In spite of progress in designing better and better heuristics (e.g., Timme 2020; Telen, Van Barel, and Verschelde 2020), the design of efficient algorithms for certified numerical continuation remains an important aspiration. With a view on closing the gap between rigorous step-length estimates and heuristics, a first observation—demonstrated experimentally by Hauenstein and Liddell (2016) and confirmed theoretically in Part I—highlights the role of higher-order derivatives. Shub and Smale's first-order step-length computation seems powerless in obtaining poly(n,δ) bounds on the number of steps: we need to get closer to Smale's γ to compute adequate step lengths (see Section 2 for a more detailed discussion).

However, estimating the higher-order derivatives occurring in γ is expensive. Thus, while using γ improves the average number of steps, it introduces a vice in the step-length computation. In Part I, we obtained a poly$(n,\delta)N$ complexity bound for estimating the variant $\hat{\gamma}_{\mathrm{Frob}}$ of γ (Proposition I.32) which, we showed, can be used to estimate step lengths. This cost is quasilinear with respect to the input size, we can hardly do better. But is N the right parameter to measure complexity? From a practical point of view, N is not so much relevant. Often N is much larger than the number of coefficients that actually define the input system, for example when the system is sparse or structured. This observation is turned to practical account by treating the input system not as a linear combination of monomials but as a black-box evaluation function, that is, as a routine that computes the value of the components of the system at any given point. Most implementations of numerical continuation do this. In this perspective, N does not play any role, and there is a need for adapting the computation of γ.

Designing algorithms for black-box inputs and analyzing their complexity for dense Gaussian random polynomial systems is interesting but misses an important point. The evaluation complexity of a random dense polynomial system is $\Theta(N)$, whereas the benefit of considering a black-box input is precisely to investigate systems with much lower evaluation complexity, and such systems have measure zero in the space of all polynomial systems. It is conceivable, even from the restricted perspective of numerical polynomial system solving, that intrinsically, polynomial

systems with low evaluation complexity behave in a different way than random dense polynomial systems. So Smale's original question of solving polynomial systems in polynomial time leads to the following refined question:

> Can we compute an approximate zero of a *structured* polynomial system F given by *black-box evaluation functions* with $\mathrm{poly}(n, \delta)$ many arithmetic operations and evaluations of F on average?

We use algebraic branching programs (ABPs), a widely studied concept in algebraic complexity theory (see §1.8), as a model of computation for polynomials with low evaluation complexity. Further, we introduce a natural model of *Gaussian random algebraic branching programs* in order to capture the aspect of randomization. The main result of this paper is an affirmative answer to the above refined question in this model.

1.3. **Polynomial systems given by black-box evaluation.** The model of computation is the BSS model, extended with rational exponentiation for convenience and a "6th type of node", as introduced by Shub and Smale (1996), that computes an exact zero in $\mathbb{P}^1$ of a bivariate homogeneous polynomial given an approximate zero (this is used in the sampling step), see Part I, §4.3.1 for a discussion. The term "black-box" refers to a mode of computation with polynomials where we assume only the ability to evaluate them at a complex point. Concretely, the polynomials are represented by programs, or BSS machines. For a black-box polynomial $f \in \mathbb{C}[z_1, \ldots, z_n]$, we denote by $L(f)$ the number of operations performed by the program representing f to evaluate f at a point in $\mathbb{C}^n$. For a polynomial system $F = (f_1, \ldots, f_n)$, we write $L(F) \doteq L(f_1) + \cdots + L(f_n)$. It is possible that evaluating F costs less than evaluating its components separately, as some computations may be shared, but we cannot save more than a factor n, so we ignore the issue. More generally, in this article, we will not enter the details of the $\mathrm{poly}(n, \delta)$ factors. The ability to evaluate first-order derivatives will also be used. For a univariate polynomial f of degree at most δ the derivative at 0 can be computed from evaluations using the formula

$$f'(0) = \frac{1}{\delta + 1} \sum_{i=0}^{\delta} \omega^{-i} f(\omega^i), \tag{1.1}$$

where $\omega \in \mathbb{C}$ is a primitive $(\delta + 1)$th root of unity. Similar formulas hold for multivariate polynomials. In practice, automatic differentiation (e.g., Baur and Strassen 1983) may be used. In any case, we can evaluate the Jacobian matrix of a black-box polynomial system F with $\mathrm{poly}(n, \delta)L(F)$ operations. Since this is below the resolution that we chose, we do not make specific assumptions on the evaluation complexity of the Jacobian matrix. Moreover, the degree of a black-box polynomial can be computed with probability 1 in the BSS model by evaluation and interpolation along a line.[4] So there is no need for the degree to be specified separately.

1.4. **The $\Gamma(f)$ number.** Beyond the evaluation complexity $L(F)$, the hardness of computing a zero of F in our setting depends on an averaged γ number. For a

[4]If the values of a univariate polynomial f at $d + 2$ independent Gaussian random points coincide with the values of a degree at most d polynomial at the same points, then f has degree at most d with probability 1, so we can compute, in the BSS model, the degree of a black-box univariate polynomial. Furthermore, the degree of a multivariate polynomial F is equal to the degree of the univariate polynomial obtained by restricting F on a uniformly distributed line passing through the origin, with probability 1.

polynomial $f \in \mathbb{C}[z_0, \ldots, z_n]$, recall that

$$\gamma(f,z) \doteq \sup_{k\geqslant 2} \left(\|\mathrm{d}_z f\|^{-1} \left\|\left\| \tfrac{1}{k!}\mathrm{d}_z^k f\right\|\right|\right)^{\frac{1}{k-1}}, \tag{1.2}$$

where the triple norm $|||A|||$ of a k-multilinear map A is defined as $\sup \frac{\|A(z_1,\ldots,z_k)\|}{\|z_1\|\cdots\|z_k\|}$. If f is homogeneous and $[z] \in \mathbb{P}^n$ is a projective point, we define $\gamma(f,[z]) \doteq \gamma(f,z)$, for some representative $z \in \mathbb{S}(\mathbb{C}^{n+1})$. The definition does not depend on the representative. By Lemma I.11 (in Part I), $\gamma(f,z) \geqslant \frac{1}{2}(\delta-1)$ if f is homogeneous of degree δ, and $\gamma(f,z) = 0$ if $\delta = 1$. For computational purposes, we prefer the *Frobenius γ number* introduced in Part I:

$$\gamma_{\mathrm{Frob}}(f,z) \doteq \sup_{k\geqslant 2} \left(\|\mathrm{d}_z f\|^{-1} \left\|\tfrac{1}{k!}\mathrm{d}_z^k f\right\|_{\mathrm{Frob}}\right)^{\frac{1}{k-1}}, \tag{1.3}$$

where $\|-\|_{\mathrm{Frob}}$ is the Frobenius norm of a multilinear map (see §I.4.2). The two variants are tightly related (Lemma I.29):

$$\gamma(f,z) \leqslant \gamma_{\mathrm{Frob}}(f,z) \leqslant (n+1)\gamma(f,z). \tag{1.4}$$

We will not need here to define, or use, the γ number of a polynomial system. For a homogeneous polynomial $f \in \mathbb{C}[z_0, \ldots, z_n]$ of degree $\delta \geqslant 2$, we define the *averaged γ number* as

$$\Gamma(f) \doteq \mathbb{E}_\zeta \left[\gamma_{\mathrm{Frob}}(f,\zeta)^2\right]^{\frac{1}{2}} \in [\tfrac{1}{2}, \infty], \tag{1.5}$$

where ζ is a uniformly distributed zero of f in $\mathbb{P}^n$. For a homogeneous polynomial system $F = (f_1, \ldots, f_n)$, we define

$$\Gamma(F) \doteq \left(\Gamma(f_1)^2 + \cdots + \Gamma(f_n)^2\right)^{\frac{1}{2}} \leqslant \sum_{i=1}^n \Gamma(f_i). \tag{1.6}$$

While $L(F)$ reflects an algebraic structure, $\Gamma(F)$ reflects a numerical aspect.

Let $d_1, \ldots, d_n$ be integers $\geqslant 2$ and let $\mathcal{H}$ be the space of homogeneous polynomial systems $(f_1, \ldots, f_n)$ with $f_i \in \mathbb{C}[z_0, \ldots, z_n]$ homogeneous of degree d_i. Let $\delta \doteq \max_i d_i$. Let $\mathcal{U}$ be the group $U(n+1)^n$ made of n copies of the group of unitary matrices of size $n+1$. For $\mathbf{u} = (u_1, \ldots, u_n) \in \mathcal{U}$ and $F = (f_1, \ldots, f_n) \in \mathcal{H}$, we define the action

$$\mathbf{u} \cdot F = \left(f_1 \circ u_1^{-1}, \ldots, f_n \circ u_n^{-1}\right). \tag{1.7}$$

It plays a major role in the setting of rigid continuation paths. Note that Γ is unitary invariant: $\Gamma(\mathbf{u} \cdot F) = \Gamma(F)$ for any $\mathbf{u} \in \mathcal{U}$. Concerning L, we have $L(\mathbf{u} \cdot F) \leqslant L(F) + O(n^3)$, using (1.7) as a formula to evaluate $\mathbf{u} \cdot F$. (Note that the matrices u_i are unitary, so the inverse is simply the Hermitian transpose.)

1.5. **Main results I.** In our first main result, we design the algorithm BoostBlackBoxSolve in the setting of rigid continuation paths (see §4.3) for computing with high probability an approximate zero of a black-box polynomial system F. We give an average analysis when the input system is $\mathbf{u} \cdot F$ where $\mathbf{u}$ is uniformly distributed and F is fixed.

Theorem 1.1. *Let F be a homogeneous polynomial system of n equations of degree at most δ in $n+1$ variables, with only regular zeros in $\mathbb{P}^n(\mathbb{C})$. On input F, given as a black-box evaluation program, and $\varepsilon > 0$, Algorithm* BoostBlackBoxSolve *computes an approximate zero of F with probability at least $1 - \varepsilon$.*

If $\mathbf{u} \in \mathcal{U}$ *is uniformly distributed, then on input* $\mathbf{u} \cdot F$ *and* ε, BoostBlackBoxSolve *performs*

$$\operatorname{poly}(n, \delta) \cdot L(F) \cdot \left(\Gamma(F) \log \Gamma(F) + \log\log \varepsilon^{-1}\right)$$

operations on average.

In addition to the foundations laid in Part I, the main underlying tool is a Monte-Carlo method for estimating Smale's γ number: with $\operatorname{poly}(n, \delta) \log \frac{1}{\varepsilon}$ evaluations of f, we can estimate $\gamma(f, z)$ within a factor $\operatorname{poly}(n, \delta)$ with probability at least $1 - \varepsilon$ (see Theorem 3.3). This turns both the computation of the step length and the whole zero-finding process into Monte-Carlo algorithms themselves and, as a consequence, BoostBlackBoxSolve departs from the simple structure of continuation algorithms described above. During execution, BoostBlackBoxSolve draws real numbers from the standard Gaussian distribution to compute the initial pair (G, ζ) and estimate various γ_{Frob}. The average cost in Theorem 1.1 is considered with respect to both this inner randomization of the algorithm, and the randomness of the input $\mathbf{u}$ (or F in Corollary 1.2 below).

BoostBlackBoxSolve actually performs the continuation procedure several times, possibly with different initial pairs, as well as a validation routine that drastically decreases the probability that the returned point is not an approximate zero of F. Its complexity analysis reflects this more complicated structure.

In contrast with many previous work, BoostBlackBoxSolve does not always succeed: its result can be wrong with a small given probability ε, but the doubly logarithmic dependence of the complexity with respect to ε is satisfactory. We do not know if it is optimal but it seems difficult, in the black-box model, to obtain an algorithm with similar complexity bounds but that succeeds (i.e., returns a certified approximate zero) with probability one: to the best of our knowledge all algorithms for certifying zeros need some global information—be it the Weyl norm of the system (Hauenstein and Sottile 2012) or evaluation in interval arithmetic (Rump and Graillat 2010)—which we cannot estimate with probability 1 in the black-box model with only $\operatorname{poly}(n, \delta)$ evaluations. So unless we add an *ad hoc* hypothesis (such as a bound on the coefficients in the monomial basis), we do not know how to certify an approximate zero in the black-box model.

Theorem 1.1 can be interpreted as an average analysis on a orbit of the action of $\mathcal{U}$ on $\mathcal{H}$. More generally, we may assume a random input $F \in \mathcal{H}$ where the distribution of F is *unitary invariant*, meaning that for any $\mathbf{u} \in \mathcal{U}$, $\mathbf{u} \cdot F$ and F have the same distribution. This leads to the following statement.

Corollary 1.2. *Let* $F \in \mathcal{H}$ *be a random polynomial system with unitary invariant distribution, let* L *be an upper bound on* $L(F)$ *and* $\Gamma = \mathbb{E}[\Gamma(F)^2]^{\frac{1}{2}}$. *On input* F *(given as a black-box evaluation program) and* $\varepsilon > 0$, *Algorithm* BoostBlackBoxSolve *computes an approximate zero of* F *with probability at least* $1 - \varepsilon$ *with*

$$\operatorname{poly}(n, \delta) \cdot L \cdot \left(\Gamma \log \Gamma + \log\log \varepsilon^{-1}\right)$$

operations on average.

The quantity $\Gamma(F)$ strongly influences the average complexity in Theorem 1.1 and Corollary 1.2 and while it is natural to expect the complexity to depend on numerical aspects of F, it is desirable to quantify this dependence by averaging over F (Smale 1997). It was shown in Part I that if $F \in \mathcal{H}$ is a Kostlan random polynomial system, then $\mathbb{E}[\Gamma(F)^2] = \operatorname{poly}(n, \delta)$ (Lemma I.38). Together with the

standard bound $L(F) = O(N)$, we immediately obtain from Corollary 1.2 the following complexity analysis, similar to the main result of Part I (Theorem I.40), but assuming only a black-box representation of the input polynomial system.

Corollary 1.3. *Let $F \in \mathcal{H}$ be a Kostlan random polynomial system. On input F and $\varepsilon > 0$, Algorithm* BoostBlackBoxSolve *computes an approximate zero of F with probability at least $1 - \varepsilon$ with $\mathrm{poly}(n, \delta) \cdot \log\log \varepsilon^{-1}$ operations and evaluations of F on average.*

Our second main result (Theorem 1.4 below) states that exact same bound for polynomial system given by independent Gaussian random algebraic branching programs. We next introduce this model.

1.6. **Algebraic branching programs.** Following Nisan (1991), an *algebraic branching program* (ABP) of degree δ is a labeled directed acyclic graph with one source and one sink, with a partition of the vertices into levels, numbered from 0 to δ, such that each edge goes from level i to level $i+1$. The source is the only vertex at level 0 and the sink is the only vertex at level δ. Each edge is labeled with a homogeneous linear form in the input variables $z_0, \dots, z_n$. An ABP *computes* the polynomial obtained as the sum over all paths from the source to the sink of the product of the linear forms by which the edges of the path are labelled. It is a homogeneous polynomial of degree δ. The *width* r of the ABP is the maximum of the cardinalities of the level sets. The *size* s of the ABP, which is defined as the number of its vertices, satisfies $r \leqslant s \leqslant (\delta - 1)r + 2$. Any homogeneous polynomial f can be computed by an ABP and the minimum size or width of an ABP computing f are important measures of the complexity of f, see §1.8.

While ABPs provide an elegant graphical way of formalizing computations with polynomials, we will use an equivalent matrix formulation. Suppose that the ith level set has r_i vertices and let $A_i(z)$ denote the weighted adjacency matrix of format $r_{i-1} \times r_i$, whose entries are the weights of the edges between vertices of level $i-1$ and level i. Thus the entries of $A_i(z)$ are linear forms in the variables $z_0, \dots, z_n$. The polynomial $f(z)$ computed by the ABP can then be expressed as the trace of iterated matrix multiplication, namely,

$$f(z) = \mathrm{tr}\,(A_1(z) \cdots A_\delta(z))\,. \tag{1.8}$$

It is convenient to relax the assumption $r_0 = r_\delta = 1$ to $r_0 = r_\delta$. Compared to the description in terms of ABPs, this adds some flexibility because the trace is invariant under cyclic permutation of the matrices $A_i(z)$.

Using the associativity of matrix multiplication, we can evaluate $f(z)$ efficiently by iterated matrix multiplication, which amounts to $O(\delta r^3)$ additions or multiplications of matrix entries; taking into account the cost $O(n)$ of evaluating a matrix entry (which is a linear forms in the variables $z_0, \dots, z_n$), we see that we can evaluate f with a total of $O(\delta r^2 n \delta r^3)$ arithmetic operations.

1.7. **Main results II.** Given positive integers $r_1, \dots, r_{\delta-1}$, we can form a random ABP (that we call *Gaussian random ABP*) of degree δ by considering a directed acyclic graph with r_i vertices in the layer i (for $1 \leqslant i \leqslant \delta - 1$), one vertex in the layers 0 and δ, and all possible edges from a layer to the next, labelled by linear forms in $z_0 \dots, z_n$ with independent and identically distributed complex Gaussian coefficients. This is equivalent to assuming that the adjacency matrices are linear forms $A_i(z) = A_{i0}z_0 + \cdots + A_{in}z_n$ with independent complex standard Gaussian matrices $A_{ij} \in \mathbb{C}^{r_{i-1} \times r_i}$.

We call a Gaussian random ABP *irreducible* if all layers (except the first and the last) have at least two vertices. The polynomial computed by an irreducible Gaussian random ABP is almost surely irreducible (Lemma 5.1), and conversely, the polynomial computed by a Gaussian random ABP that is not irreducible is not irreducible; which justifies the naming.

Recall the numerical parameter Γ entering the complexity of numerical continuation in the rigid setting, see (1.5) and Theorem 1.1. The second main result in this article is an upper bound on the expectation of $\Gamma(f)$, when f is computed by a Gaussian random ABP. Remarkably, the bound does not depend on the sizes r_i of the layers defining the Gaussian random ABP; in particular it is independent of its width!

Theorem 1.4. *If f is the random polynomial computed by an irreducible Gaussian random ABP of degree δ, then*

$$\mathbb{E}\left[\Gamma(f)^2\right] \leqslant \tfrac{3}{4}\delta^3(\delta+n)\log\delta.$$

The distribution of the polynomial computed by a Gaussian random ABP is unitarily invariant so, as a consequence of Corollary 1.2, we obtain polynomial complexity bounds for solving polynomial systems made of Gaussian random ABP.

Corollary 1.5. *If $f_1,\dots,f_n$ are independent irreducible Gaussian random ABPs of degree at most δ and evaluation complexity at most L, then* BoostBlackBoxSolve*, on input $f_1,\dots,f_n$ and $\varepsilon>0$ computes a zero of $(f_1,\dots,f_n)$ with probability at least $1-\varepsilon$ in*

$$\mathrm{poly}(n,\delta)\cdot L\cdot\log\log\varepsilon^{-1}$$

operations on average.

This result provides an answer to the refined Smale's problem raised at the end of §1.2, where "structured" is interpreted as "low evaluation complexity in the ABP model".

The polynomial systems computed by ABPs of with r form a zero measure subset of $\mathcal{H}$ when n and δ are large enough. More precisely, they form a subvariety of $\mathcal{H}$ of dimension at most $r^2\delta n$ while the dimension of $\mathcal{H}$ grows superpolynomially with n and δ. Note also that a polynomial f computed by a Gaussian random ABP may be almost surely singular (in the sense that the projective hypersurface that it defines is singular), see Lemma 5.2. This strongly contrasts with previously considered stochastic model of polynomial systems.

Lastly, it would be interesting to describe the limiting distribution of the polynomial computed by a Gaussian random ABP as the size of the layers goes to infinity. Since this question is out of the scope of this article, we leave it open.

1.8. **On the role of algebraic branching programs.** To motivate our choice of the model of ABPs, we point out here their important role in algebraic complexity theory, notably in Valiant's algebraic framework of NP-completeness (Valiant 1979, 1982), see also Bürgisser (2000). This model features the complexity class VBP, which models efficiently computable polynomials as sequences of multivariate complex polynomials f_n, where the degree of f_n is polynomially bounded in n and the homogeneization of f_n can be computed by an ABP of width polynomially bounded in n. It is known (Toda 1992; Malod and Portier 2008) that the sequence of determinants of generic $n\times n$ matrices is complete for the class VBP: this means the determinants have efficient computations in this model and moreover,

any $(f_n) \in \mathrm{VBP}$ can be tightly reduced to a sequence of determinants in the sense that f_n can be written as the determinant of a matrix, whose entries are affine linear forms, and such that the size of the matrix is polynomially bounded in n. The related complexity class VP consists of the sequences of multivariate complex polynomials f_n, such that the degree of f_n grows at most polynomially in n and such that f_n can be computed by an arithmetic circuit (equivalently, straightline program) of size polynomially bounded in n. While it is clear that $\mathrm{VBP} \subseteq \mathrm{VP}$, it is a longstanding open question whether equality holds. However, after relaxing "polynomially bounded'" to "quasi-polynomially bounded" [5], the classes collapse (e.g., see Malod and Portier (2008)). These results should make clear the relevance and universality of the model of ABPs. Moreover, Valiant (1979) defined another natural complexity class VNP, formalizing efficiently definable polynomials for which the sequence of permanents of generic matrices is complete. Valiant's conjecture $\mathrm{VBP} \neq \mathrm{VNP}$ is a version of the famous $\mathrm{P} \neq \mathrm{NP}$ conjecture.

1.9. **Organization of paper.** In Section 2 we first recall the basics of the complexity analysis of numerical continuation algorithms and summarize the results obtained in Part I. Section 3 is devoted to numerical continuation algorithms when the functions are given by a black-box. We introduce here a sampling algorithm to estimate γ_{Frob} with high probability in this setting. Section 4 is devoted to the complexity analysis of the new algorithm on a random input $\mathbf{u} \cdot F$. In particular, in §4.3, we consider the problem of certifying an approximate zero in the black-box model and we prove Theorem 1.1. Finally, Section 5 presents the proof of Theorem 1.4, our second main result.

2. Numerical continuation with few steps

2.1. **The classical setting.** Numerical continuation algorithms have been so far the main tool for the complexity analysis of numerical solving of polynomial systems. We present here the main line of the theory as developed by Shub and Smale (1993c,a, 1996, 1994), Beltrán and Pardo (2009, 2011), and Beltrán (2011). The general idea to solve a polynomial system $F \in \mathcal{H}$ consists of embedding F in a one-parameter continuous family $(F_t)_{t\in[0,1]}$ of polynomial systems such that $F_1 = F$ and a zero of F_0, say $\zeta_0 \in \mathbb{P}^n$ is known. Then, starting from $t = 0$ and $z = \zeta_0$, t and z are updated to track a zero of F_t all along the path from F_0 to F_1, as follows:

while $t < 1$ **do** $t \leftarrow t + \Delta t$; $z \leftarrow \mathrm{Newton}(F_t, z)$ **end while**,

where Δt needs to be defined. The idea is that z always stays close to ζ_t, the zero of F_t obtained by continuing ζ_0. To ensure correctness, the increment Δt should be chosen small enough. But the bigger Δt is, the fewer iterations will be necessary, meaning a better complexity. The size of Δt is typically controlled with, on the one hand, effective bounds on the variations of the zeros of F_t as t changes, and on the other hand, effective bounds on the convergence of Newton's iteration. The general principle to determine Δt is the following, in very rough terms because a precise argument generally involve lengthy computations. The increment Δt should be small enough so that ζ_t is in the basin of attraction around $\zeta_{t+\Delta t}$ of Newton's iteration for $F_{t+\Delta t}$. This leads to the rule-of-thumb $\|\Delta\zeta_t\|\rho(F_{t+\Delta t}, \zeta_{t+\Delta t}) \lesssim 1$, where $\Delta\zeta_t = \zeta_{t+\Delta t} - \zeta_t$ and $\rho(F_t, \zeta_t)$ is the inverse of the radius of the basin of

[5] *Quasi-polynomially bounded* in n means bounded by $2^{(\log n)^c}$ for some constant c.

attraction of Newton's iteration. A condition that we can rewrite as

$$\frac{1}{\Delta t} \gtrsim \rho(F_t, \zeta_t) \left\| \frac{\Delta \zeta_t}{\Delta t} \right\|, \tag{2.1}$$

assuming that

$$\rho(F_{t+\Delta t}, \zeta_{t+\Delta t}) \simeq \rho(F_t, \zeta_t). \tag{2.2}$$

The factor $\frac{\Delta \zeta_t}{\Delta t}$ is almost the derivative $\dot{\zeta}_t$ of ζ_t with respect to t. It is generally bounded using a *condition number* $\mu(F_t, \zeta_t)$, that is the largest variation of the zero ζ_t after a pertubation of F_t in $\mathcal{H}$, so that

$$\left\| \frac{\Delta \zeta_t}{\Delta t} \right\| \simeq \|\dot{\zeta}_t\| \leqslant \mu(F_t, \zeta_t) \|\dot{F}_t\|, \tag{2.3}$$

where $\dot{F}_t$ (resp. $\dot{\zeta}_t$) is the derivative of F (resp. ζ_t) with respect to t, and the right-hand side is effectively computable. The parameter $\rho(F_t, \zeta_t)$ is much deeper. Smale's α-theory has been a preferred tool to deal with it in many complexity analyses. The number γ takes a prominent role in the theory and controls the convergence of Newton's iteration (Smale 1986): $\rho(F_t, \zeta_t) \lesssim \gamma(F_t, \zeta_t)$. (For the definition of $\gamma(F, \zeta)$, e.g., see Eq. (8) in Part I.) So we obtain the condition

$$\frac{1}{\Delta t} \gtrsim \gamma(F_t, \zeta_t) \mu(F_t, \zeta_t) \|\dot{F}_t\| \tag{2.4}$$

that ensures the correctness of the algorithm. A rigorous argument requires a nice behavior of both factors $\gamma(F_t, \zeta_t)$ and $\mu(F_t, \zeta_t)$ as t varies, this is a crucial point, especially in view of the assumption (2.2). The factor $\|\dot{F}_t\|$ is generally harmless; the factor $\mu(F_t, \zeta_t)$ is important but the variations with respect to t are generally easy to handle; however the variations of $\gamma(F_t, \zeta_t)$ are more delicate. This led Shub and Smale (1993c) to consider the upper bound (called "higher-derivative estimate")

$$\gamma(F, z) \lesssim \mu(F, z), \tag{2.5}$$

with the same μ as above, and the subsequent correctness condition

$$\frac{1}{\Delta t} \gtrsim \mu(F_t, \zeta_t)^2 \|\dot{F}_t\|. \tag{2.6}$$

Choosing at each iteration Δt to be the largest possible value allowed by (2.6), we obtain a numerical continuation algorithm, with adaptive step length, whose number K of iterations is bounded, as shown first by Shub (2009), by

$$K \lesssim \int_0^1 \mu(F_t, \zeta_t)^2 \|\dot{F}_t\| \mathrm{d}t. \tag{2.7}$$

It remains to choose the starting system F_0, with a built-in zero ζ_0, and the path from F_0 to F_1. For complexity analyses, the most common choice of path is a straight-line segment in the whole space of polynomial systems $\mathcal{H}$. For the choice of the starting system F_0, Beltrán and Pardo (2009, 2008) have shown that a Kostlan random system is a relevant choice and that there is a simple algorithm to sample a random system with a known zero. If F_1 is also a random Gaussian system, then all the intermediate systems F_t are also random Gaussian, and using (2.7), we obtain a bound, following Beltrán and Pardo, on the expected number of iterations in the numerical continuation from F_0 to F_1:

$$\mathbb{E}_{F_0, \zeta_0, F_1}[K] \simeq \mathbb{E}_{F, \zeta}[\mu(F, \zeta)^2] \simeq \dim \mathcal{H}, \tag{2.8}$$

where ζ is a random zero of F. The dimension of $\mathcal{H}$ is the number of coefficients in F, it is the *input size*. For n equations of degree δ in n variables, we compute

$$\dim \mathcal{H} = n\binom{n+\delta}{n}. \tag{2.9}$$

This is larger than any polynomial in n and δ (as n and δ go to ∞), but is much smaller than δ^n, the generic number of solutions of such a system. The cost of an iteration (computing the step size and performing one Newton's iteration) is also bounded by the input size. So we have an algorithm whose average complexity is polynomial in the input size. This is a major complexity result because it breaks the $\mathrm{poly}(\delta^n)$ barrier set by algorithms that compute all solutions simultaneously. However, the bound (2.8) on the expected number of iterations is still much larger than what heuristic algorithms seem to achieve.

A first idea to design a faster algorithm would be to search for a better continuation path in order to lower the right-hand side in (2.7). Such paths do exist and can give a $\mathrm{poly}(n,\delta)$ bound on $\mathbb{E}[K]$ (Beltrán and Shub 2009). Unfortunately, their computation requires, in the current state of the art, to solve the target system first. A second approach focuses on sharpening the correctness condition (2.6), that is, on making bigger continuation steps. The comparison of (2.6) with heuristics shows that there is room for improvement (Beltrán and Leykin 2012, 2013). In devising this condition, two inequalities are too generous. Firstly, Inequality (2.3) bounds the variation of ζ_t by the worst-case variation. The average worst-case variation can only grow with the dimension of the parameter space, $\dim \mathcal{H}$, and it turns out to be much bigger than the average value of $\|\dot\zeta_t\|$, which is $\mathrm{poly}(n,\delta)$. This was successfully exploited by Armentano et al. (2016) to obtain the bound $\mathbb{E}[K] \lesssim \sqrt{\dim \mathcal{H}}$ for random Gaussian systems. They used straight-line continuation paths but a finer computation of the step size. The other inequality that turns out to be too coarse is (2.5): the higher derivatives need to be handled more accurately.

2.2. **Rigid continuation paths.** In Part I, we introduced rigid continuation paths to obtain, in the case of random Gaussian systems, the bound

$$\mathbb{E}[K] \leqslant \mathrm{poly}(n,\delta). \tag{2.10}$$

To solve a polynomial system $F = (f_1,\dots,f_n) \in \mathcal{H}$ in $n+1$ homogeneous variables, we consider continuation paths having the form

$$F_t \doteq \left(f_1 \circ u_1^{-1}(t),\dots,f_n \circ u_n^{-1}(t)\right), \tag{2.11}$$

where $u_1(t),\dots,u_n(t) \in U(n+1)$ are unitary matrices depending on the parameter t, with $u_i(1) = \mathrm{id}$. The parameter space for the numerical continuation is not $\mathcal{H}$ anymore but $U(n+1)^n$, denoted $\mathcal{U}$, a real manifold of dimension n^3. For $\mathbf{u} = (u_1,\dots,u_n) \in \mathcal{U}$ and $F \in \mathcal{H}$, we denote

$$\mathbf{u}\cdot F \doteq \left(f_1 \circ u_1^{-1},\dots,f_n \circ u_n^{-1}\right) \in \mathcal{H}. \tag{2.12}$$

We developed in this setting an analogue of Beltrán and Pardo's algorithm. Firstly, we sample uniformly $\mathbf{v} \in \mathcal{U}$ together with a zero of the polynomial system $\mathbf{v}\cdot F$. The same kind of construction as in the Gaussian case makes it possible to perform this operation without solving any polynomial system (only n univariate equations). Then, we construct a path $(\mathbf{u}_t)_{t\in[0,1]}$ in $\mathcal{U}$ between $\mathbf{v}$ and $\mathbf{1}_{\mathcal{U}}$, and perform numerical continuation using $F_t \doteq \mathbf{u}_t \cdot F$. The general strategy sketched in §2.1 applies but the rigid setting features important particularities. The most salient of which is the average conditioning, that is, the average worst-case variation of ζ_t with respect

to infinitesimal variations of $\mathbf{u}_t$. It is now poly(n) (see §I.3.2), mostly because the dimension of the parameter space is poly(n). Besides, the way the continuation path is designed preserves the geometry of the equations. This is reflected in a better behavior of $\gamma(F_t, \zeta_t)$ as t varies, which makes it possible to use an upper bound much finer than (2.5), that we called the *split γ number*. In the case of a random Gaussian input, we obtained in the end a poly(n, δ) bound on the average number of iterations for performing numerical continuation along rigid paths.

2.3. **The split γ number.** Computing a good upper bound of the γ number is the key to make bigger continuation steps. We recall here the upper bound introduced in Part I. The *incidence condition number* of $F = (f_1, \dots, f_n)$ at z is

$$\kappa(F, z) \doteq \left\| (\mathrm{d}_z F_z)^\dagger \right\|, \tag{2.13}$$

where $\dagger$ denotes the Moore–Penrose pseudoinverse and F_z the normalized system

$$F_z \doteq \left(\frac{f_1}{\|\mathrm{d}_z f_1\|}, \dots, \frac{f_n}{\|\mathrm{d}_z f_n\|} \right). \tag{2.14}$$

When z is a zero of F, this quantity depends only on the angles formed by the tangent spaces at z of the n hypersurfaces $\{f_i = 0\}$ (see §I.2.1 and §I.3 for more details). It is closely related to the intersection condition number introduced by Bürgisser (2017). In the context of rigid paths, it is also the natural condition number: the variation of a zero ζ of a polynomial system $\mathbf{u} \cdot F$ under a perturbation of $\mathbf{u}$ is bounded by $\kappa(\mathbf{u} \cdot F, \zeta)$ (Lemma I.16). Moreover, F being fixed, if $\mathbf{u} \in \mathcal{U}$ is uniformly distributed and if ζ is a uniformly distributed zero of $\mathbf{u} \cdot F$, then $\mathbb{E}[\kappa(\mathbf{u} \cdot F, \zeta)^2] \leqslant 6n^2$ (Proposition I.17).

The *split γ number* is defined as

$$\hat{\gamma}(F, z) \doteq \kappa(F, z) \left(\gamma(f_1, z)^2 + \cdots + \gamma(f_n, z)^2 \right)^{\frac{1}{2}}. \tag{2.15}$$

It tightly upper bounds $\gamma(F, z)$ in that (Theorem I.13)

$$\gamma(F, z) \leqslant \hat{\gamma}(F, z) \leqslant n\kappa(F, z)\gamma(F, z). \tag{2.16}$$

Whereas $\gamma(F, z)$ does not behave nicely as a function of F, the split variant behaves well in the rigid setting: F being fixed, the function $\mathcal{U} \times \mathbb{P}^n \to \mathbb{R}$, $(\mathbf{u}, z) \mapsto \hat{\gamma}(\mathbf{u} \cdot F, z)^{-1}$ is 13-Lipschitz continuous (Lemma I.21).[6] This makes it possible to perform numerical continuation. Note that we need not compute γ exactly, an estimate within a fixed ratio is enough. For computational purposes, we rather use the variant γ_{Frob}, defined in (1.3), in which the operator norm is replaced by a Hermitian norm. It induces a *split γ_{Frob} number*

$$\hat{\gamma}_{\mathrm{Frob}}(F, z) \doteq \kappa(F, z) \left(\gamma_{\mathrm{Frob}}(f_1, z)^2 + \cdots + \gamma_{\mathrm{Frob}}(f_n, z)^2 \right)^{\frac{1}{2}} \tag{2.17}$$

as in (2.15). Algorithm 1 describes the computation of an approximate zero of a polynomial system $\mathbf{u} \cdot F$, given a zero of some $\mathbf{v} \cdot F$. (It is the same as Algorithm I.2, with $\hat{\gamma}_{\mathrm{Frob}}$ for g and $C = 15$, which gives the constant 240 that appears in Algorithm 1.) As an analogue of (2.7), Theorem I.23 bounds the number K of continuation steps performed by Algorithm 1 as an integral over the continuation path:

$$K \leqslant 325 \int_0^T \kappa(\mathbf{w}_t \cdot F, \zeta_t) \hat{\gamma}_{\mathrm{Frob}}(\mathbf{w}_t \cdot F, \zeta_t) \|\dot{\mathbf{w}}_t\| \mathrm{d}t. \tag{2.18}$$

[6] Note that the importance of such a Lipschitz property has been highlighted by Demmel (1987). It implies that $1/13\gamma$ is upper bounded on $\mathcal{U} \times \mathbb{P}^n$ by the distance to the subset of all pairs $(\mathbf{u}, \zeta)$ where ζ is a singular zero of $\mathbf{u} \cdot F$.

Algorithm 1. Rigid numerical continuation, original version
Input: $F \in \mathcal{H}$, $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$ and $z \in \mathbb{P}^n$
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, $\mathbf{u}$, $\mathbf{v}$, z)
  $(\mathbf{w}_t)_{0 \leqslant t \leqslant T} \leftarrow$ a 1-Lipschitz continuous path from $\mathbf{v}$ to $\mathbf{u}$ in $\mathcal{U}$
  $t \leftarrow 0$
  while true **do**
    for i from 1 to n **do**
      $w \leftarrow i$th component of $\mathbf{w}_t$
      $g_i \leftarrow \gamma_{\mathrm{Frob}}(f_i \circ w^{-1}, z)$ ▷ See (1.3).
    end for
    $t \leftarrow t + \left(240\, \kappa(\mathbf{w}_t, z)^2 \left(\sum_{i=1}^n g_i^2\right)^{\frac{1}{2}}\right)^{-1}$ ▷ See (2.13) and (2.17).
    if $t \geqslant T$ **then**
      return z
    end if
    $z \leftarrow \mathrm{Newton}(\mathbf{w}_t \cdot F, z)$ ▷ Newton iteration
  end while
end function

Based on this bound, we obtained in Part I the following average analysis. Let $F = (f_1, \ldots, f_n)$ be a random polynomial system. We only assume that the distribution of F is *unitary invariant*: for any $\mathbf{u} \in \mathcal{U}$, the system $\mathbf{u} \cdot F$ has the same distribution as F. This property holds, for example, when F is a random Gaussian system, or when $F = \mathbf{v} \cdot G$ for a fixed system G and a random uniformly distributed $\mathbf{v} \in \mathcal{U}$. Under this hypothesis, the number K of continuation steps performed by Algorithm 1 satisfies $\mathbb{E}[K] \leqslant 9000 n^3\, \mathbb{E}[\Gamma(F)^2]^{\frac{1}{2}}$ (Theorem I.27, with $\mathfrak{g}_i = \gamma_{\mathrm{Frob}}$ and $C' = 5$, according to Lemma I.31). In case we cannot compute γ_{Frob} exactly, but instead an upper bound A such that $\gamma_{\mathrm{Frob}} \leqslant A \leqslant M\gamma_{\mathrm{Frob}}$, for some fixed $M \geqslant 1$, the algorithm works as well, but the bound on the average number of continuation steps is multiplied by M (see Remark I.28):

$$\mathbb{E}[K] \leqslant 9000 n^3 M\, \Gamma(F). \tag{2.19}$$

To obtain an interesting complexity result for a given class of unitary invariant distributions, based on numerical continuation along rigid paths and Inequality (2.19), we need, firstly, to specify how to compute or approximate γ_{Frob} at a reasonable cost, and secondly, to estimate $\mathbb{E}[\Gamma(F)^2]^{\frac{1}{2}}$. For the application to dense Gaussian systems, considered in Part I, γ_{Frob} is computed directly, using the monomial representation of the system to compute all higher derivatives, and the estimation of $\Gamma(F)$ is mostly standard. Using the monomial representation is not efficient anymore in the black-box model. We will rely instead on a probabilistic estimation of γ_{Frob}, within a factor $\mathrm{poly}(n, \delta)$. However, this estimation may fail with small probability, compromising the correctness of the result.

3. Fast numerical continuation for black-box functions

3.1. **Weyl norm.** We recall here how to characterize the Weyl norm of a homogeneous polynomial as an expectation, which is a key observation behind algorithm GammaProb to approximate $\gamma_{\mathrm{Frob}}(f, z)$ by random sampling.

Let $f \in \mathbb{C}[z_0, \ldots, z_n]$ be a homogeneous polynomial of degree $\delta > 0$. In the monomial basis, f decomposes as $\sum_\alpha c_\alpha z^\alpha$, where $\alpha = (\alpha_0, \ldots, \alpha_n)$ is a multi-index. The Weyl norm of f is defined as

$$\|f\|_W^2 \doteq \sum_\alpha \frac{\alpha_0! \cdots \alpha_n!}{\delta!} |c_\alpha|^2 . \tag{3.1}$$

The following statement seems to be classical.

Lemma 3.1. *Let f be a homogeneous polynomial of degree δ.*

(i) For a uniformly distributed w in the Euclidean unit ball of $\mathbb{C}^{n+1}$ we have

$$\|f\|_W^2 = \binom{n+1+\delta}{\delta} \mathbb{E}\left[|f(w)|^2\right].$$

(ii) For a uniformly distributed z in the unit sphere of $\mathbb{C}^{n+1}$ we have

$$\|f\|_W^2 = \binom{n+\delta}{\delta} \mathbb{E}\left[|f(z)|^2\right].$$

Proof. Let H be the space of homogeneous polynomials of degree δ in $z_0, \ldots, z_n$. Both left-hand and right-hand sides of the first stated equality define a norm on H coming from a Hermitian inner product. The monomial basis is orthogonal for both: this is obvious for Weyl's norm. For the L^2-norm, this is (Rudin 1980, Proposition 1.4.8). So it only remains to check that the claim holds true when f is a monomial. By (Rudin 1980, Proposition 1.4.9(2)), if $w^\alpha = w_0^{\alpha_0} \cdots w_n^{\alpha_n}$ is a monomial of degree δ, we have

$$\mathbb{E}\left[|w^\alpha|^2\right] = \frac{(n+1)!\alpha_0! \cdots \alpha_n!}{(n+1+\delta)!} = \frac{(n+1)!\delta!}{(n+1+\delta)!} \cdot \frac{\alpha_0! \cdots \alpha_n!}{\delta!}, \tag{3.2}$$

$$= \tbinom{n+1+\delta}{\delta}^{-1} \|w^\alpha\|_W^2 . \tag{3.3}$$

which is the claim. The second equality follows similarly from (Rudin 1980, Proposition 1.4.9(1)). □

The following inequalities will also be useful.

Lemma 3.2. *For any homogeneous polynomial $f \in \mathbb{C}[z_0, \ldots, z_n]$ of degree δ,*

$$\binom{n+\delta}{\delta}^{-1} \|f\|_W^2 \leqslant \max_{z \in \mathbb{S}(\mathbb{C}^{n+1})} |f(z)|^2 = \max_{w \in B(\mathbb{C}^{n+1})} |f(w)|^2 \leqslant \|f\|_W^2.$$

Proof. The first inequality follows directly from the second equality of Lemma 3.1. It is clear that the maximum is reached on the boundary. For the second inequality, we may assume (because of the unitary invariance of $\| - \|_W$) that the maximum of $|f|$ on the unit ball is reached at $(1, 0, \ldots, 0)$. Besides, the coefficient $c_{\delta,0,\ldots,0}$ of f is $f(1, 0, \ldots, 0)$. Therefore,

$$\max_{w \in B} |f(w)|^2 = |f(1, 0, \ldots, 0)|^2 = |c_{\delta,0,\ldots,0}|^2 \leqslant \|f\|_W^2. \qquad \square$$

3.2. **Probabilistic evaluation of the gamma number.** The main reason for introducing the Frobenius norm in the γ number, instead of the usual operator norm, is the equality (Lemma I.30)

$$\frac{1}{k!} \left\|\mathrm{d}_z^k f\right\|_{\mathrm{Frob}} = \|f(z + \bullet)_k\|_W, \tag{3.4}$$

Algorithm 2. Probabilistic estimation of γ_{Frob}

Input: $f \in \mathbb{C}[x_0, \dots, x_n]$ of degree $\leqslant \delta$, given as black-box evaluation program, $z \in \mathbb{C}^{n+1}$, and $\varepsilon > 0$

Output: $\Gamma \in \mathbb{R}$

Postcondition: $\gamma_{\mathrm{Frob}}(f, z) \leqslant \Gamma \leqslant 192\, n^2 \delta\, \gamma_{\mathrm{Frob}}(f, z)$ with probability at least $1 - \varepsilon$.

function GammaProb(f, z, ε)
  $h \leftarrow f(z + \bullet)$ (as black-box evaluation program)
  $s \leftarrow \left\lceil 1 + \log_2 \frac{\delta}{\varepsilon} \right\rceil$
  for i from 1 to s **do**
    $w_i \leftarrow$ random uniformly distributed element of B (unit ball of $\mathbb{C}^{n+1}$)
    compute $h_2(w_i), \dots, h_{\deg f}(w_i)$, where h_k is the degree k component of h
    ▷ Lemma 3.4
  end for
  compute $\mathrm{d}_0 h$
  return $\max\limits_{2 \leqslant k \leqslant \delta} \left(\dfrac{(32nk)^k}{\|\mathrm{d}_0 h\|^2} \cdot \binom{n+1+k}{k} \dfrac{1}{s} \displaystyle\sum_{i=1}^{s} |h_k(w_i)|^2 \right)^{\frac{1}{2k-2}}$.
end function

where $\|f(z + \bullet)_k\|_W$ is the Weyl norm of the homogeneous component of degree k of the shifted polynomial $x \mapsto f(z + x)$. It follows that

$$\gamma_{\mathrm{Frob}}(f, z) = \sup_{k \geqslant 2} \left(\|\mathrm{d}_z f\|^{-1} \, \|f(z + \bullet)_k\|_W \right)^{\frac{1}{k-1}}. \tag{3.5}$$

This equality opens up interesting ways for estimating γ_{Frob}, and therefore γ. We used it to compute γ_{Frob} efficiently when f is a dense polynomial given in the monomial basis, see §I.4.3.3. In that context, we would compute the shift $f(z + \bullet)$ in the same monomial basis in quasilinear time as $\min(n, \delta) \to \infty$. From there, the quantities $\|f(z + \bullet)_k\|_W$ can be computed in linear time. In the black-box model, however, the monomial expansions (of either f or $f(z + \bullet)$) cannot fit into a $\mathrm{poly}(n, \delta) L(f)$ complexity bound, because the number of monomials of degree δ in $n + 1$ variables is not $\mathrm{poly}(n, \delta)$. Nonetheless, we can obtain a good enough approximation of $\|f(z + \bullet)_k\|_W$ with a few evaluations but a nonzero probability of failure. This is the purpose of Algorithm 2, which we analyze in the next theorem.

Theorem 3.3. *Given $f \in \mathbb{C}[x_0, \dots, x_n]$ as a black-box function, an upper bound δ on its degree, a point $z \in \mathbb{C}^{n+1}$, and some $\varepsilon > 0$, algorithm* GammaProb *computes some $\Gamma \geqslant 0$ such that*

$$\gamma_{\mathrm{Frob}}(f, z) \leqslant \Gamma \leqslant 192 n^2 \delta \cdot \gamma_{\mathrm{Frob}}(f, z)$$

with probability at least $1 - \varepsilon$, using $O\left(\delta \log\left(\frac{\delta}{\varepsilon}\right)(L(f) + n + \log \delta)\right)$ operations.

Moreover, for any $t \geqslant 1$,

$$\mathbb{P}\left[\Gamma \leqslant \frac{\gamma_{\mathrm{Frob}}(f, z)}{t} \right] \leqslant \varepsilon^{1 + \frac{1}{2} \log_2 t}.$$

Note that we currently do not know how to estimate γ_{Frob} within an arbitrarily small factor. The key in Theorem 3.3 is to write each $\|f(z + \bullet)_k\|_W^2$ as an expectation (this is classical, see §3.1) and to approximate it by sampling (there are some obstacles). We assume that $z = 0$ by changing f to $f(z + \bullet)$, which is harmless because the evaluation complexity is changed to $L(f) + O(n)$. Furthermore, the homogeneous components f_k of f are accessible as black-box functions; this is the content of the next lemma.

Lemma 3.4. *Given $w \in \mathbb{C}^{n+1}$, one can compute $f_0(w), \dots, f_\delta(w)$, with $O(\delta(L(f) + n + \log \delta))$ arithmetic operations.*

Proof. We first compute all $f(\xi^i w)$, for $0 \leqslant i \leqslant \delta$ for some primitive root of unity ξ of order $\delta + 1$. This takes $(\delta + 1)L(f) + O(\delta n)$ arithmetic operations. Since

$$f(\xi^i w) = \sum_{k=0}^{\delta} \xi^{ik} f_k(w), \tag{3.6}$$

we recover the numbers $f_k(w)$ with the inverse Fourier transform,

$$f_k(w) = \frac{1}{\delta + 1} \sum_{i=0}^{\delta} \xi^{-ik} f(\xi^i w). \tag{3.7}$$

We may assume that δ is a power of two (δ is only required to be an upper bound on the degree of f), and the fast Fourier transform algorithm has an $O(\delta \log \delta)$ complexity bound to recover the $f_k(z)$. (With slightly more complicated formulas, we can also use $\xi = 2$ to keep close to the pure BSS model.) □

We now focus on the probabilistic estimation of $\|f_k\|_W$ via a few evaluations of f_k. Let $B \doteq B(\mathbb{C}^{n+1})$ denote the Euclidean unit ball in $\mathbb{C}^{n+1}$ and let $w \in B$ be a uniformly distributed random variable. By Lemma 3.1 we have

$$\|f_k\|_W^2 = \binom{n+1+k}{k} \mathbb{E}\left[|f_k(w)|^2\right]. \tag{3.8}$$

The expectation in the right-hand side can be estimated with finitely many samples of $|f_k(w)|^2$. To obtain a rigorous confidence interval, we study some statistical properties of $|f_k(w)|^2$. Let $w_1, \dots, w_s$ be independent uniformly distributed variables in B, and let

$$\hat{\mu}_k^2 \doteq \frac{1}{s} \sum_{i=1}^{s} |f_k(w_i)|^2 \tag{3.9}$$

denote their *empirical mean*. Let $\mu_k^2 \doteq \mathbb{E}[|f_k(w)|^2] = \mathbb{E}[\hat{\mu}_k^2]$ be the mean that we want to estimate. (Note that both μ_k and $\hat{\mu}_k$ depend on f_k; we supressed this dependence in the notation.)

The next proposition shows that $\hat{\mu}_k^2$ estimates μ_k^2 within a $\mathrm{poly}(n,k)^k$ factor with very few samples. The upper bound is obtained by a standard concentration inequality (Hoeffding's inequality). The lower bound is more difficult, and very specific to the current setting, because we need to bound μ_k^2 away from zero with only a small number of samples. Concentration inequalities do not apply because the standard deviation may be larger than the expectation, so a confidence interval whose radius is comparable to the standard deviation (which is what we can hope for with a small number of samples) may contain negative values.

Proposition 3.5. *For any $0 \leqslant k \leqslant \delta$, we have, with probability at least $1 - 2^{1-s}$,*

$$(32nk)^{-k} \mu_k^2 \leqslant \hat{\mu}_k^2 \leqslant (6n)^k \mu_k^2,$$

where s is the number of samples.

Before proceeding with the proof, we state two lemmas, the principle of which comes from Ji, Kollar, and Shiffman (1992, Lemma 8).

Lemma 3.6. *Let $g \in \mathbb{C}[z]$ be a univariate polynomial of degree k and let $c \in \mathbb{C}$ be its leading coefficient. For any $\eta > 0$,*

$$\operatorname{vol}\left\{z \in \mathbb{C} \;\middle|\; |g(z)|^2 \leqslant \eta\right\} \leqslant \pi k \left(|c|^{-2}\eta\right)^{\frac{1}{k}}.$$

Proof. Let $u_1, \ldots, u_k \in \mathbb{C}$ be the roots of g, with multiplicities, so that

$$g(z) = c(z - u_1)\cdots(z - u_k). \tag{3.10}$$

The distance of some $z \in \mathbb{C}$ to the set $S \doteq \{u_1, \ldots, u_k\}$ is the minimum of all $|z - u_i|$. In particular

$$\operatorname{dist}(z, S)^k \leqslant \prod_{i=1}^{k} |z - u_i| = |c|^{-1} |g(z)|. \tag{3.11}$$

Therefore,

$$\left\{z \in \mathbb{C} \;\middle|\; |g(z)|^2 \leqslant \eta\right\} \subset \bigcup_{i=1}^{k} B\left(u_i, |c|^{-\frac{1}{k}} \eta^{\frac{1}{2k}}\right), \tag{3.12}$$

where $B(u_i, r) \subseteq \mathbb{C}$ is the disk of radius r around u_i. The volume of $B(u_i, r)$ is πr^2, so the claim follows directly. □

Lemma 3.7. *If $w \in B$ is a uniformly distributed random variable, then for all $\eta > 0$,*

$$\mathbb{P}\left[|f_k(w)|^2 \leqslant \eta \max_{\mathbb{S}} |f_k|^2\right] \leqslant (n+1)k\eta^{\frac{1}{k}},$$

where $\max_{\mathbb{S}} |f_k|$ is the maximum value of $|f_k|$ on the unit sphere in $\mathbb{C}^{n+1}$.

Proof. Let c be the coefficient of x_n^k in f_k. It is the value of f_k at $(0, \ldots, 0, 1)$. Up to a unitary change of coordinates, $|f_k|$ reaches a maximum at $(0, \ldots, 0, 1)$ so that $c = \max_{\mathbb{S}} |f_k|$. Up to scaling, we may further assume that $c = 1$. For any $(p_0, \ldots, p_{n-1}) \in \mathbb{C}^n$,

$$\operatorname{vol}\left\{z \in \mathbb{C} \;\middle|\; |f_k(p_0, \ldots, p_{n-1}, z)|^2 \leqslant \eta\right\} \leqslant \pi k \eta^{1/k}, \tag{3.13}$$

by Lemma 3.6 applied to the polynomial $g(z) = f_k(p_0, \ldots, p_{n-1}, z)$, which, by construction, is monic. It follows, from the inclusion $B(\mathbb{C}^{n+1}) \subseteq B(\mathbb{C}^n) \times \mathbb{C}$, that

$$\operatorname{vol}\left\{w \in B(\mathbb{C}^{n+1}) \;\middle|\; |f_k(w)|^2 \leqslant \eta\right\} \tag{3.14}$$

$$\leqslant \operatorname{vol}\left\{(p_0, \ldots, p_{n-1}, z) \in B(\mathbb{C}^n) \times \mathbb{C} \;\middle|\; |f_k(p_0, \ldots, p_{n-1}, z)|^2 \leqslant \eta\right\} \tag{3.15}$$

$$\leqslant \operatorname{vol} B(\mathbb{C}^n) \cdot \pi k \eta^{\frac{1}{k}}. \tag{3.16}$$

Using $\operatorname{vol} B(\mathbb{C}^n) = \frac{\pi^n}{n!}$ and dividing both sides by $\operatorname{vol} B(\mathbb{C}^{n+1})$ concludes the proof. □

Lemma 3.8. *For any $\eta > 0$, we have*

$$\mathbb{P}\left[\hat{\mu}_k^2 \leqslant \eta \mu_k^2\right] \leqslant \left(8nk\eta^{\frac{1}{k}}\right)^{\frac{s}{2}}.$$

Proof. Put $M \doteq \max_{\mathbb{S}} |f_k|$. If $\hat{\mu}_k^2 \leqslant \eta M^2$ then at least $\lceil s/2 \rceil$ samples among $|f(w_1)|^2, \ldots, |f(w_s)|^2$ satisfy $|f(w_i)|^2 \leqslant 2\eta M^2$. By the union bound and Lemma 3.7 we obtain,

$$\mathbb{P}\left[\hat{\mu_k}^2 \leqslant \eta M^2\right] \leqslant \binom{s}{\lceil s/2 \rceil} \mathbb{P}\left[|f(w)|^2 \leqslant 2\eta M^2\right]^{\lceil s/2 \rceil} \tag{3.17}$$

$$\leqslant 2^s \left((n+1)k\eta^{\frac{1}{k}}\right)^{\frac{s}{2}} \tag{3.18}$$

$$\leqslant \left(8nk\eta^{\frac{1}{k}}\right)^{\frac{s}{2}}. \tag{3.19}$$

To conclude, we note that $\mu_k \leqslant M$. □

Proof of Proposition 3.5. With $\eta \doteq (32nk)^{-k}$, Lemma 3.8 gives

$$\mathbb{P}\left[\hat{\mu}_k^2 \leqslant \eta\mu_k^2\right] \leqslant \left(8nk\eta^{\frac{1}{k}}\right)^{\frac{s}{2}} = 2^{-s}. \tag{3.20}$$

It follows that

$$\mathbb{P}\left[\mu_k^2 \leqslant (32nk)^k\, \hat{\mu}_k^2\right] \geqslant 1 - 2^{-s}, \tag{3.21}$$

which is the stated left-hand inequality.

For the right-hand inequality, we apply Hoeffding's inequality (e.g., Boucheron, Lugosi, and Massart 2013, Theorem 2.8). The variable $s\hat{\mu}_k^2$ is a sum of s independent variables lying in the interval $[0, M^2]$, where we again abbreviate $M \doteq \max_{\mathbb{S}} |f_k|$. Accordingly, for any $C \geqslant 1$,

$$\mathbb{P}\left[\hat{\mu}_k^2 \geqslant C\mu_k^2\right] = \mathbb{P}\left[s\hat{\mu}_k^2 - s\mu_k^2 \geqslant (C-1)s\mu_k^2\right] \leqslant \exp\left(-\frac{2(C-1)^2 s^2 \mu_k^4}{sM^4}\right). \tag{3.22}$$

By Lemma 3.2 combined with (3.8), we have

$$M^2 \leqslant \binom{n+1+k}{k}\mu_k^2. \tag{3.23}$$

Applying this bound, we obtain

$$\mathbb{P}\left[\hat{\mu}_k^2 \geqslant C\mu_k^2\right] \leqslant \exp\left(-\frac{2(C-1)^2 s}{\binom{n+1+k}{k}^2}\right). \tag{3.24}$$

We choose $C = (6n)^k$ and simplify further using the inequality $\binom{m+k}{k} \leqslant \frac{(m+k)^k}{k!} \leqslant (e(m+k)/k)^k$ and $e(n+1+k)/k \leqslant e(n+3)/2$ (use $k \geqslant 2$) to obtain

$$\frac{C-1}{\binom{n+1+k}{k}} \geqslant \frac{(6n)^k - 1}{\left(\frac{e(n+3)}{2}\right)^k} \geqslant \left(\frac{12n}{e(n+3)}\right)^k - \left(\frac{2}{e(n+3)}\right)^k \tag{3.25}$$

$$\geqslant \left(\frac{3}{e}\right)^2 - \left(\frac{2}{4e}\right)^2 \geqslant \sqrt{\frac{1}{2}\log 2}. \tag{3.26}$$

We obtain therefore

$$\mathbb{P}\left[\hat{\mu}_k^2 \geqslant (6n)^k \mu_k^2\right] \leqslant \exp(-\log(2)s) = 2^{-s}. \tag{3.27}$$

Combined with (3.21), the union bound implies

$$\mathbb{P}\left[\hat{\mu}_k^2 \leqslant (32nk)^{-k}\mu_k^2 \text{ or } \hat{\mu}_k^2 \geqslant (6n)^k \mu_k^2\right] \leqslant 2 \cdot 2^{-s} \tag{3.28}$$

and the proposition follows. □

Proof of Theorem 3.3. Recall that we assume that $z = 0$. Proposition 3.5 can be rephrased as follows: with probability at least $1 - 2^{1-s}$, we have

$$\mu_k^2 \leqslant (32nk)^k \hat{\mu}_k^2 \leqslant (192n^2k)^k \mu_k^2 \tag{3.29}$$

Defining

$$c_k^2 \doteq \binom{n+1+k}{k}(32nk)^k \hat{\mu}_k^2, \tag{3.30}$$

using that by (3.8)

$$\|f_k\|_W^2 = \binom{n+1+k}{k}\mu_k^2, \tag{3.31}$$

and applying the union bound, we therefore see that

$$\|f_k\|_W^2 \leqslant c_k^2 \leqslant (192n^2k)^k \cdot \|f_k\|_W^2 \tag{3.32}$$

holds for all $2 \leqslant k \leqslant \delta$, with probability at least $1 - \delta 2^{1-s}$. If we chose $s = \left\lceil 1 + \log_2 \frac{\delta}{\varepsilon} \right\rceil$, then $\delta 2^{1-s} \leqslant \varepsilon$. Recall from (3.4) and (3.5) that

$$\gamma_{\mathrm{Frob}}(f,z) = \max_{\delta \geqslant k \geqslant 2} \left(\|\mathrm{d}_0 f\|^{-1} \|f_k\|_W \right)^{\frac{1}{k-1}}. \tag{3.33}$$

Noting that $(192n^2k)^{\frac{k}{k-1}} \leqslant (192n^2\delta)^2$, for $2 \leqslant k \leqslant \delta$, we conclude that the random variable

$$\Gamma \doteq \max_{2 \leqslant k \leqslant \delta} \left(\|\mathrm{d}_0 f\|^{-1} c_k \right)^{\frac{1}{k-1}}, \tag{3.34}$$

which is returned by Algorithm 2, indeed satisfies

$$\gamma_{\mathrm{Frob}}(f,z) \leqslant \Gamma \leqslant 192n^2\delta \cdot \gamma_{\mathrm{Frob}}(f,z) \tag{3.35}$$

with probability at least $1 - \varepsilon$, which proves the first assertion.

For the assertion on the number of operations, it suffices to note that by Lemma 3.4, the computation of $d_0 f$ and of $\hat{\mu}_2, \ldots, \hat{\mu}_\delta$ can be done with $O(s\delta(L(f) + n + \log \delta))$ arithmetic operations.

It only remains to check, for any $t \geqslant 1$, the tail bound

$$\mathbb{P}\left[\Gamma \leqslant \frac{\gamma_{\mathrm{Frob}}(f,z)}{t} \right] \leqslant \varepsilon^{1 + \frac{1}{2}\log_2 t}. \tag{3.36}$$

Unfolding the definitions (3.33) and (3.34) and using again (3.31), we obtain

$$\mathbb{P}\left[\Gamma \leqslant \frac{\gamma_{\mathrm{Frob}}(f,z)}{t} \right] \leqslant \sum_{k=2}^{\delta} \mathbb{P}\left[(32nk)^k \hat{\mu}_k^2 \leqslant t^{-2(k-1)} \mu_k^2 \right] \tag{3.37}$$

$$\leqslant \sum_{k=2}^{\delta} \left(8nk \cdot \left((32nk)^k t^{2(k-1)} \right)^{-\frac{1}{k}} \right)^{\frac{s}{2}}, \quad \text{by Lemma 3.8,} \tag{3.38}$$

$$= \sum_{k=2}^{\delta} \left(\frac{1}{4} t^{-2\frac{k-1}{k}} \right)^{\frac{s}{2}} \leqslant \delta 2^{-s} t^{-\frac{s}{2}}. \tag{3.39}$$

Since $s = \left\lceil 1 + \log_2 \frac{\delta}{\varepsilon} \right\rceil$, we have $\delta 2^{-s} \leqslant \varepsilon$. Furthermore, $s \geqslant -\log_2 \varepsilon$, so

$$t^{-\frac{s}{2}} \leqslant t^{\frac{1}{2}\log_2 \varepsilon} = \varepsilon^{\frac{1}{2}\log_2 t}, \tag{3.40}$$

which proves (3.36). □

3.3. **A Monte-Carlo continuation algorithm.** We deal here with the specifics of a numerical continuation with a step-length computation that may be wrong.

The probabilistic algorithm for the evaluation of the step length can be plugged into the rigid continuation algorithm (Algorithm 1). There is no guarantee, however, that the randomized computations of the γ_{Frob} fall within the confidence interval described in Theorem 3.3 and, consequently, there is no guarantee that the corresponding step-length estimation is accurate. If step lengths are underestimated, we don't control anymore the complexity: as the step lengths go to zero, the number of steps goes to infinity. Overestimating a single step length, instead, may undermine the correctness of the result, and the subsequent behavior of the algorithm is

Algorithm 3. Bounded-time numerical continuation routine for black-box input
Input: $F \in \mathcal{H}$ (given as black-box), $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$, $z \in \mathbb{P}^n$, $K_{\max} > 0$ and $\varepsilon > 0$
Precondition: z is a zero of $\mathbf{v} \cdot F$.
Output: $w \in \mathbb{P}^n$ or FAIL.
Postcondition: If some $w \in \mathbb{P}^n$ is output then w is an approximate zero of $\mathbf{u} \cdot F$ with probability $\geqslant 1 - \varepsilon$.

function BOUNDEDBLACKBOXNC(F, $\mathbf{u}$, $\mathbf{v}$, z, $K_{\max}$, ε)
  $\eta \leftarrow (nK_{\max})^{-1}\varepsilon$
  $(\mathbf{w}_t)_{0 \leqslant t \leqslant T} \leftarrow$ a 1-Lipschitz continuous path from $\mathbf{v}$ to $\mathbf{u}$ in $\mathcal{U}$
  $t \leftarrow 0$
  for k from 1 to $K_{\max}$ **do**
    for i from 1 to n **do**
      $w \leftarrow i$th component of $\mathbf{w}_t$
      $g_i \leftarrow$ GAMMAPROB$(f_i \circ w^{-1}, z, \eta)$ ▷ Algorithm 2
    end for
    $t \leftarrow t + \left(240\,\kappa(\mathbf{w}_t, z)^2 \left(\sum_{i=1}^n g_i^2\right)^{\frac{1}{2}}\right)^{-1}$
    if $t \geqslant T$ **then**
      return z
    end if
    $z \leftarrow \mathrm{Newton}(\mathbf{w}_t \cdot F, z)$ ▷ Newton iteration
  end for
  return FAIL
end function

unknown (it may even not to terminate). So we introduce a limit on the number of continuation steps. Algorithm 3 is a corresponding modification of Algorithm 1. When reaching the limit on the number of steps, this algorithm halts with a failure notification.

Proposition 3.9. *On input F, $\mathbf{u}$, $\mathbf{v}$, z, $K_{\max}$, and ε, such that z is a zero of $\mathbf{v} \cdot F$, Algorithm* BOUNDEDBLACKBOXNC *either fails or returns some $w \in \mathbb{P}^n$. In the latter case, w is an approximate zero of $\mathbf{u} \cdot F$ with probability at least $1 - \varepsilon$. The total number of operations is* $\mathrm{poly}(n, \delta) \cdot K_{\max} \log\left(K_{\max}\varepsilon^{-1}\right) \cdot L(F)$.

Proof. Assume $w \in \mathbb{P}^n$ is returned which is not an approximate zero of F. This implies that one of the estimations of $\gamma_{\mathrm{Frob}}(f, z)$, computed by the GAMMAPROB subroutines yielded a result that is smaller than the actual value of $\gamma_{\mathrm{Frob}}(f, z)$. There are at most $nK_{\max}$ such estimations, so by Theorem 3.3, this happens with probability at most $nK_{\max}\eta$, which by choice of η is exactly ε.

The total number of operations is bounded by $K_{\max}$ times the cost of an iteration. The cost of an iteration is dominated by the evaluation of the g_i, which is bounded by $O(\delta \log(\delta n K_{\max}\varepsilon^{-1})(L(F) + n + \log\delta))$ by Theorem 3.3 and the choice of η, and the Newton iteration, which costs $\mathrm{poly}(n, \delta)L(F)$. □

In case Algorithm 3 fails, it is natural to restart the computation with a higher iteration limit. This is Algorithm 4. We can compare its complexity to that of Algorithm 1, which assumes an exact computation of γ. Let $K(F, \mathbf{u}, \mathbf{v}, z)$ be a bound for the number of iterations performed by Algorithm 1 on input F, $\mathbf{u}$, $\mathbf{v}$ and z, allowing an overestimation of the step length up to a factor $192n^2\delta$ (in view of Theorem 3.3).

Algorithm 4. Numerical continuation for black-box input
Input: $F \in \mathcal{H}$ (given as black-box), $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$, $z \in \mathbb{P}^n$ and $\varepsilon \in (0, \frac{1}{4}]$.
Precondition: z is a zero of $\mathbf{v} \cdot F$.
Output: $w \in \mathbb{P}^n$.
Postcondition: w is an approximate zero of $\mathbf{u} \cdot F$ with probability $\geqslant 1 - \varepsilon$.

function BlackBoxNC(F, $\mathbf{u}$, $\mathbf{v}$, z, ε)
  $K_{\max} \leftarrow 1$
  repeat
    $K_{\max} \leftarrow 2K_{\max}$
    $w \leftarrow$ BoundedBlackBoxNC $(F, \mathbf{u}, \mathbf{v}, z, K_{\max}, \varepsilon)$ ▷ Algorithm 3
  until $w \neq$ Fail
  return w
end function

Proposition 3.10. *On input F, $\mathbf{u}$, $\mathbf{v}$, z and $\varepsilon \in (0, \frac{1}{4}]$, such that z is a zero of $\mathbf{v} \cdot F$, and $K(F, \mathbf{u}, \mathbf{v}, z) < \infty$, Algorithm 4 terminates almost surely and returns an approximate zero of $\mathbf{u} \cdot F$ with probability at least $1 - \varepsilon$. The average total number of operations is* $\operatorname{poly}(n, \delta) \cdot L(F) \cdot K \log\left(K\varepsilon^{-1}\right)$, *with $K = K(F, \mathbf{u}, \mathbf{v}, z)$.*

Proof. Let $K \doteq K(F, \mathbf{u}, \mathbf{v}, z)$. By definition of K, if all approximations lie in the desired confidence interval, then BoundedBlackBoxNC terminates after at most K iterations. So as soon as $K_{\max} \geqslant K$, BoundedBlackBoxNC may return Fail only if the approximation of some γ_{Frob} is not correct. This happens with probability at most ε at each iteration of the main loop in Algorithm 4, independently. So the number of iterations is finite almost surely. That the result is correct with probability at least $1 - \varepsilon$ follows from Proposition 3.9.

We now consider the total cost. At the mth iteration, we have $K_{\max} = 2^m$, so the cost of the mth iteration is $\operatorname{poly}(n, \delta) \cdot 2^m \log(2^m \varepsilon^{-1}) \cdot L(f)$, by Proposition 3.9. Put $\ell \doteq \lceil \log_2 K \rceil$. If the mth iteration is reached for some $m > \ell$, then all the iterations from ℓ to $m - 1$ have failed. This has a probability $\leqslant \varepsilon^{m-\ell}$ to happen, so, if I denotes the number of iterations, we have

$$\mathbb{P}[I \geqslant m] \leqslant \min\left(1, \varepsilon^{m-\ell}\right). \tag{3.41}$$

The total expected cost is therefore bounded by

$$\mathbb{E}[\text{cost}] \leqslant \operatorname{poly}(n, \delta) L(f) \sum_{m=1}^{\infty} 2^m \log(2^m \varepsilon^{-1}) \mathbb{P}[I \geqslant m] \tag{3.42}$$

$$\leqslant \operatorname{poly}(n, \delta) L(f) \sum_{m=1}^{\infty} 2^m \log(2^m \varepsilon^{-1}) \min\left(1, \varepsilon^{m-\ell}\right). \tag{3.43}$$

The claim follows easily from splitting the sum into two parts, $1 \leqslant m < \ell$ and $m > \ell$, and applying the bounds (with $c = \log \varepsilon^{-1}$)

$$\sum_{m=1}^{\ell-1} 2^m (m + c) \leqslant (\ell + c) 2^\ell \tag{3.44}$$

and, for $\varepsilon \in (0, \frac{1}{4})$,

$$\sum_{m=\ell}^{\infty} 2^m (m + c) \varepsilon^{m-\ell} \leqslant \frac{(\ell + c) 2^\ell}{(1 - 2\varepsilon)^2} \leqslant 4(\ell + c) 2^\ell. \tag{3.45}$$

□

Algorithm 5. Zero finding for black-box input
Input: $F \in \mathcal{H}$ (given as black-box) and $\varepsilon \in (0, \frac{1}{4}]$.
Output: $w \in \mathbb{P}^n$.
Postcondition: w is an approximate zero of F with probability $\geqslant 1 - \varepsilon$.

function BlackBoxSolve(F, ε)
  Sample $(\mathbf{v}, z)$ in the rigid solution variety of F ▷ Algorithm I.1
  return BlackBoxNC $(F, \mathbf{1}_{\mathcal{U}}, \mathbf{v}, z, \varepsilon)$ ▷ Algorithm 4
end function

4. Condition based complexity analysis

We recall from Part I (§2) the *rigid solution variety* corresponding to a polynomial system $F = (f_1, \ldots, f_n)$, which consists of the pairs $(\mathbf{u}, z) \in \mathcal{U} \times \mathbb{P}^n$ such that $(\mathbf{u} \cdot F)(z) = 0$, which means $f_1(u^{-1}z) = 0, \ldots, f_n(u^{-1}z) = 0$. To solve a given polynomial system $F \in \mathcal{H}$, we sample an initial pair in the rigid solution variety corresponding to F (Algorithm I.1) and perform a numerical continuation using Algorithm 4. This gives Algorithm 5.

Theorem 4.1. *On input $F \in \mathcal{H}$ (given as a black-box evaluation program) with only regular zeros, and $\varepsilon > 0$, Algorithm* BlackBoxSolve *computes an approximate zero of F with probability at least $1 - \varepsilon$.*

If $\mathbf{u} \in \mathcal{U}$ is uniformly distributed, then on input $\mathbf{u} \cdot F$ and ε, Algorithm BlackBoxSolve *performs*

$$\operatorname{poly}(n, \delta) \cdot L(F) \cdot \Gamma(F)\big(\log \Gamma(F) + \log \varepsilon^{-1}\big)$$

operations on average.

Termination and correctness (with probability at least $1 - \varepsilon$), are clear by Proposition 3.10. We next focus on proving the complexity bound. Note that the statement of Theorem 4.1 is similar to that of Theorem 1.1. The only difference lies in the complexity bound, whose dependence on ε^{-1} is logarithmic in the former and doubly logarithmic in the latter.

4.1. Complexity of sampling the rigid solution variety.

Toward the proof of Theorems 1.1 and 4.1, we first review the complexity of sampling the initial pair for the numerical continuation. In the rigid setting, this sampling boils down to sampling hypersurfaces, which in turn amounts to computing roots of univariate polynomials (see Part I, §2.4). Some technicalities are required to connect known results about root-finding algorithms to our setting, and especially the parameter $\Gamma(F)$, but the material is very classical.

Proposition 4.2. *Given $F \in \mathcal{H}$ as a black-box evaluation program, we can sample $\mathbf{v} \in \mathcal{U}$ and $\zeta \in \mathbb{P}^n$ such that $\mathbf{v}$ is uniformly distributed and ζ is a uniformly distributed zero of $\mathbf{v} \cdot F$, with $\operatorname{poly}(n, \delta) \cdot (L(F) + \log\log \Gamma(F))$ operations on average.*

Proof. This follows from Proposition I.10 and Proposition 4.3 below. □

Proposition 4.3. *For any $f \in \mathbb{C}[z_0, \ldots, z_n]$ homogeneous of degree $\delta \geqslant 2$, given as a black-box evaluation program, one can sample a uniformly distributed point in the zero set $V(f)$ of f by a probabilistic algorithm with $\operatorname{poly}(n, \delta) \cdot (L(f) + \log\log \Gamma(f))$ operations on average.*

Proof. Following Corollary I.9, we can compute a uniformly distributed zero of f by first sampling a line $\ell \subset \mathbb{P}^n$ uniformly distributed in the Grasmannian of lines, and

then sampling a uniformly distributed point in the finite set $\ell \cap V(f)$. To do this, we consider the restriction $f|_\ell$, which, after choosing a orthonormal basis of ℓ, is a bivariate homogeneous polynomial, and compute its roots. The representation of $f|_\ell$ in a monomial basis can be computed by $\delta+1$ evaluations of f and interpolation, at a cost $O\left(\delta(L(f)+n+\log\delta)\right)$, as in Lemma 3.4. By Lemma 4.4 below, computing the roots takes

$$\text{(4.1)} \qquad \operatorname{poly}(\delta)\log\log\left(\max_{\zeta\in\ell\cap V(f)}\gamma(f|_\ell,\zeta)\right)$$

operations on average. We assume a 6th type of node to refine approximate roots into exact roots (recall the discussion in §1.3). Then we have, by the definition (1.5) of $\Gamma(f|_\ell)$,

$$\text{(4.2)} \qquad \max_{\zeta\in\ell\cap V(f)}\gamma(f|_\ell,\zeta)^2 \leqslant \sum_{\zeta\in\ell\cap V(f)}\gamma_{\mathrm{Frob}}(f|_\ell,\zeta)^2 = \delta\,\Gamma(f|_\ell)^2.$$

Note that $\delta\Gamma(f|_\ell)^2 \geqslant \delta\frac{1}{4}(\delta-1)^2 \geqslant \frac{1}{2} \geqslant \frac{1}{e}$ since $\gamma(f|_\ell,\zeta)\geqslant\frac{1}{2}(\delta-1)$ by Lemma 11 of Part I. By Jensen's inequality, using the concavity of $\log\log$ on $[e^{-1},\infty)$, we obtain

$$\text{(4.3)} \qquad \mathbb{E}_\ell\left[\log\log\left(\delta\,\Gamma(f|_\ell)^2\right)\right] \leqslant \log\log\left(\delta\mathbb{E}_\ell\left[\Gamma(f|_\ell)^2\right]\right).$$

Finally, Lemma 4.5 below gives

$$\text{(4.4)} \qquad \log\log\left(\delta\mathbb{E}_\ell\left[\Gamma(f|_\ell)^2\right]\right) \leqslant \log\log\left(2n\delta\,\Gamma(f)^2\right)$$

and the claim follows. □

Lemma 4.4. *Let $g\in\mathbb{C}[z_0,z_1]$ be a homogeneous polynomial of degree δ without multiple zeros. One can compute, with a probabilistic algorithm, δ approximate zeros of g, one for each zero of g, with $\operatorname{poly}(\delta)\log\log\gamma_{\max}$ operations on average, where $\gamma_{\max}\doteq\max_{\zeta\in V(g)}\gamma(g,\zeta)$.*

Proof. The proof essentially relies on the following known fact due to Renegar (1987) (see also Pan 2001, Thm. 2.1.1 and Cor. 2.1.2, for tighter bounds). Let $f\in\mathbb{C}[t]$ be a given polynomial of degree δ, $R>0$ be a known upper bound on the modulus of the roots $\xi_1,\dots,\xi_\delta\in\mathbb{C}$ of f, and $\varepsilon>0$ be given. We can compute from this data with $\operatorname{poly}(\delta)\log\log\frac{R}{\varepsilon}$ operations approximations $x_1,\dots,x_n\in\mathbb{C}$ of the zeros such that $|\xi_i-x_i|\leqslant\varepsilon$.

To apply this result to the given homogeneous polynomial g, we first apply a uniformly random unitary transformation $u\in U(2)$ to the given g and dehomogenize $u\cdot g$, obtaining the univariate polynomial $f\in\mathbb{C}[t]$.

We first claim that with probability at least $3/4$ we have: $(*)$ $|\xi_i|\leqslant 2\sqrt{\delta}$ for all zeros $\xi_i\in\mathbb{C}$ of f. This can be seen as follows. We measure distances in $\mathbb{P}^1$ with respect to the projective (angular) distance. The disk of radius θ around a point in $\mathbb{P}^1$, has measure at most $\pi(\sin\theta)^2$ (Bürgisser and Cucker 2013, Lemma 20.8). Let $\sin\theta=(2\sqrt{\delta})^{-1}$. Then a uniformly random point p in $\mathbb{P}^1$ lies in a disk of radius θ around a root of f with probability at most $\delta(\sin\theta)^2\leqslant 1/4$. Write $0\doteq[1:0]$ and $\infty\doteq[0:1]$ and note that $\operatorname{dist}(0,p)+\operatorname{dist}(p,\infty)=\pi/2$ for any $p\in\mathbb{P}^1$. Since $u^{-1}(\infty)$ is uniformly distributed, we conclude that with probability at least $3/4$, each zero $\zeta\in\mathbb{P}^1$ of g satisfies $\operatorname{dist}(\zeta,u^{-1}(\infty))\geqslant\theta$, which means $\operatorname{dist}(\zeta,u^{-1}(0))\leqslant\pi/2-\theta$. The latter easily implies for the corresponding affine root $\xi=\zeta_1/\zeta_0$ of f that $|\xi|\leqslant(\tan\theta)^{-1}\leqslant(\sin\theta)^{-1}=2\sqrt{\delta}$, hence $(*)$ holds.

The maximum norm of a zero of $f\in\mathbb{C}[t]$ can be computed with a small relative error with $O(\delta\log\delta)$ operations (Pan 1996, Fact 2.2(b)), so we can test the property $(*)$. We repeatedly sample a new $u\in U(2)$ until $(*)$ holds. Each iteration

succeeds with probability at least $\frac{3}{4}$ of success, so there are at most two iterations on average.

For a chosen $\varepsilon > 0$, we can now compute with Renegar's algorithm the roots of f, up to precision ε with $\mathrm{poly}(\delta)\log\log\frac{1}{\varepsilon}$ operations (where the $\log\log 2\sqrt{\delta}$ is absorbed by $\mathrm{poly}(\delta)$). By homogeneizing and transforming back with u^{-1}, we obtain approximations $p_1,\dots,p_\delta$ of the projective roots $\zeta_1,\dots,\zeta_\delta$ of g up to precision ε, measured in projective distance.

The remaining difficulty is that the p_i might not be approximate roots of g, in the sense of Smale. However, suppose that for all i we have

$$\varepsilon\gamma(g,p_i) \leqslant \tfrac{1}{11}. \tag{4.5}$$

Using that $z \mapsto \gamma(g,z)^{-1}$ is 5-Lipschitz continuous on $\mathbb{P}^1$ (Lairez 2020, Lemma 31), we see that $\varepsilon\gamma(g,\zeta_i) \leqslant \frac{1}{6}$ for all i. This is known to imply that p_i is an approximate zero of p_i (Shub and Smale 1993c, and Theorem I.12 for the constant). On the other hand, using again the Lipschitz property, we are sure that Condition (4.5) is met as soon as $\varepsilon\gamma_{\max} \leqslant \frac{1}{16}$.

So starting with $\varepsilon = \frac{1}{2}$, we compute points $p_1,\dots,p_\delta$ approximating $\zeta_1,\dots,\zeta_\delta$ up to precision ε until (4.5) is met for all p_i, squaring ε after each unsuccessful iteration. Note that Renegar's algorithm need not be restarted when ε is refined. We have $\varepsilon\gamma_{\max} \leqslant \frac{1}{16}$ after at most $\log\log(16\gamma_{\max})$ iterations. Finally, note that we do not need to compute exactly γ, an approximation within factor 2 is enough, with appropriate modifications of the constants, and this is achieved by γ_{Frob}, see (1.4), which we can compute in $\mathrm{poly}(\delta)$ operations. □

Lemma 4.5. *Let $f \in \mathbb{C}[z_0,\dots,z_n]$ be homogeneous of degree δ and let $\ell \subset \mathbb{P}^n$ be a uniformly distributed random projective line. Then $\mathbb{E}_\ell\left[\Gamma(f|_\ell)^2\right] \leqslant 2n\,\Gamma(f)^2$.*

Proof. Let $\ell \subset \mathbb{P}^n$ be a uniformly distributed random projective line and let $\zeta \in \ell$ be uniformly distributed among the zeros of $f|_\ell$. Then ζ is also a uniformly distributed zero of f, see Corollary I.9. Let θ denote the angle between the tangent line $T_\zeta\ell$ and the line $T_\zeta V(f)^\perp$ normal to $V(f)$ at ζ. By an elementary geometric reasoning, we have $\|\mathrm{d}_\zeta f|_\ell\| = \|\mathrm{d}_\zeta f\|\cos\vartheta(\ell,\zeta)$. Moreover, $\|\mathrm{d}^k_\zeta f|_\ell\|_{\mathrm{Frob}} \leqslant \|\mathrm{d}^k_\zeta f\|_{\mathrm{Frob}}$. So it follows that

$$\gamma_{\mathrm{Frob}}(f|_\ell,\zeta)^2 \leqslant \gamma_{\mathrm{Frob}}(f,\zeta)^2\cos(\theta)^{-2}. \tag{4.6}$$

In order to bound this, we consider now a related, but different distribution. As above, let ζ be a uniformly distributed zero of f. Consider now a uniformly distributed random projective line ℓ' passing through ζ. The two distributions (ℓ,ζ) and (ζ,ℓ') are related by Lemma I.5 as follows: for any integrable function h of ℓ and ζ, we have

$$\mathbb{E}_{\ell,\zeta}[h(\ell,\zeta)] = c\,\mathbb{E}_{\zeta,\ell'}[h(\ell',\zeta)\det{}^\perp(T_\zeta\ell', T_\zeta V(f))], \tag{4.7}$$

where c is some normalization constant and where $\det^\perp(T_\zeta\ell', T_\zeta V(f))$ is defined in I.§.2.1. It is only a matter of unfolding definitions to see that it is equal to $\cos\theta'$, where θ' denotes the angle between $T_\zeta\ell'$ and $T_\zeta V(f)^\perp$. With $h = 1$, we obtain $c = \mathbb{E}\left[\cos\theta'\right]^{-1}$ and therefore we get

$$\mathbb{E}_{\ell,\zeta}[h(\ell,\zeta)] = \mathbb{E}_{\zeta,\ell'}[h(\ell',\zeta)\cos\theta']\,\mathbb{E}[\cos\theta']^{-1}. \tag{4.8}$$

We analyze now the distribution of θ': $\cos(\theta')^2$ is a beta-distributed variable with parameters 1 and $n-1$: indeed, $\cos(\theta')^2 = |u_1|^2/\|u\|^2$ where $u \in \mathbb{C}^n$ is a Gaussian random vector, and it is well known that the distribution of this quotient

of χ^2-distributed random variables is a beta-distributed variable. Generally, the moments of a beta-distributed random variable Z with parameters α, β satisfy

$$\mathbb{E}[Z^r] = \frac{B(\alpha + r, \beta)}{B(\alpha, \beta)}, \tag{4.9}$$

where B is the Beta function and $r > -\alpha$. In particular, for $r > -1$,

$$\mathbb{E}_{\zeta,\ell'}[\cos(\theta')^{2r}] = \frac{B(1 + r, n - 1)}{B(1, n - 1)}, \tag{4.10}$$

and hence

$$\mathbb{E}\left[\cos(\theta')^{-1}\right] \mathbb{E}\left[\cos(\theta')\right]^{-1} = \frac{B(\frac{1}{2}, n - 1)}{B(\frac{3}{2}, n - 1)} = 2n - 1. \tag{4.11}$$

Continuing with (4.8), we obtain

$$\begin{aligned}
\mathbb{E}_\ell\left[\Gamma(f|_\ell)^2\right] &= \mathbb{E}_{\ell,\zeta}\left[\gamma_{\mathrm{Frob}}(f|_\ell, \zeta)^2\right], && \text{by (1.5)}, && (4.12)\\
&\leqslant \mathbb{E}_{\ell,\zeta}\left[\gamma_{\mathrm{Frob}}(f, \zeta)^2 \cos(\theta)^{-2}\right], && \text{by (4.6)}, && (4.13)\\
&= \mathbb{E}_{\zeta,\ell'}\left[\gamma_{\mathrm{Frob}}(f, \zeta)^2 \cos(\theta')^{-1}\right] \mathbb{E}\left[\cos(\theta')\right]^{-1}, && \text{by (4.8)}, && (4.14)\\
&= \mathbb{E}_\zeta\left[\gamma_{\mathrm{Frob}}(f, \zeta)^2\right] \mathbb{E}_{\ell'}\left[\cos(\theta')^{-1}\right] \mathbb{E}\left[\cos(\theta')\right]^{-1} && && (4.15)\\
&= \Gamma(f)^2(2n - 1), && \text{by (4.11)} && (4.16)
\end{aligned}$$

the second last equality (4.15) since the random variable θ' is independent from ζ. This concludes the proof. □

4.2. **Proof of Theorem 4.1.** Termination and correctness of Algorithm BlackBoxSolve are clear by Proposition 3.10. We now study the average complexity of BlackBoxSolve$(\mathbf{u} \cdot F, \varepsilon)$, where $\mathbf{u} \in \mathcal{U}$ is uniformly distributed. Recall that $\Gamma(\mathbf{u} \cdot F) = \Gamma(F)$, by unitary invariance of γ_{Frob}, and $L(\mathbf{u} \cdot F) = L(F) + O(n^3)$.

The sampling operation costs at most $\mathrm{poly}(n, \delta) \cdot L(F) \cdot \log\log\Gamma(F)$ on average, by Proposition 4.2. The expected cost of the continuation phase is $\mathrm{poly}(n, \delta) \cdot L(F) \cdot K(\log K + \log \varepsilon^{-1})$, by Proposition 3.10, where $K = K(\mathbf{u} \cdot F, \mathbf{1}_\mathcal{U}, \mathbf{v}, z)$ and $(\mathbf{v}, z)$ is the sampled initial pair. By unitary invariance,

$$K(\mathbf{u} \cdot F, \mathbf{1}_\mathcal{U}, \mathbf{v}, z) = K(F, \mathbf{u}, \mathbf{v}', z), \tag{4.17}$$

where $\mathbf{v}' = \mathbf{v}\mathbf{u}$. Moreover, since $\mathbf{v}$ is uniformly distributed and independent from $\mathbf{u}$, $\mathbf{v}'$ is also uniformly distributed and independent from $\mathbf{u}$, and z is a uniformly distributed zero of $\mathbf{v}' \cdot F$. So the following proposition concludes the proof of Proposition 4.1.

Proposition 4.6. *Let $\mathbf{u}, \mathbf{v} \in \mathcal{U}$ be independent and uniformly distributed random variables, let ζ be a uniformly distributed zero of $\mathbf{v} \cdot F$ and let $K = K(F, \mathbf{u}, \mathbf{v}, \zeta)$. Then we have $\mathbb{E}[K] \leqslant \mathrm{poly}(n, \delta)\Gamma(F)$ and $\mathbb{E}[K \log K] \leqslant \mathrm{poly}(n, \delta) \cdot \Gamma(F) \log\Gamma(F)$.*

Sketch of proof. The first bound $\mathbb{E}[K] \leqslant \mathrm{poly}(n, \delta)\Gamma(F)$ was shown in Theorem I.27. Following *mutatis mutandis* the proof of Theorem I.25 (the only additional fact needed is Proposition 4.7 below for $a = 3/2$), we obtain that

$$\mathbb{E}\left[K^{\frac{3}{2}}\right] \leqslant \mathrm{poly}(n, \delta)\Gamma(F)^{\frac{3}{2}}. \tag{4.18}$$

Next, we observe that the function $h : x \mapsto x^{\frac{2}{3}}(1 + \log x^{\frac{2}{3}})$ is concave on $[1, \infty)$. By Jensen's inequalities, it follows that

$$\mathbb{E}[K \log K] \leqslant \mathbb{E}\left[h(K^{\frac{3}{2}})\right] \leqslant h\left(\mathbb{E}[K^{\frac{3}{2}}]\right) \leqslant \mathrm{poly}(n, \delta)\Gamma(F) \log\Gamma(F), \tag{4.19}$$

which gives the claim. □

The following statement extends Proposition I.17 to more general exponents. The proof technique is more elementary and the result, although not as tight, good enough for our purpose.

Proposition 4.7. *Let $M \in \mathbb{C}^{n\times(n+1)}$ be a random matrix whose rows are independent uniformly distributed vectors in $\mathbb{S}(\mathbb{C}^{n+1})$, and let $\sigma_{\min}(M)$ be the smallest singular value of M. For all $a \in [1,2)$,*

$$\mathbb{E}\left[\sigma_{\min}(M)^{-2a}\right] \leqslant \frac{n^{1+2a}}{2-a},$$

and, equivalently with the notations of Proposition I.17,

$$\mathbb{E}\left[\kappa(\mathbf{u},\zeta)^{2a}\right] \leqslant \frac{n^{1+2a}}{2-a}.$$

Proof. For short, let σ denote $\sigma_{\min}(M)$. Let $u_1,\dots,u_n$ be the rows of M. By definition, there is a unit vector $x \in \mathbb{C}^n$ such that

$$\|x_1u_1 + \cdots + x_nu_n\|^2 = \sigma^2. \tag{4.20}$$

If V_i denotes the subspace of $\mathbb{C}^{n+1}$ spanned by all u_j except u_i, and b_i denotes the squared Euclidean distance of u_i to V_i, then (4.20) implies $b_i \leqslant |x_i|^{-2}\sigma^2$ for all i. Moreover, since x is a unit vector, there is at least one i such that $|x_i|^2 \geqslant \frac{1}{n}$. Hence $n\sigma^2 \geqslant \min_i b_i$ and therefore

$$\mathbb{E}\left[\sigma^{-2a}\right] \leqslant n^a\mathbb{E}\left[\max_i b_i^{-a}\right] \leqslant n^a\sum_{i=1}^{n}\mathbb{E}\left[b_i^{-a}\right]. \tag{4.21}$$

To analyze the distribution of b_i consider, for fixed V_i, a standard Gaussian vector p_i in V_i, and an independent standard Gaussian vector q_i in $V_i^\perp$. (Note $\dim V_i^\perp = 4$.) Since u_i is uniformly distributed in the sphere, it has the same distribution as $(p_i+q_i)/\sqrt{\|p_i\|^2+\|q_i\|^2}$. In particular, b_i has the same distribution as $\|q_i\|^2/(\|p_i\|^2+\|q_i\|^2)$, which is a Beta distribution with parameters $2, n-1$, since $\|p_i\|^2$ and $\|q_i\|^2$ are independent χ^2-distributed random variables with $2n-2$ and 4 degrees of freedom, respectively. By (4.9) we have for the moments, using $a<2$,

$$\mathbb{E}\left[b_i^{-a}\right] = \frac{B(2-a,n-1)}{B(2,n-1)} = \frac{\Gamma(2-a)\Gamma(n+1)}{\Gamma(n+1-a)}. \tag{4.22}$$

We obtain

$$\mathbb{E}\left[b_i^{-a}\right] = \frac{\Gamma(3-a)}{2-a}\cdot n\cdot\frac{\Gamma(n)}{\Gamma(n+1-a)}, \qquad \text{using twice } \Gamma(x+1)=x\Gamma(x), \tag{4.23}$$

$$\leqslant \frac{1}{2-a}\cdot n\cdot n^{a-1}, \qquad \text{by Gautschi's inequality.} \tag{4.24}$$

In combination with (4.21) this gives the result. □

4.3. **Confidence boosting and proof of Theorem 1.1.** We may leverage the quadratic convergence of Newton's iteration to increase the confidence in the result of Algorithm 5 and reduce the dependence on ε (the maximum probability of failure) from $\log\frac{1}{\varepsilon}$ down to $\log\log\frac{1}{\varepsilon}$, so that we can choose $\varepsilon = 10^{-10^{100}}$ without afterthoughts, at least in the BSS model. On a physical computer, the working precision should be comparable with ε, which imposes some limitations. A complete certification, without possibility of error, with $\mathrm{poly}(n,\delta)$ evaluations of F, seems difficult to reach in the black-box model: with only $\mathrm{poly}(n,\delta)$ evaluations, we cannot

Algorithm 6. Boosting the confidence for approximate zeros
Input: $F = (f_1, \dots, f_n) \in \mathcal{H}$ (given as black-box), $w \in \mathbb{P}^n$ and $\varepsilon \in (0, \frac{1}{2})$.
Output: $z \in \mathbb{P}^n$ or Fail
Postcondition: If Boost returns a point z, then it is an approximate zero of F with probability $\geqslant 1 - \varepsilon$.

function Boost(F, w, ε)
  $k \leftarrow \left\lceil \max\left(1 + \log_2 \log_2 \left(20n^2\delta\alpha_0^{-1}\right), 1 + \log_2 \log_2 \varepsilon^{-1}\right)\right\rceil$
  $z \leftarrow \mathcal{N}_F^k(w)$
  $c \leftarrow \kappa(F, z)\left(\sum_{i=1}^n \textsc{GammaProb}(f_i, z, \frac{1}{4n})^2\right)^{\frac{1}{2}}$ ▷ Algorithm 2
  if $2^{2^{k-1}}\beta(F, z)c \leqslant \alpha_0$ **then**
    return z
  else
    return Fail
  end if
end function

distinguish a polynomial system F from the infinitely many other systems with the same evaluations.

To describe this boosting procedure we first recall some details about α-theory and Part I. Let $F \in \mathcal{H}$ be a polynomial system and $z \in \mathbb{P}^n$ be a projective point. Let $\mathcal{N}_F(z)$ denote the projective Newton iteration (and $\mathcal{N}_F^k(z)$ denote the composition of k projective Newton iterations). Let

$$\beta(F, z) \doteq d_{\mathbb{P}}\left(z, \mathcal{N}_F(z)\right). \tag{4.25}$$

There is an absolute constant α_0 such that for any $z \in \mathbb{P}^n$, if $\beta(F, z)\gamma(F, z) \leqslant \alpha_0$, then z is an approximate zero of F (Dedieu and Shub 1999, Theorem 1). This is one of many variants of the alpha-theorem of Smale (1986). There may be differences in the definition of γ or β, or even the precise definition of approximate zero, but they only change the constant α_0.

It is important to be slightly more precise about the output of Algorithm 5 (when all estimates are correct, naturally): by the design of the numerical continuation (see Proposition I.22 with $C = 15$ and $A = \frac{1}{4C}$), the output point $w \in \mathbb{P}^n$ satisfies

$$d_{\mathbb{P}}(w, \zeta)\hat{\gamma}_{\mathrm{Frob}}(F, \zeta) \leqslant \frac{1}{4 \cdot 15} = \frac{1}{60}, \tag{4.26}$$

for some zero ζ of F, where $\hat{\gamma}_{\mathrm{Frob}}$ is the split Frobenius γ number (see §2.3). This implies (see Theorem I.12), using $\gamma \leqslant \hat{\gamma}_{\mathrm{Frob}}$, that

$$d_{\mathbb{P}}\left(\mathcal{N}_F^k(w), \zeta\right) \leqslant 2^{1-2^k} d_{\mathbb{P}}(w, \zeta). \tag{4.27}$$

The last important property we recall is the 15-Lipschitz continuity of the function $z \in \mathbb{P}^n \mapsto \hat{\gamma}_{\mathrm{Frob}}(F, z)^{-1}$ (Lemmas I.26 and I.31).

Algorithm 6 checks the criterion $\beta(F, z)\gamma(F, z) \leqslant \alpha_0$ after having refined the presumed approximate zero with a few Newton's iterations. If the input point is indeed an approximate zero, then $\beta(F, z)$ will be very small and it will satisfy the criterion above even with a very gross approximation of $\gamma(F, z)$.

Proposition 4.8. *On input $F \in \mathcal{H}$, $w \in \mathbb{P}^n$, and $\varepsilon \in (0, \frac{1}{2})$, Algorithm* Boost *outputs some $z \in \mathbb{P}^n$ (succeeds) or fails after $\mathrm{poly}(n, \delta)L(F)\log\log\varepsilon^{-1}$ operations. If w satisfies* (4.26)*, then it succeeds with probability at least $\frac{3}{4}$. If it succeeds, then the output point is an approximate zero of F with probability at least $1 - \varepsilon$.*

Proof. We use the notations (k, z, and c) of Algorithm 6. Assume first that (4.26) holds for w and some zero ζ of F. By (4.27) and (4.26),

$$d_{\mathbb{P}}(z,\zeta)\hat{\gamma}_{\mathrm{Frob}}(F,\zeta) \leqslant \frac{1}{60}. \tag{4.28}$$

Using the Lipschitz continuity and (4.28),

$$\hat{\gamma}_{\mathrm{Frob}}(F,z) \leqslant \frac{\hat{\gamma}_{\mathrm{Frob}}(F,\zeta)}{1-15d_{\mathbb{P}}(z,\zeta)\hat{\gamma}_{\mathrm{Frob}}(F,\zeta)} \leqslant \frac{4}{3}\hat{\gamma}_{\mathrm{Frob}}(F,\zeta), \tag{4.29}$$

and it follows from (4.27) and (4.26) again that

$$\beta(F,z)\hat{\gamma}_{\mathrm{Frob}}(F,z) \leqslant \frac{4}{3}\left(d_{\mathbb{P}}(z,\zeta)+d_{\mathbb{P}}(\mathcal{N}_F(z),\zeta)\right)\hat{\gamma}_{\mathrm{Frob}}(F,\zeta) \tag{4.30}$$

$$\leqslant \frac{4}{3}\left(2^{1-2^k}+2^{1-2^{k+1}}\right)d_{\mathbb{P}}(w,\zeta)\hat{\gamma}_{\mathrm{Frob}}(F,\zeta) \tag{4.31}$$

$$\leqslant \frac{1}{10}2^{-2^k}. \tag{4.32}$$

Besides, by Theorem 3.3, we have with probability at least $\frac{3}{4}$,

$$c \leqslant 192n^2\delta\cdot\hat{\gamma}_{\mathrm{Frob}}(F,z). \tag{4.33}$$

(Note that the computation of c involves n calls to GammaProb, each returning a result outside the specified range with probability at most $\frac{1}{4n}$. So the n computations are correct with probability at least $\frac{3}{4}$.) It follows from (4.32) and (4.33), along with the choice of k, that

$$2^{2^{k-1}}\beta(F,z)c \leqslant \tfrac{192}{10}n^2\delta 2^{-2^{k-1}} \leqslant \alpha_0, \tag{4.34}$$

with probability at least $\frac{3}{4}$. We conclude, assuming (4.26), that Algorithm BlackBoxSolve succeeds with probability at least $\frac{3}{4}$.

Assume now that the algorithm succeeds but z, the output point, is not an approximate zero of F. On the one hand, z is not an approximate zero, so

$$\beta(F,z)\hat{\gamma}_{\mathrm{Frob}}(F,z) > \alpha_0, \tag{4.35}$$

and on the other hand, the algorithm succeeds, so $2^{2^{k-1}}\beta(F,z)c \leqslant \alpha_0$, and then

$$2^{2^{k-1}}c \leqslant \hat{\gamma}_{\mathrm{Frob}}(F,z). \tag{4.36}$$

By definition (2.17) of $\hat{\gamma}_{\mathrm{Frob}}(F,z)$, and since $c=\kappa(F,z)(\Gamma_1^2+\cdots+\Gamma_n^2)^{\frac{1}{2}}$, where Γ_i denotes the value returned by the call to GammaProb$(f_i,z,\frac{1}{4n})$, we get

$$2^{2^{k-1}}(\Gamma_1^2+\cdots+\Gamma_n^2)^{\frac{1}{2}} \leqslant \left(\gamma_{\mathrm{Frob}}(f_1,z)^2+\cdots+\gamma_{\mathrm{Frob}}(f_n,z)^2\right)^{\frac{1}{2}}. \tag{4.37}$$

This implies that, for some i,

$$2^{2^{k-1}}\Gamma_i \leqslant \hat{\gamma}_{\mathrm{Frob}}(f_i,z). \tag{4.38}$$

By choice of k, $2^{2^{k-1}} \geqslant \varepsilon^{-1}$, and using the tail bound in Theorem 3.3, with $t=\varepsilon^{-1}$, (4.38) may only happen with probability at most

$$\frac{1}{4n}\Big(\frac{1}{4n}\Big)^{\frac{1}{2}\log_2 t} \leqslant \left(\frac{1}{4}\right)^{\frac{1}{2}\log_2 t} = 2^{-\log_2 t} = \varepsilon. \tag{4.39}$$

The complexity bound is clear since a Newton iteration requires only $\mathrm{poly}(n,\delta)L(F)$ operations. □

The combination of BlackBoxSolve and Boost leads to Algorithm 7, Boost-BlackBoxSolve.

Algorithm 7. Boosted zero finder for black-box input
Input: $F \in \mathcal{H}$ (given as black-box), $\varepsilon \in (0, \frac{1}{2})$
Output: $z \in \mathbb{P}^n$
Postcondition: z is an approximate zero of F with probability $\geqslant 1 - \varepsilon$.

```
function BoostBlackBoxSolve(F, ε)
    repeat
        w ← BlackBoxSolve(F, 1/4)        ▷ Algorithm 5
        z ← Boost(F, w, ε)               ▷ Algorithm 6
    until z ≠ Fail
    return z
end function
```

Proof of Theorem 1.1. The correctness, with probability at least $1 - \varepsilon$, is clear, by the correctness of Boost. An iteration of Algorithm 7 succeeds if and only if Boost succeeds. If (4.26) holds (which it does with probability at least $\frac{3}{4}$), then Boost succeeds with probability at least $\frac{3}{4}$. So each iteration of Algorithm 7 succeeds with probability at least $\frac{1}{2}$, and the expected number of iterations is therefore at most two. Furthermore, on input $\mathbf{u} \cdot F$, the average cost of each iteration is $\operatorname{poly}(n, \delta) L(F) \Gamma(F) \log \Gamma(F)$ for BlackBoxSolve and $\operatorname{poly}(n, \delta) L(F) \log \log \varepsilon^{-1}$ for Boost. □

Proof of Corollary 1.2. Let $\mathbf{u} \in \mathcal{U}$ be uniformly distributed and independent from F. By hypothesis, $\mathbf{u} \cdot F$ and F have the same distribution, so we study $\mathbf{u} \cdot F$ instead. Then Theorem 1.1 applies and we obtain, for fixed $f \in \mathcal{H}$ and random $u \in \mathcal{U}$, that BoostBlackBoxSolve terminates after

$$\operatorname{poly}(n, \delta) \cdot L \cdot \left(\mathbb{E}\left[\Gamma(F) \log \Gamma(F) \right] + \log \log \varepsilon^{-1} \right) \tag{4.40}$$

operations on average. With the concavity on $[1, \infty)$ of the function $h : x \mapsto x^{\frac{1}{2}} \log x^{\frac{1}{2}}$, Jensen's inequality ensures that

$$\mathbb{E}\left[\Gamma(F) \log \Gamma(F) \right] = \mathbb{E}\left[h(\Gamma(F)^2) \right] \leqslant h\left(\mathbb{E}[\Gamma(F)^2] \right), \tag{4.41}$$

which gives the complexity bound. □

5. Probabilistic analysis of algebraic branching programs

The goal of this section is to prove our second main result, Theorem 1.4. Recall from §1.7 the notion of a Gaussian random ABP. We first state a result that connects the notions of irreducible Gaussian random ABPs with that of irreducible polynomials.

Lemma 5.1. *Let f be the homogeneous polynomial computed by an irreducible Gaussian random ABP in the variables $z_0, \ldots, z_n$. If $n \geqslant 2$ then f is almost surely irreducible.*

Proof. The proof is by induction on the degree δ, the base case $\delta = 1$ being clear. So suppose $\delta \geqslant 2$. In the given ABP replace the label of each edge e by a new variable y_e. Let G denote the modified ABP and g the polynomial computed by G. The polynomial f is obtained as a restriction of g to a generic linear subspace, so, by Bertini's theorem, it suffices to prove that g is irreducible (recall $n \geqslant 2$).

Let s denote the source vertex and t the target vertex of G. There is a path from s to t: let $e = (s, v)$ be its first edge. We remove s and all vertices in the first layer different from v, making v the source vertex of a new ABP denoted H. It is

irreducible: if the layers of G have the sizes $1, r_1, \ldots, r_{\delta-1}, 1$, then the layers of H have the sizes $1, r_2, \ldots, r_{\delta-1}, 1$. The paths of H from source to target are in bijective correspondence with the paths of G from v to t. Therefore, $g = y_e p + q$, where p is the polynomial computed by H, and q corresponds to the paths from s to t which avoid v. By induction hypothesis, p is irreducible. Clearly, $q \neq 0$ because $r_1 > 0$, and p does not divide q since the variable corresponding to an edge leaving v does not appear in q (such edge exists due to $\delta \geqslant 2$). We conclude that p and q are relatively prime. Moreover, the variable y_e does neither appear in p nor in q, so it follows that g is irreducible. □

We also remark that a random polynomial computed by a Gaussian random ABP may define a random hypersurface in $\mathbb{P}^n$ that is always singular. It is rather uncommon in our field to be able to study stochastic models featuring singularities almost surely, so it is worth a lemma.

Lemma 5.2. *If $f \in \mathbb{C}[z_0, \ldots, z_n]$ is the polynomial computed by a algebraic branching program with at most n edges, then the hypersurface $V(f) \subset \mathbb{P}^n$ is singular.*

Proof. Let e be the number of edges of the algebraic branching program computing f. After a linear change of variables, we may assume that f depends only on $z_0, \ldots, z_{e-1}$. The singular locus of $V(f)$ is defined by the vanishing of the partial derivatives $\frac{\partial}{\partial z_i} f$. But these derivatives are identically 0 for $i \geqslant e$, so that the singular locus is defined by at most e equations. So it is nonempty. □

As already mentioned before, the distribution of a polynomial computed by a Gaussian random ABP is best understood in terms of matrices. This calls for the introduction of some terminology. For any δ-tuple $\mathbf{r} = (r_1, \ldots, r_\delta)$, let $M_{\mathbf{r}}(n+1)$ (and $M_{\mathbf{r}}$ for short) denote the space of all δ-tuples of matrices $(A_1(z), \ldots, A_\delta(z))$, of respective size $r_\delta \times r_1$, $r_1 \times r_2$, $\ldots$, $r_{\delta-1} \times r_\delta$, with degree one homogeneous entries in $z = (z_0, \ldots, z_n)$. (It is convenient to think of $r_0 = r_\delta$.) We have $\dim_{\mathbb{C}} M_{\mathbf{r}} = (n+1) \sum_{i=1}^{\delta} r_{i-1} r_i$. For $A \in M_{\mathbf{r}}$, we define the degree δ homogeneous polynomial

$$f_A(z) \doteq \operatorname{tr}\left(A_1(z) \cdots A_\delta(z)\right). \tag{5.1}$$

A Hermitian norm is defined on $M_{\mathbf{r}}$ by

$$\|A\|^2 \doteq \sum_{i=1}^{\delta} \sum_{j=0}^{n} \|A_i(e_j)\|^2_{\mathrm{Frob}},$$

where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{C}^{n+1}$, with a 1 at index j ($0 \leqslant j \leqslant n$). The standard Gaussian probability on $M_{\mathbf{r}}$ is defined by the density $\pi^{-\dim_{\mathbb{C}} M_{\mathbf{r}}} \exp(-\|A\|^2) \mathrm{d}A$. The distribution of the polynomial computed by a Gaussian random ABP with layer sizes $(r_1, \ldots, r_{\delta-1})$ is the distribution of f_A, where A is standard Gaussian in $M_{(r_1, \ldots, r_{\delta-1}, 1)}$.

The following statement is the main ingredient of the proof of Theorem 1.4. It can be seen as an analogue of Lemma I.37. (Note that $r_\delta = 1$, the case of interest of ABPs, is included.)

Proposition 5.3. *Assume that $r_1, \ldots, r_{\delta-1} \geqslant 2$. Let $A \in M_{\mathbf{r}}$ be standard Gaussian and let $\zeta \in \mathbb{P}^n$ be a uniformly distributed projective zero of f_A. For any $k \geqslant 2$, we*

have

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

Theorem 1.4 easily follows from Proposition 5.3.

Proof of Theorem 1.4. Let $A \in M_{\mathbf{r}}$ be standard Gaussian so that $f = f_A$. The proof follows exactly the lines of the proof of Lemma I.38 and the intermediate Lemma I.37. We bound the supremum in the definition (1.3) of γ_{Frob} by a sum:

$$\mathbb{E}\left[\gamma_{\mathrm{Frob}}(f_A,\zeta)^2\right] \leqslant \sum_{k=2}^{\delta}\mathbb{E}\left[\left(\|\mathrm{d}_\zeta f_A\|^{-1}\left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|_{\mathrm{Frob}}\right)^{\frac{2}{k-1}}\right] \tag{5.2}$$

$$\leqslant \sum_{k=2}^{\delta}\mathbb{E}\left[\|\mathrm{d}_\zeta f_A\|^{-2}\left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}\right]^{\frac{1}{k-1}} \tag{5.3}$$

$$\leqslant \sum_{k=2}^{\delta}\tfrac{1}{4}\delta^2(\delta+n)\left(1+\frac{\delta-1}{k-1}\right), \text{ by Proposition 5.3,} \tag{5.4}$$

$$\leqslant \tfrac{3}{4}\delta^3(\delta+n)\log\delta, \tag{5.5}$$

using Jensen's inequality for (5.3) and $1+\sum_{k=2}^{\delta}\frac{1}{k-1} \leqslant 2+\log(\delta-1) \leqslant 3\log\delta$ for (5.5). □

The remaining of this article is devoted to the proof of Proposition 5.3.

5.1. **A coarea formula.** The goal of this subsection is to establish a consequence of the coarea formula (Federer 1959, Theorem 3.1) that is especially useful to estimate $\Gamma(f)$ for a random polynomial f. This involves a certain identity of normal Jacobians of projections that appears so frequently that it is worthwhile to provide the statement in some generality.

Let us first introduce some useful notations. For a linear map $h : E \to F$ between two Euclidean spaces we define its *Euclidean determinant* as

$$\mathrm{Edet}(h) \doteq \det(h\circ h^{\mathsf{t}})^{\frac{1}{2}}, \tag{5.6}$$

where $h^{\mathsf{t}} : F \to E$ is the transpose of h. If $p : U \to V$ is a linear map between Hermitian spaces, then $\mathrm{Edet}(p)$ is defined by the induced Euclidean structures on U and V and it is well known that

$$\mathrm{Edet}(p) = \det(p\circ p^*), \tag{5.7}$$

where $p^* : V \to U$ is the Hermitian transpose (and det is the determinant over $\mathbb{C}$).

The *normal Jacobian* of a smooth map φ between Riemannian manifolds at a given point x is defined as the Euclidean determinant of the derivative of the map at that point:

$$\mathrm{NJ}_x\,\varphi \doteq \mathrm{Edet}\left(\mathrm{d}_x\varphi\right). \tag{5.8}$$

Lemma 5.4. *Let E and F be Euclidean (resp. Hermitian) spaces, let V be a subspace of $E\times F$ and let $p : E\times F \to E$ and $q : E\times F \to F$ be the canonical projections. Then* $\mathrm{Edet}(p|_V) = \mathrm{Edet}(q|_{V^\perp})$ *and* $\mathrm{Edet}(q|_V) = \mathrm{Edet}(p|_{V^\perp})$.

Proof. By symmetry, it suffices to show the first equality. Let $v_1,\dots,v_r,w_1,\dots,w_s$ be an orthonormal basis of $E\times F$ such that $v_1,\dots,v_r$ is a basis of V and $w_1,\dots,w_s$

is basis of $V^\perp$. After fixing orthonormal bases for E and F (and the corresponding basis of $E \times F$), consider the orthogonal (resp. unitary) matrix U with the columns $v_1, \dots, v_r, w_1, \dots, w_s$. We decompose U as a block matrix

$$U \doteq \left[\begin{array}{c|c} V_E & W_E \\ \hline V_F & W_F \end{array}\right] \doteq \left[\begin{array}{c|c|c|c|c|c} p(v_1) & \dots & p(v_r) & p(w_1) & \dots & p(w_s) \\ \hline q(v_1) & \dots & q(v_r) & q(w_1) & \dots & q(w_s) \end{array}\right]. \tag{5.9}$$

Using $UU^* = I$ and $U^*U = I$ we see that $V_E V_E^* + W_E W_E^* = I$ and $W_E^* W_E + W_F^* W_F = I$. It follows from Sylvester's determinant identity $\det(I + AB) = \det(I + BA)$ that

$$\det(V_E V_E^*) = \det(I - W_E W_E^*) = \det(I - W_E^* W_E) = \det(W_F^* W_F). \tag{5.10}$$

By definition, we have $\operatorname{Edet}(p|_V)^\eta = \det(V_E V_E^*)$ with $\eta = 1$ in the Euclidean situation and $\eta = 2$ in the Hermitian situation. Similarly, $\operatorname{Edet}(q|_{V^\perp})^\eta = \det(W_F W_F^*)$. Therefore, indeed $\operatorname{Edet}(p|_V) = \operatorname{Edet}(q|_{V^\perp})$. □

Corollary 5.5. *In the setting of Lemma 5.4, suppose V is a real (or complex) hyperplane in $E \times F$ with nonzero normal vector $(v, w) \in E \times F$. Then*

$$\frac{\operatorname{Edet}(p|_V)}{\operatorname{Edet}(q|_V)} = \left(\frac{\|w\|}{\|v\|}\right)^\eta,$$

where $\eta = 1$ in the Euclidean situation and $\eta = 2$ in the Hermitian situation.

Proof. $V^\perp$ is spanned by (v, w) and therefore, $\operatorname{Edet}(p|_{V^\perp}) = \left(\|v\|/\sqrt{\|v\|^2 + \|w\|^2}\right)^\eta$, and $\operatorname{Edet}(q|_{V^\perp}) = \left(\|w\|/\sqrt{\|v\|^2 + \|w\|^2}\right)^\eta$. Now apply Lemma 5.4. □

We consider now the abstract setting of a family (f_A) of homogeneous polynomials of degree δ in the variables $z_0, \dots, z_n$, parameterized by elements A of a Hermitian manifold M through a holomorphic map $A \in M \mapsto f_A$. Let $\mathcal{V}$ be the solution variety $\{(A, \zeta) \in M \times \mathbb{P}^n \mid f_A(\zeta) = 0\}$ and $\pi_1 : \mathcal{V} \to M$ and $\pi_2 : \mathcal{V} \to \mathbb{P}^n$ be the restrictions of the canonical projections. We can identify the fiber $\pi_1^{-1}(A)$ with the zero set $V(f_A)$ in $\mathbb{P}^n$. Moreover, the fiber $\pi_2^{-1}(\zeta)$ can be identified with $M_\zeta \doteq \{A \in M \mid f_A(\zeta) = 0\}$. For fixed $\zeta \in \mathbb{P}^n$, we consider the map $M \to \mathbb{C}$, $A \mapsto f_A(\zeta)$ and its derivative at A,

$$\partial_A f(\zeta) : T_A M \to \mathbb{C}. \tag{5.11}$$

Moreover, for fixed $A \in M$, we consider the map $f_A \colon \mathbb{C}^{n+1} \to \mathbb{C}$ and its derivative at ζ,

$$\mathrm{d}_\zeta f_A : T_\zeta \mathbb{P}^n \to \mathbb{C}, \tag{5.12}$$

restricted to the tangent space $T_\zeta \mathbb{P}^n$, that we identify with the orthogonal complement of $\mathbb{C}\zeta$ in $\mathbb{C}^{n+1}$ with respect to the standard Hermitian inner product.

Proposition 5.6. *For any measurable function $\Theta : \mathcal{V} \to [0, \infty)$, we have*

$$\int_M \mathrm{d}A \int_{V(f_A)} \mathrm{d}\zeta\ \Theta(A, \zeta) \|\partial_A f(\zeta)\|^2 = \int_{\mathbb{P}^n} \mathrm{d}\zeta \int_{M_\zeta} \mathrm{d}A\ \Theta(A, \zeta) \|\mathrm{d}_\zeta f_A\|^2.$$

Here $\mathrm{d}A$ denotes the Riemannian volume measure on M and M_ζ, respectively.

Proof. As in (Bürgisser and Cucker 2013, Lemma 16.9), the tangent space of $\mathcal{V}$ at $(A, \zeta) \in \mathcal{V}$ can be expressed as

$$V \doteq T_{A,\zeta}\mathcal{V} = \left\{(\dot{A}, \dot{\zeta}) \in T_A M \times T_\zeta \mathbb{P}^n \;\middle|\; \mathrm{d}_\zeta f_A(\dot{\zeta}) + \partial_A f(\zeta)(\dot{A}) = 0\right\}. \tag{5.13}$$

If $\partial_A f(\zeta)$ and $\mathrm{d}_\zeta f_A$ are not both zero, then V is a hyperplane in the product $E \times F \doteq T_A M \times T_\zeta \mathbb{P}^n$ of Hermitian spaces and V has the normal vector $(\partial_A f(\zeta), \mathrm{d}_\zeta f_A)$, upon identification of spaces with their duals. If we denote by p and q the canonical projections of V onto E and F, then $\mathrm{d}_{A,\zeta}\pi_1 = p|_V$ and $\mathrm{d}_{A,\zeta}\pi_2 = q|_V$, hence

$$\mathrm{NJ}_{A,\zeta}(\pi_1) = \mathrm{Edet}(p|_V), \quad \mathrm{NJ}_{A,\zeta}(\pi_2) = \mathrm{Edet}(q|_V). \tag{5.14}$$

By Corollary 5.5, we therefore have

$$\frac{\mathrm{NJ}_{A,\zeta}(\pi_1)}{\mathrm{NJ}_{A,\zeta}(\pi_2)} = \frac{\mathrm{Edet}(p|_V)}{\mathrm{Edet}(q|_V)} = \frac{\|\mathrm{d}_\zeta f_A\|^2}{\|\partial_A f(\zeta)\|^2}. \tag{5.15}$$

The coarea formula (Federer 1959, Theorem 3.1) applied to $\pi_1 : \mathcal{V} \to M$ asserts,

$$\int_{\mathcal{V}} \mathrm{d}(A,\zeta)\Theta(A,\zeta)\|\mathrm{d}_\zeta f_A\|^2 \,\mathrm{NJ}_{A,\zeta}(\pi_1) = \int_M \mathrm{d}A \int_{V(f_A)} \mathrm{d}\zeta \;\Theta(A,\zeta)\|\mathrm{d}_\zeta f_A\|^2. \tag{5.16}$$

(Note that $\mathcal{V}$ may have singularities, so we actually apply the coarea formula to its smooth locus.) On the other hand, the coarea formula applied to $\pi_2 : \mathcal{V} \to \mathbb{P}^n$ gives

$$\int_{\mathcal{V}} \mathrm{d}(A,\zeta)\Theta(A,\zeta)\|\partial_A f(\zeta)\|^2 \,\mathrm{NJ}_{A,\zeta}(\pi_2) = \int_{\mathbb{P}^n} \mathrm{d}\zeta \int_{M_\zeta} \mathrm{d}A \;\Theta(A,\zeta)\|\partial_A f(\zeta)\|^2. \tag{5.17}$$

By (5.15) we have

$$\mathrm{NJ}_{A,\zeta}(p)\|\partial_A f(\zeta)\|^2 = \mathrm{NJ}_{A,\zeta}(q)\|\mathrm{d}_\zeta f_A\|^2, \tag{5.18}$$

so all the four integrals above are equal. □

5.2. **A few lemmas on Gaussian random matrices.** We present here some auxiliary results on Gaussian random matrices, centering around the new notion of the *anomaly* of a matrix. This will be crucial for the proof of Theorem 1.4.

We endow the space $\mathbb{C}^r$ with the probability density $\pi^{-r}e^{-\|x\|^2}\mathrm{d}x$, where $\|x\|$ is the usual Hermitian norm, and call a random vector $x \in \mathbb{C}^r$ with this probability distribution *standard Gaussian*. This amounts to say that the real and imaginary parts of x are independent centered Gaussian with variance $\frac{1}{2}$. Note that $\mathbb{E}_x\left[\|x\|^2\right] = r$. This convention slightly differs from some previous writings with a different scaling, where the distribution used is $(2\pi)^{-r}e^{-\frac{1}{2}\|x\|^2}\mathrm{d}x$. This choice seems more natural since it avoids many spurious factors. Similarly, the matrix space $\mathbb{C}^{r\times s}$ is endowed with the probability density $\pi^{-rs}\exp(-\|R\|^2_{\mathrm{Frob}})\mathrm{d}R$, and we call a random matrix with this probability distribution *standard Gaussian* as well. (In the random matrix literature this is called complex Ginibre ensemble.)

Lemma 5.7. *For $P \in \mathbb{C}^{r\times s}$ fixed and $x \in \mathbb{C}^s$ standard Gaussian, we have*

$$\mathbb{E}_x\left[\|Px\|^2\right] = \|P\|^2_{\mathrm{Frob}}, \quad \mathbb{E}_x\left[\|Px\|^{-2}\right] \geqslant \|P\|^{-2}_{\mathrm{Frob}}, \quad \mathbb{E}_x\left[\|x\|^{-2}\right] = \frac{1}{s-1}.$$

Proof. By the singular value decomposition and unitary invariance, we may assume that P equals $\mathrm{diag}(\sigma_1, \dots, \sigma_{\min(r,s)})$, with zero columns or zero rows appended. Then $\|Px\|^2 = \sum_i \sigma_i^2|x_i|^2$, hence $\mathbb{E}_x\left[\|Px\|^2\right] = \sum_i \sigma_i^2 \mathbb{E}_{x_i}\left[\|x_i\|^2\right] = \sum_i \sigma_i^2 = \|P\|^2_{\mathrm{Frob}}$.

For the second assertion, we note that for a nonnegative random variable Z, we have by Jensen's inequality that $\mathbb{E}\left[Z\right]^{-1} \leqslant \mathbb{E}\left[Z^{-1}\right]$, since $x \mapsto x^{-1}$ is convex on $(0,\infty)$. The second assertion follows by applying this to $Z \doteq \|Px\|^2$ and using the first assertion.

For the third assertion, we note $\|x\|^2 = \frac{1}{2}\chi^2_{2s}$, where χ^2_{2s} stands for a chi-square distribution with $2s$ degrees of freedom. It is known that $\mathbb{E}[\chi^{-2}_{2s}] = 1/(2s-2)$. □

We define the *anomaly* of a matrix $P \in \mathbb{C}^{r\times s}$ as the quantity

$$\theta(P) \doteq \mathbb{E}_x\left[\frac{\|P\|^2_{\mathrm{Frob}}}{\|Px\|^2}\right] \in [1,\infty), \tag{5.19}$$

where $x \in \mathbb{C}^s$ is a standard Gaussian random vector. Note that $\theta(P) \geqslant 1$ by Lemma 5.7. Moreover, by the same lemma, $\theta(I_r) = r/(r-1)$. This quantity $\theta(P)$ is easily seen to be finite if $\operatorname{rk} P > 1$; it grows logarithmically to infinity as P approaches a rank 1 matrix.

Lemma 5.8. *Let $P \in \mathbb{C}^{r\times s}$ and $Q \in \mathbb{C}^{t\times u}$ be fixed matrices and $X \in \mathbb{C}^{s\times t}$ be a standard Gaussian random matrix. Then*

$$\mathbb{E}_X\left[\frac{\|P\|^2_{\mathrm{Frob}}\|Q\|^2_{\mathrm{Frob}}}{\|PXQ\|^2_{\mathrm{Frob}}}\right] \leqslant \theta(P).$$

Proof. Up to left and right multiplications of Q by unitary matrices, we may assume that Q is diagonal, with nonnegative real numbers $\sigma_1,\dots,\sigma_{\min(t,u)}$ on the diagonal (and we define $\sigma_i = 0$ for $i > \min(t,u)$). This does not change the left-hand side because the Frobenius norm is invariant by left and right multiplications with unitary matrices, and the distribution of X is unitary invariant as well.

Let $e^u_1,\dots,e^u_u$ (reps. $e^t_1,\dots,e^t_t$) be the canonical basis of $\mathbb{C}^u$ (resp. $\mathbb{C}^t$). Observe that

$$\|PXQ\|^2_{\mathrm{Frob}} = \sum_{i=1}^{u}\|PXQe^u_i\|^2 = \sum_{i=1}^{t}\sigma_i^2\|PXe^t_i\|^2. \tag{5.20}$$

Noting that $\|Q\|^2_{\mathrm{Frob}} = \sigma_1^2+\cdots+\sigma_t^2$, the convexity of $x\mapsto x^{-1}$ on $(0,\infty)$ gives

$$\left(\frac{1}{\|Q\|^2_{\mathrm{Frob}}}\sum_{i=1}^{t}\sigma_i^2\|PXe^t_i\|^2\right)^{-1} \leqslant \frac{1}{\|Q\|^2_{\mathrm{Frob}}}\sum_{i=1}^{t}\frac{\sigma_i^2}{\|PXe^t_i\|^2}. \tag{5.21}$$

Since X is standard Gaussian, $Xe^t_i \in \mathbb{C}^s$ is also standard Gaussian. Therefore, by definition of θ, we have for any $1 \leqslant i \leqslant t$,

$$\mathbb{E}_X\left[\frac{\|P\|^2_{\mathrm{Frob}}}{\|PXe^t_i\|^2}\right] = \theta(P). \tag{5.22}$$

It follows that

$$\begin{aligned}
\mathbb{E}_X\left[\frac{\|P\|^2_{\mathrm{Frob}}\|Q\|^2_{\mathrm{Frob}}}{\|PXQ\|^2_{\mathrm{Frob}}}\right] &\leqslant \frac{1}{\|Q\|^2_{\mathrm{Frob}}}\sum_{i=1}^{t}\sigma_i^2\mathbb{E}\left[\frac{\|P\|^2_{\mathrm{Frob}}}{\|PXe^t_i\|^2}\right], && \text{by (5.20) and (5.21)},\\
&= \frac{1}{\|Q\|^2_{\mathrm{Frob}}}\sum_{i=1}^{t}\sigma_i^2\theta(P), && \text{by (5.22)},\\
&= \theta(P),
\end{aligned}$$

which concludes the proof. □

Lemma 5.9. *Let $P \in \mathbb{C}^{r\times s}$ be fixed, $t > 1$, and $X \in \mathbb{C}^{s\times t}$ be a standard Gaussian random matrix. Then*

$$\mathbb{E}_X\left[\theta(PX)\right] = \frac{1}{t-1} + \theta(P).$$

Furthermore, if $X_1,\dots,X_m$ are standard Gaussian matrices of size $r_0\times r_1$, $r_1\times r_2,\dots,r_{m-1}\times r_m$, respectively, where $r_0,\dots,r_m > 1$, then

$$\mathbb{E}_{X_1,\dots,X_m}\left[\theta(X_1\cdots X_m)\right] = 1 + \sum_{i=0}^{m}\frac{1}{r_i-1}.$$

Proof. Let $x \in \mathbb{C}^t$ be a standard Gaussian random vector, so that

$$\mathbb{E}_X\left[\theta(PX)\right] = \mathbb{E}_{X,x}\left[\frac{\|PX\|_{\mathrm{Frob}}^2}{\|PXx\|^2}\right]. \tag{5.23}$$

We first compute the expectation conditionally on x. So we fix x and write $x = \|x\|u_1$ for some unit vector u_1. We choose other unit vectors $u_2, \dots, u_t$ to form an orthonormal basis of $\mathbb{C}^t$. Since $\|PX\|_{\mathrm{Frob}}^2 = \sum_{i=1}^t \|PXu_i\|^2$, we obtain

$$\frac{\|PX\|_{\mathrm{Frob}}^2}{\|PXx\|^2} = \frac{1}{\|x\|^2} + \sum_{i=2}^{t} \frac{\|PXu_i\|^2}{\|PXu_1\|^2\|x\|^2}. \tag{5.24}$$

Since X is standard Gaussian, the vectors Xu_i are standard Gaussian and independent. So we obtain, using Lemma 5.7,

$$\mathbb{E}_X\left[\frac{\|PXu_i\|^2}{\|PXu_1\|^2}\right] = \mathbb{E}_X[\|PXu_i\|^2]\,\mathbb{E}_X\left[\frac{1}{\|PXu_1\|^2}\right] \tag{5.25}$$

$$= \|P\|_{\mathrm{Frob}}^2\,\mathbb{E}_X\left[\frac{1}{\|PXu_1\|^2}\right] = \theta(P). \tag{5.26}$$

Combining with (5.24), we obtain

$$\mathbb{E}_X\left[\frac{\|PX\|_{\mathrm{Frob}}^2}{\|PXx\|^2}\right] = \frac{1}{\|x\|^2} + \sum_{i=2}^{t} \frac{\theta(P)}{\|x\|^2} = \frac{1}{\|x\|^2}\big(1 + (t-1)\theta(P)\big). \tag{5.27}$$

When we take the expectation over x, the first claim follows with the third statement of Lemma 5.7.

The second claim follows by induction on m. The base case $m = 1$ follows from writing $\mathbb{E}_{X_1}\left[\theta(X_1)\right] = \mathbb{E}_{X_1}\left[\theta(I_{r_0}X_1)\right]$, the first part of Lemma 5.7, and $\theta(I_{r_0}) = 1 + \frac{1}{r_0-1}$. For the induction step $m > 1$, we first fix $X_1, \dots, X_{m-1}$ and obtain from the first assertion

$$\mathbb{E}_{X_m}\left[\theta(X_1 \cdots X_{m-1}X_m)\right] = \frac{1}{r_m - 1} + \theta(X_1 \cdots X_{m-1}). \tag{5.28}$$

Taking the expectation over $X_1, \dots, \cdots, X_{m-1}$ and applying the induction hypothesis implies the claim. □

Lemma 5.10. *For any fixed $P, Q \in \mathbb{C}^{r\times r}$ and $X \in \mathbb{C}^{r\times r}$ standard Gaussian, we have*

(i) $\mathbb{E}\left[\left|\mathrm{tr}(XQ)\right|^2\right] = \|Q\|_{\mathrm{Frob}}^2$,

(ii) $\mathbb{E}\left[\|PXQ\|_{\mathrm{Frob}}^2\right] = \|P\|_{\mathrm{Frob}}^2\,\|Q\|_{\mathrm{Frob}}^2$.

Proof. By unitarily invariance of the distribution of X and the Frobenius norm, we can assume that P and Q are diagonal matrices. Then the claims reduce to easy computations. □

5.3. **Proof of Proposition 5.3.** We now carry out the estimation of

$$\mathbb{E}\left[\|\mathrm{d}_\zeta f_A\|^{-2}\left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f_A\right\|_{\mathrm{Frob}}^2\right], \tag{5.29}$$

where $A \in M_{\mathbf{r}}$ is standard Gaussian and $\zeta \in \mathbb{P}^n$ is a uniformly distributed zero of f_A. The computation is lengthy but the different ingredients arrange elegantly.

5.3.1. *Conditioning A on ζ.* As often in this kind of average analysis, the first step is to consider the conditional distribution of A given ζ, reversing the natural definition where ζ is defined conditionally on A. This is of course the main purpose of Proposition 5.6. Consider the Hermitian vector space $M \doteq M_{\mathbf{r}}$ and let $\mathrm{d}'A =$

$\pi^{-\dim_{\mathbb{C}} M_{\mathbf{r}}} e^{-\sum_i \|A_i\|^2} \mathrm{d}A$ denote the Gaussian probability measure on $M_{\mathbf{r}}$. It is a classical fact (e.g., Howard 1993, p. 20) that the volume of a hypersurface of degree δ in $\mathbb{P}^n$ equals $\delta \operatorname{vol} \mathbb{P}^{n-1}$; this applies in particular to $V(f_A)$. By Proposition 5.6, we have

$$\begin{aligned} &\mathbb{E}\left[\|\mathrm{d}_\zeta f_A\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}\right] \\ &\quad = \int_M \mathrm{d}'A\ (\operatorname{vol} V(f_A))^{-1} \int_{V(f_A)} \mathrm{d}\zeta\ \|\mathrm{d}_\zeta f_A\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}} \qquad (5.30) \\ &\quad = \left(\delta \operatorname{vol} \mathbb{P}^{n-1}\right)^{-1} \int_{\mathbb{P}^n} \mathrm{d}\zeta \int_{M_\zeta} \mathrm{d}'A\ \|\partial_A f(\zeta)\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}. \qquad (5.31) \end{aligned}$$

Here $\mathrm{d}'A$ denotes the Gaussian measure on M and M_ζ, respectively.

We focus on the inner integral over M_ζ for some fixed ζ. Everything being unitarily invariant, this integral actually does not depend on ζ. So we fix $\zeta \doteq [1:0:\cdots:0]$. We next note that $\operatorname{vol} \mathbb{P}^n = \frac{\pi}{n} \operatorname{vol} \mathbb{P}^{n-1}$ and we obtain

$$\mathbb{E}\left[\|\mathrm{d}_\zeta f_A\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}\right] = \frac{\pi}{\delta n} \int_{M_\zeta} \mathrm{d}'A\ \|\partial_A f(\zeta)\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}. \qquad (5.32)$$

Recall that the entries of $A_i = A_i(z)$ are linear forms in $z_0, z_1, \ldots, z_n$. We define

$$B_i \doteq A_i(\zeta) \in \mathbb{C}^{r_{i-1} \times r_i}, \quad A_i(z) = z_0 B_i + C_i(z_1, \ldots, z_n), \qquad (5.33)$$

where the entries of the matrix $C_i(z_1, \ldots, z_n)$ are linear forms $z_1, \ldots, z_n$. This yields an orthogonal decomposition $M_{\mathbf{r}}(n+1) \simeq M_{\mathbf{r}}(1) \oplus M_{\mathbf{r}}(n)$ with respect to the Hermitian norm on $M_{\mathbf{r}}$, where $A = B + C$ with

$$B = (B_1, \ldots, B_\delta) \in \prod_{i=1}^{\delta} \mathbb{C}^{r_{i-1} \times r_i} \simeq M_{\mathbf{r}}(1), \quad C = (C_1, \ldots, C_\delta) \in M_{\mathbf{r}}(n). \qquad (5.34)$$

Consider the function $f(\zeta) : M_{\mathbf{r}}(n+1) \to \mathbb{C}$, $A \mapsto f_A(\zeta)$. By (5.1) we have $f_A(\zeta) = \operatorname{tr}(A_1(\zeta), \ldots, A_\delta(\zeta)) = \operatorname{tr}(B_1 \cdots B_\delta)$. The derivative of $f(\zeta)$ is given by

$$\partial_A f(\zeta)(\dot{A}) = \sum_{i=1}^{\delta} \operatorname{tr}\left(B_1 \cdots B_{i-1} \dot{B}_i B_{i+1} \cdots B_\delta\right) = \sum_{i=1}^{\delta} \operatorname{tr}(\dot{B}_i \hat{B}_i), \qquad (5.35)$$

where $\dot{A} = \dot{B} + \dot{C}$ and (invariance of the trace under cylic permutations)

$$\hat{B}_i \doteq B_{i+1} \cdots B_\delta\, B_1 \cdots B_{i-1} \qquad (5.36)$$

Hence the induced norm of the linear form $\partial_A f(\zeta)$ on the Hermitian space $M_{\mathbf{r}}$ satisfies

$$\|\partial_A f(\zeta)\|^2 = \sum_{i=1}^{\delta} \|\hat{B}_i\|^2_{\mathrm{Frob}}. \qquad (5.37)$$

The equation defining the fiber M_ζ can be written as $\operatorname{tr}(B_1 \cdots B_\delta) = 0$. We have $M_\zeta \simeq W \times M_{\mathbf{r}}(n)$, where W denotes the space of δ-tuples of complex matrices (of respective size $r_0 \times r_1$, $r_1 \times r_2$, etc.) that satisfy this condition. Using this identification, the projection

$$M_\zeta \to W,\ (A_1(z), \ldots, A_\delta(z)) \mapsto (B_1, \ldots, B_\delta) = (A_1(\zeta), \ldots, A_\delta(\zeta)) \qquad (5.38)$$

is given by evaluation at ζ. With (5.37), this implies that

$$\int_{M_\zeta} \mathrm{d}'A\ \|\partial_A f(\zeta)\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}} = \int_W \mathrm{d}'B \int_{M_{\mathbf{r}}(n)} \mathrm{d}'C\ \|\partial_A f(\zeta)\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}} \qquad (5.39)$$

$$= \int_W \frac{\mathrm{d}'B}{\|\hat{B}_1\|^2 + \cdots + \|\hat{B}_\delta\|^2} \int_{M_{\mathbf{r}}(n)} \mathrm{d}'C \left\| \tfrac{1}{k!} \mathrm{d}^k_\zeta f_A \right\|^2_{\mathrm{Frob}}. \tag{5.40}$$

As before, we denote by $\mathrm{d}'B$ and $\mathrm{d}'C$ the Gaussian probability measures on the respective spaces.

5.3.2. *Computation of the inner integral.* We now study $\|\mathrm{d}^k_\zeta f_A\|^2_{\mathrm{Frob}}$ to obtain an expression for the integral $\int \mathrm{d}'C \|\frac{1}{k!}\mathrm{d}^k_\zeta f_A\|^2_{\mathrm{Frob}}$ that appears in (5.40). The goal is Equation (5.56).

Recall that $\zeta = (1, 0, \ldots, 0)$. Let $g(z) \doteq f_A(\zeta + z)$ and write g_k for the kth homogeneous component of g. By Lemma I.30, we have

$$\left\| \tfrac{1}{k!} \mathrm{d}^k_\zeta f_A \right\|_{\mathrm{Frob}} = \|g_k\|_W. \tag{5.41}$$

By expanding a multilinear product, we compute with (5.33) that

$$g(z_0, \ldots, z_n) = \operatorname{tr}\left(((1+z_0)B_1 + C_1) \cdots ((1+z_0)B_\delta + C_\delta)\right) \tag{5.42}$$

$$= \sum_{I \subseteq \{1,\ldots,\delta\}} (1+z_0)^{\delta - \#I} h_I(z_1, \ldots, z_n), \tag{5.43}$$

where $h_I(z_1, \ldots, z_n) \doteq \operatorname{tr}\left(U^I_1 \cdots U^I_\delta\right)$ with

$$U^I_i \doteq \begin{cases} C_i(z_1, \ldots, z_n) & \text{if } i \in I \\ B_i & \text{otherwise.} \end{cases} \tag{5.44}$$

Note that h_I is of degree $\#I$ in $z_1, \ldots, z_n$. Hence the homogeneous part g_k satisfies

$$g_k(z_0, \ldots, z_\delta) = \sum_{m=1}^{k} \binom{\delta - m}{k - m} z_0^{k-m} \sum_{\#I = m} h_I(z_1, \ldots, z_n). \tag{5.45}$$

The contribution for $m = 0$ vanishes by assumption:

$$\binom{\delta}{k} z_0^k h_\varnothing = \binom{\delta}{k} z_0^k \operatorname{tr}(B_1 \cdots B_k) = 0. \tag{5.46}$$

All the terms of the outer sum in (5.45) over m have disjoint monomial support, so they are orthogonal for the Weyl inner product; see §3.1. Moreover for any homogeneous polynomial $p(z_1, \ldots, z_n)$ of degree $m \leqslant k$, the definition of the Weyl norm easily implies $\binom{k}{m}\|z_0^{k-m} p\|^2_W = \|p\|^2_W$. It follows that

$$\|g_k\|^2_W = \sum_{m=1}^{k} \binom{\delta - m}{k - m}^2 \binom{k}{m}^{-1} \left\| \sum_{\#I = m} h_I \right\|^2_W. \tag{5.47}$$

For two different subsets $I, I' \subseteq \{1, \ldots, \delta\}$, there is at least one index i such that C_i occurs in h_I and not in $h_{I'}$, so that the Weyl inner product $\langle h_I, h_{I'} \rangle_W$ depends linearly on C_i and then, by symmetry, $\int \mathrm{d}'C \langle h_I, h_{I'} \rangle_W = 0$. It follows that

$$\int \mathrm{d}'C \ \|g_k\|^2_W = \sum_{m=1}^{k} \binom{\delta - m}{k - m}^2 \binom{k}{m}^{-1} \sum_{\#I = m} \int \mathrm{d}'C \, \|h_I\|^2_W. \tag{5.48}$$

For computing $\int \mathrm{d}'C \, \|h_I\|^2_W$, with $\#I = m > 0$, we proceed as follows. From Lemma 3.1 (h_I is a homogeneous polynomial in n variables of degree m), we obtain that

$$\|h_I\|^2_W = \binom{m + n - 1}{m} \frac{1}{\operatorname{vol} \mathbb{S}(\mathbb{C}^n)} \int_{\mathbb{S}(\mathbb{C}^n)} \mathrm{d}z \, |h_I(z)|^2. \tag{5.49}$$

Then, given that the tuple $(C_1, \dots, C_\delta)$ is standard Gaussian in $M_{\mathbf{r}}(n)$, the matrices $C_1(z), \dots, C_\delta(z)$ are independent standard Gaussian random matrices, for any $z \in \mathbb{S}(\mathbb{C}^r)$. Let $I \subseteq \{1, \dots, \delta\}$ be such that $1 \in I$ (without loss of generality, because the indices are defined up to cyclic permutation). Then we have $h_I(z_1, \dots, z_n) = \operatorname{tr}\left(C_1(z) U_2^I \cdots U_\delta^I\right)$. Integrating over C_1, Lemma 5.10(i) shows for a fixed $z \in \mathbb{S}(\mathbb{C}^{n+1})$ that

$$\int \mathrm{d}'C_1 \ |h_I(z)|^2 = \|U_2^I \cdots U_\delta^I\|_{\mathrm{Frob}}^2. \tag{5.50}$$

Integrating further with respect to C_i with $i \notin I$ is trivial since $\|U_2^I \cdots U_\delta^I\|_{\mathrm{Frob}}^2$ does not depend on these C_i. To integrate with respect to C_i with $i \in I$, we use Lemma 5.10(ii) to obtain

$$\int \mathrm{d}'C_1 \mathrm{d}'C_i \ |h_I(z)|^2 = \|U_2^I \cdots U_{i-1}^I\|_{\mathrm{Frob}}^2 \, \|U_{i+1}^I \cdots U_\delta^I\|_{\mathrm{Frob}}^2. \tag{5.51}$$

After integrating with respect to the remaining C_i in the same way, we obtain

$$\int \mathrm{d}'C \ |h_I(z)|^2 = P_I(B), \tag{5.52}$$

where $P_I(B)$ does not depend on z and is defined as follows. Let $I = \{i_1, \dots, i_m\}$, with $1 = i_1 < \cdots < i_m$. Then

$$P_I(B) \doteq \|B_2 \cdots B_{i_2-1}\|_{\mathrm{Frob}}^2 \|B_{i_2+1} \cdots B_{i_3-1}\|_{\mathrm{Frob}}^2 \cdots \|B_{i_m+1} \cdots B_\delta\|_{\mathrm{Frob}}^2. \tag{5.53}$$

More generally, if $i_1 \neq 1$, $P_I(B)$ is defined as above with the first and last factors replaced, respectively, by

$$\|B_{i_1+1} \cdots B_{i_2-1}\|_{\mathrm{Frob}}^2 \text{ and } \|B_{i_m+1} \cdots B_\delta B_1 \cdots B_{i_1-1}\|_{\mathrm{Frob}}^2, \tag{5.54}$$

and (5.52) still holds. Averaging (5.52) with respect to $z \in \mathbb{S}(\mathbb{C}^n)$, we obtain with (5.49)

$$\int \mathrm{d}'C \ \|h_I\|_W^2 = \binom{m+n-1}{m} P_I(B). \tag{5.55}$$

Combining further with (5.41) and (5.48), we obtain

$$\int \mathrm{d}'C \ \left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f_A\right\|_{\mathrm{Frob}}^2 = \sum_{m=1}^{k} \binom{\delta-m}{k-m}^2 \binom{k}{m}^{-1} \binom{m+n-1}{m} \sum_{\#I=m} P_I(B). \tag{5.56}$$

Combining with (5.40), this leads to

$$\begin{aligned} \int_{M_\zeta} \mathrm{d}'A \ \|\partial_A f(\zeta)\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}_\zeta^k f_A\right\|^2 = \\ \sum_{m=1}^{k} \binom{\delta-m}{k-m}^2 \binom{k}{m}^{-1} \binom{m+n-1}{m} \sum_{\#I=m} \int_W \mathrm{d}'B \frac{P_I(B)}{\|\hat{B}_1\|_{\mathrm{Frob}}^2 + \cdots + \|\hat{B}_\delta\|_{\mathrm{Frob}}^2}. \end{aligned} \tag{5.57}$$

Recall that $\hat{B}_i = B_{i+1} \cdots B_\delta B_1 \cdots B_{i-1}$.

5.3.3. *Computation of the integral over W.* We now consider the integral

$$\int_W \mathrm{d}'B \frac{P_I(B)}{\|\hat{B}_1\|_{\mathrm{Frob}}^2 + \cdots + \|\hat{B}_\delta\|_{\mathrm{Frob}}^2}, \tag{5.58}$$

which appears in the right-hand side of (5.57). The goal is the bound (5.68). To simplify notation, we assume $1 \in I$ but this does not change anything, up to cyclic permutation of the indices. We apply the coarea formula to the projection $q\colon W \to F$, $B \mapsto (B_2, \dots, B_\delta)$, where $F \doteq \mathbb{C}^{r_1 \times r_2} \times \cdots \times \mathbb{C}^{r_{\delta-1} \times r_\delta}$. Since the

complex hypersurface W is defined by the condition $\operatorname{tr}(B_1 \cdots B_\delta) = 0$, we have

$$T_B W = \left\{ (\dot B_1, \dots, \dot B_\delta) \;\middle|\; \sum_{i=1}^{\delta} \operatorname{tr}(\dot B_i \hat B_i) = 0 \right\} \subseteq \mathbb{C}^{r_\delta \times r_1} \times F; \tag{5.59}$$

this is the same computation as for (5.35). In particular, the normal space of W is spanned by $(\hat B_1^*, \dots, \hat B_\delta^*)$, where * denotes the Hermitian transpose. It follows from Lemma 5.4 (used as in Corollary 5.5) that the normal Jacobian $\mathrm{NJ}_B(q)$ of q at some $B \in W$ is given by

$$\mathrm{NJ}_B(q) = \frac{\|\hat B_1\|^2}{\|\hat B_1\|_{\mathrm{Frob}}^2 + \cdots + \|\hat B_\delta\|_{\mathrm{Frob}}^2}. \tag{5.60}$$

The coarea formula then gives

$$\int_W \mathrm{d}'B \frac{P_I(B)}{\|\hat B_1\|_{\mathrm{Frob}}^2 + \cdots + \|\hat B_\delta\|_{\mathrm{Frob}}^2} = \int_F \mathrm{d}'B_2 \cdots \mathrm{d}'B_\delta \int_{\operatorname{tr}(B_1 \cdots B_\delta)=0} \mathrm{d}'B_1 \, \frac{P_I(B)}{\|\hat B_1\|_{\mathrm{Frob}}^2}. \tag{5.61}$$

Note that the inner integrand does not depend on B_1. Moreover, for fixed $B_2, \dots, B_\delta$, the condition $\operatorname{tr}(B_1 \cdots B_\delta) = 0$ restricts B_1 to a hyperplane in $\mathbb{C}^{r_0 \times r_1}$. Due to the unitary invariance of the standard Gaussian measure, the position of the hyperplane does not matter and we obtain

$$\int_{\operatorname{tr}(B_1 \cdots B_\delta)=0} \mathrm{d}'B_1 = \int_{\mathbb{C}^{r_0 r_1 - 1}} \mathrm{d}'B_1 = \frac{1}{\pi}. \tag{5.62}$$

It follows that

$$\int_W \mathrm{d}'B \frac{P_I(B)}{\|\hat B_1\|_{\mathrm{Frob}}^2 + \cdots + \|\hat B_\delta\|_{\mathrm{Frob}}^2} = \frac{1}{\pi} \int_F \mathrm{d}'B_2 \cdots \mathrm{d}'B_\delta \; \frac{P_I(B)}{\|\hat B_1\|_{\mathrm{Frob}}^2} \tag{5.63}$$

$$= \frac{1}{\pi} \int_F \mathrm{d}'B_2 \cdots \mathrm{d}'B_\delta \frac{\prod_{k=1}^{m-1} \|B_{i_k+1} \cdots B_{i_{k+1}-1}\|_{\mathrm{Frob}}^2 \cdot \|B_{i_m+1} \cdots B_\delta\|_{\mathrm{Frob}}^2}{\|B_2 \cdots B_{i_2} \cdots B_{i_3} \cdot \cdots \cdot B_{i_m} \cdots B_\delta\|_{\mathrm{Frob}}^2}, \tag{5.64}$$

where $I = \{i_1, \dots, i_m\}$ with $i_1 = 1$. If $m = 1$, that is $I = \{1\}$, then the integrand simplifies to 1.

Recall the anomaly $\theta(A)$ of a matrix defined in (5.19). When $m > 1$, we take expectations over $B_{i_2}, \dots, B_{i_m}$ and repeatedly apply Lemma 5.8, to obtain[7]

$$\int_F \mathrm{d}'B_2 \cdots \mathrm{d}'B_\delta \; \frac{P_I(B)}{\|\hat B_1\|_{\mathrm{Frob}}^2} \leqslant \prod_{k=1}^{m-1} \int \mathrm{d}'B_{i_k+1} \cdots \mathrm{d}'B_{i_{k+1}-1} \theta(B_{i_k+1} \cdots B_{i_{k+1}-1}). \tag{5.65}$$

Every block $B_{i_k+1} \cdots B_{i_{k+1}-1}$ appears except the last block $B_{i_m+1} \cdots B_\delta$. If one of the parameters r_i is 1, then, by cyclic permutation of the indices, we may assume that it appears in the last block (indeed, by the hypothesis $r_1, \dots, r_{\delta-1} \geqslant 2$, there is at most one i with $r_i = 1$).

---

[7] Let us exemplify the computations (5.65)–(5.66) on a particular case: $\delta = 6$ and $I = \{1, 4, 5\}$. In this case $P_I(B) = \|B_2 B_3\|_{\mathrm{Frob}}^2 \|\mathbf{1}\|_{\mathrm{Frob}}^2 \|B_6\|_{\mathrm{Frob}}^2$, where $\mathbf{1}$ is the identity matrix of size $r_4 \times r_4$. Then, by (5.63) and two applications of Lemma 5.8 (first for integrating w.r.t B_4 then B_5),

$$\begin{aligned}
\int \mathrm{d}'B_2 \cdots \mathrm{d}'B_6 \; \frac{P_I(B)}{\|\hat B_1\|^2} &= \int \mathrm{d}'B_2 \cdots \mathrm{d}'B_6 \frac{\|B_2 B_3\|_{\mathrm{Frob}}^2 \, \|B_5 B_6\|_{\mathrm{Frob}}^2}{\|B_2 B_3 B_4 B_5 B_6\|_{\mathrm{Frob}}^2} \frac{\|\mathbf{1}\|_{\mathrm{Frob}}^2 \|B_6\|^2}{\|B_5 B_6\|_{\mathrm{Frob}}^2} \\
&\leqslant \int_E \mathrm{d}'B_2 \, \mathrm{d}'B_3 \, \mathrm{d}'B_5 \, \mathrm{d}'B_6 \; \theta(B_2 B_3) \frac{\|\mathbf{1}\|_{\mathrm{Frob}}^2 \|B_6\|^2}{\|B_5 B_6\|_{\mathrm{Frob}}^2} \\
&\leqslant \left(\int \mathrm{d}'B_2 \, \mathrm{d}'B_3 \; \theta(B_2 B_3) \right) \theta(\mathbf{1}) \\
&= \left(1 + \frac{1}{r_1 - 1} + \frac{1}{r_2 - 1} + \frac{1}{r_3 - 1} \right) \left(1 + \frac{1}{r_4 - 1} \right),
\end{aligned}$$

the last by Lemma 5.9.

So we can apply Lemma 5.9 and obtain

$$\int_F \mathrm{d}'B_2 \cdots \mathrm{d}'B_\delta \, \frac{P_I(B)}{\|\hat{B}_1\|^2_{\mathrm{Frob}}} \leqslant \prod_{j=1}^{m-1} \left(1 + \sum_{j=i_k}^{i_{k+1}-1} \frac{1}{r_j - 1}\right) \tag{5.66}$$

$$\leqslant \left(1 + \frac{1}{m-1} \sum_{j=i_1}^{i_m-1} \frac{1}{r_j - 1}\right)^{m-1}, \tag{5.67}$$

using the inequality of arithmetic and geometric means. Since $r_j > 1$ for $j \leqslant i_m - 1$, we further obtain

$$\int_F \mathrm{d}'B_2 \cdots \mathrm{d}'B_\delta \, \frac{P_I(B)}{\|\hat{B}_1\|^2_{\mathrm{Frob}}} \leqslant \left(1 + \frac{\delta - 1}{m-1}\right)^{m-1}. \tag{5.68}$$

5.3.4. *Conclusion.* Combining (5.32), (5.57), (5.63), and (5.68), we obtain (note the cancellation of π),

$$\mathbb{E}\left[\|\mathrm{d}_\zeta f_A\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}\right] \tag{5.69}$$

$$\leqslant \frac{1}{\delta n} \sum_{m=1}^{k} \binom{\delta - m}{k - m}^2 \binom{k}{m}^{-1} \binom{m + n - 1}{m} \binom{\delta}{m} \left(1 + \frac{\delta - 1}{m-1}\right)^{m-1}.$$

By reordering the factorials, we have

$$\binom{\delta - m}{k - m}^2 \binom{k}{m}^{-1} \binom{\delta}{m} = \binom{\delta - m}{k - m} \binom{\delta}{k}. \tag{5.70}$$

As in the proof of Lemma I.37, we observe the identity

$$\sum_{m=0}^{k} \binom{\delta - m}{k - m} \binom{m + n - 1}{m} = \binom{\delta + n}{k}. \tag{5.71}$$

Moreover, since $m \leqslant k$,

$$\left(1 + \frac{\delta - 1}{m-1}\right)^{m-1} \leqslant \left(1 + \frac{\delta - 1}{k-1}\right)^{k-1} \tag{5.72}$$

(including $m = 1$ where the left-hand side is 1). Equations (5.69), (5.70), (5.71) and (5.72) give

$$\mathbb{E}\left[\|\mathrm{d}_\zeta f_A\|^{-2} \left\|\tfrac{1}{k!}\mathrm{d}^k_\zeta f_A\right\|^2_{\mathrm{Frob}}\right] \leqslant \frac{1}{n\delta} \binom{\delta}{k} \binom{\delta + n}{k} \left(1 + \frac{\delta - 1}{k-1}\right)^{k-1}. \tag{5.73}$$

This gives the first inequality of Proposition 5.3. For the second we argue as in the proof of Lemma I.37: the maximum value of $\left[\frac{1}{n\delta}\binom{\delta}{k}\binom{\delta+n}{k}\right]^{\frac{1}{k-1}}$ with $k \geqslant 2$ is reached at $k = 2$. Hence, for any $k \geqslant 2$,

$$\left[\frac{1}{n\delta}\binom{\delta}{k}\binom{\delta + n}{k}\right]^{\frac{1}{k-1}} \leqslant \left[\frac{1}{n\delta}\binom{\delta}{2}\binom{\delta + n}{2}\right]^{\frac{1}{k-1}} \leqslant \left[\frac{1}{4}\delta^2(\delta + n)\right]^{\frac{1}{k-1}}. \tag{5.74}$$

This concludes the proof of Proposition 5.3.

References

D. Armentano, C. Beltrán, P. Bürgisser, F. Cucker, and M. Shub (2016). "Condition Length and Complexity for the Solution of Polynomial Systems". In: *Found. Comput. Math.* DOI: 10/ggck9h.

W. Baur and V. Strassen (1983). "The Complexity of Partial Derivatives". In: *Theor. Comput. Sci.* 22.3, pp. 317–330. DOI: 10/dhrzs5.

C. Beltrán (2011). "A Continuation Method to Solve Polynomial Systems and Its Complexity". In: *Numer. Math.* 117.1, pp. 89–113. DOI: 10/c8cs5s.

C. Beltrán and A. Leykin (2012). "Certified Numerical Homotopy Tracking". In: *Exp. Math.* 21.1, pp. 69–83. DOI: 10/ggck73.

– (2013). "Robust Certified Numerical Homotopy Tracking". In: *Found. Comput. Math.* 13.2, pp. 253–295. DOI: 10/ggck74.

C. Beltrán and L. M. Pardo (2008). "On Smale's 17th Problem: A Probabilistic Positive Solution". In: *Found. Comput. Math.* 8.1, pp. 1–43. DOI: 10/b94hmw.

– (2009). "Smale's 17th Problem: Average Polynomial Time to Compute Affine and Projective Solutions". In: *J. Amer. Math. Soc.* 22.2, pp. 363–385. DOI: 10/c32q5b.

– (2011). "Fast Linear Homotopy to Find Approximate Zeros of Polynomial Systems". In: *Found. Comput. Math.* 11.1, pp. 95–129. DOI: 10/ffch6h.

C. Beltrán and M. Shub (2009). "Complexity of Bezout's Theorem. VII. Distance Estimates in the Condition Metric". In: *Found. Comput. Math.* 9.2, pp. 179–195. DOI: 10/fdrfmz.

S. Boucheron, G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press. DOI: 10/dnmr.

P. Bürgisser (2000). *Completeness and Reduction in Algebraic Complexity Theory*. Algorithms and Computation in Mathematics. Springer-Verlag. DOI: 10/d9n4.

– (2017). "Condition of Intersecting a Projective Variety with a Varying Linear Subspace". In: *SIAM J. Appl. Algebra Geom.* 1.1, pp. 111–125. DOI: 10/ggck9p.

P. Bürgisser and F. Cucker (2011). "On a Problem Posed by Steve Smale". In: *Ann. of Math. (2)* 174.3, pp. 1785–1836. DOI: 10/djcp42.

– (2013). *Condition: The Geometry of Numerical Algorithms*. Grundlehren Der Mathematischen Wissenschaften 349. Springer. DOI: 10/dsfq.

J.-P. Dedieu and M. Shub (1999). "Multihomogeneous Newton Methods". In: *Math. Comp.* 69.231, pp. 1071–1099. DOI: 10/c39qj9.

J. W. Demmel (1987). "On Condition Numbers and the Distance to the Nearest Ill-Posed Problem". In: *Numer. Math.* 51.3, pp. 251–289.

H. Federer (1959). "Curvature Measures". In: *Trans. Am. Math. Soc.* 93.3, pp. 418–491. DOI: 10/d5bdvc.

M. Giusti, G. Lecerf, and B. Salvy (2001). "A Gröbner Free Alternative for Polynomial System Solving". In: *J. Complexity* 17.1, pp. 154–211. DOI: 10/fpzjtc.

J. D. Hauenstein and A. C. Liddell (2016). "Certified Predictor–Corrector Tracking for Newton Homotopies". In: *J. Symb. Comput.* 74, pp. 239–254. DOI: 10/ggck7j.

J. D. Hauenstein and F. Sottile (2012). "Algorithm 921: alphaCertified: Certifying Solutions to Polynomial Systems". In: *ACM Trans. Math. Softw.* 38.4, pp. 1–20. DOI: 10/ggck9q.

R. Howard (1993). "The Kinematic Formula in Riemannian Homogeneous Spaces". In: *Mem. Amer. Math. Soc.* 106.509. DOI: 10/ggck9w.

S. Ji, J. Kollar, and B. Shiffman (1992). "A Global Lojasiewicz Inequality for Algebraic Varieties". In: *Trans. Am. Math. Soc.* 329.2, p. 813. DOI: 10/ddzp62.

P. Lairez (2017). "A Deterministic Algorithm to Compute Approximate Roots of Polynomial Systems in Polynomial Average Time". In: *Found. Comput. Math.* 17.5, pp. 1265–1292. DOI: 10/ggck6w.

– (2020). "Rigid Continuation Paths I. Quasilinear Average Complexity for Solving Polynomial Systems". In: *J. Amer. Math. Soc.* 33.2, pp. 487–526. DOI: 10/ggck65.

Y. N. Lakshman (1991). "A Single Exponential Bound on the Complexity of Computing Gröbner Bases of Zero Dimensional Ideals". In: *Effective Methods in*

Algebraic Geometry. Ed. by T. Mora and C. Traverso. Birkhäuser, pp. 227–234. DOI: 10/dm2j9z.

G. Malod and N. Portier (2008). "Characterizing Valiant's Algebraic Complexity Classes". In: *J. Complex.* Computational Algebraic Geometry Workshop 24.1, pp. 16–38. DOI: 10/d5hmhz.

N. Nisan (1991). "Lower Bounds for Non-Commutative Computation". In: *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*. STOC '91. ACM, pp. 410–418. DOI: 10/dw5zfd.

V. Y. Pan (2001). "Univariate Polynomials: Nearly Optimal Algorithms for Factorization and Rootfinding". In: *Proceedings of the 2001 International Symposium on Symbolic and Algebraic Computation*. ISSAC '01. ACM, pp. 253–267. DOI: 10/bjqkgg.

V. Pan (1996). "Optimal and Nearly Optimal Algorithms for Approximating Polynomial Zeros". In: *Comput. Math. Appl.* 31.12, pp. 97–138. DOI: 10/bfnwbq.

J. Renegar (1987). "On the Worst-Case Arithmetic Complexity of Approximating Zeros of Polynomials". In: *J. Complex.* 3.2, pp. 90–113. DOI: 10/fqrc56.

– (1989). "On the Worst-Case Arithmetic Complexity of Approximating Zeros of Systems of Polynomials". In: *SIAM J. Comput.* 18.2, pp. 350–370. DOI: 10/df2t8j.

W. Rudin (1980). *Function theory in the unit ball of* $\mathbb{C}^n$. Grundlehren der Mathematischen Wissenschaften 241. Springer.

S. M. Rump and S. Graillat (2010). "Verified Error Bounds for Multiple Roots of Systems of Nonlinear Equations". In: *Numer. Algorithms* 54.3, pp. 359–377. DOI: 10/dfbbjz.

M. Shub (2009). "Complexity of Bezout's Theorem. VI. Geodesics in the Condition (Number) Metric". In: *Found. Comput. Math.* 9.2, pp. 171–178. DOI: 10/cr5t6q.

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

– (1993b). "Complexity of Bezout's Theorem. III. Condition Number and Packing". In: *J. Complexity* 9.1, pp. 4–14. DOI: 10/d9hw6h.

– (1993c). "Complexity of Bézout's Theorem. I. Geometric Aspects". In: *J. Amer. Math. Soc.* 6.2, pp. 459–501. DOI: 10/fk6z2g.

– (1994). "Complexity of Bezout's Theorem. V. Polynomial Time". In: *Theoret. Comput. Sci.* 133.1, pp. 141–164. DOI: 10/fp47hg.

– (1996). "Complexity of Bezout's Theorem. IV. Probability of Success; Extensions". In: *SIAM J. Numer. Anal.* 33.1, pp. 128–148. DOI: 10/dwtpvj.

S. Smale (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, pp. 185–196.

– (1997). "Complexity Theory and Numerical Analysis". In: *Acta Numer.* 6, pp. 523–551. DOI: 10/d834qt.

– (1998). "Mathematical Problems for the next Century". In: *Math. Intell.* 20.2, pp. 7–15. DOI: 10/bwg3j3.

S. Telen, M. Van Barel, and J. Verschelde (2020). *A Robust Numerical Path Tracking Algorithm for Polynomial Homotopy Continuation*. arXiv: 1909.04984.

S. Timme (2020). *Mixed Precision Path Tracking for Polynomial Homotopy Continuation*. arXiv: 1902.02968.

S. Toda (1992). "Classes of Arithmetic Circuits Capturing the Complexity of Computing the Determinant". In: *IEICE Trans. Inf. Syst.* E75-D.1, pp. 116–124.

L. G. Valiant (1979). "Completeness Classes in Algebra". In: *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*. STOC '79. ACM, pp. 249–261. DOI: 10/cbp8bz.

– (1982). "Reducibility by Algebraic Projections". In: *Logic and Algorithmic (Zurich, 1980)*. Vol. 30. Monograph. Enseign. Math. Univ. Genève, pp. 365–380.

Institut für Mathematik, Technische Universität Berlin, Germany
Email address: pbuerg@math.tu-berlin.de

Department of Mathematics, City University of Hong Kong, Hong Kong
Email address: macucker@cityu.edu.hk

Inria, France
Email address: pierre.lairez@inria.fr