PROBABILISTIC CONDITION NUMBER ESTIMATES FOR REAL POLYNOMIAL SYSTEMS I: A BROADER FAMILY OF DISTRIBUTIONS

ALPEREN ERGÜR, GRIGORIS PAOURIS, AND J. MAURICE ROJAS

Abstract. We consider the sensitivity of real roots of polynomial systems with respect to perturbations of the coefficients. In particular — for a version of the condition number defined by Cucker, Krick, Malajovich, and Wschebor — we establish new probabilistic estimates that allow a much broader family of measures than considered earlier. We also generalize further by allowing over-determined systems. Along the way, we derive new Lipshitz estimates for polynomial maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, extending earlier work of Kellog on the case $m=1$, which may be of independent interest.

In Part II, we study smoothed complexity and how sparsity (in the sense of restricting which monomial terms can appear) can help further improve earlier condition number estimates.

1. Introduction

When designing algorithms for polynomial system solving, it quickly becomes clear that complexity is governed by more than simply the number of variables and degrees of the equations. Numerical solutions are meaningless without further information on the spacing of the roots, not to mention their sensitivity to perturbation. A mathematically elegant means of capturing this sensitivity is the notion of *condition number* (see, e.g., [3, 6] and our discussion below).

A subtlety behind complexity bounds incorporating the condition number is that computing it — even within a large multiplicative error — is provably as hard as computing the numerical solution one seeks in the first place (see, e.g., [14] for a precise statement in the linear case). However, it is now known that the condition number admits *probabilistic* bounds, thus enabling its use in average-case analysis, high probability analysis, and smoothed analysis of numerical algorithms. In fact, this probabilistic approach has revealed (see, e.g., [2, 5, 18]) that, in certain settings, numerical solving can be done in polynomial-time on average, in spite of numerical solving having exponential worst-case complexity.

The numerical approximation of complex roots provides an instructive example of how one can profit from randomization.

First, there are classical reductions showing that deciding the existence of complex roots for systems of polynomials in $\bigcup_{m,n\in\mathbb{N}}(\mathbb{Z}[x_1,\ldots,x_n])^m$ is already **NP**-hard. However, classical algebraic geometry (e.g., Bertini's Theorem and Bézout's Theorem [24]) tells us that, with probability 1, the number of complex roots of a *random* system of homogeneous polynomials, $P:=(p_1,\ldots,p_m)\in\mathbb{C}[x_1,\ldots,x_n]$ (with each p_i having fixed positive degree d_i), is 0, $\prod_{i=1}^n d_i$, or infinite, according as $m>n-1$, $m=n-1$, or $m<n-1$. (Any probability measure absolutely continuous with respect to Lebesgue measure will do in the preceding statement.)

Secondly, examples like $P:=(x_1-x_2^2, x_2-x_3^2,\ldots,x_{n-1}-x_n^2,(2x_n-1)(3x_n-1))$, which has affine roots $\left(2^{-2^{n-1}},\ldots,2^{-2^0}\right)$ and $\left(3^{-2^{n-1}},\ldots,3^{-2^0}\right)$, reveal that the number of digits necessary to distinguish the coordinates of roots of P may be exponential in n (among other parameters). However, it is now known via earlier work on discriminants and random polynomial systems (see, e.g., [8, Thm. 5]) that the number of digits needed to separate roots of P is polynomial in n *with high probability*, assuming the coefficients are rational,

A.E. was partially supported by NSF grant CCF-1409020 and NSF CAREER grant DMS-1151711. G.P. was partially supported by BSF grant 2010288 and NSF CAREER grant DMS-1151711. J.M.R. was partially supported by NSF grant CCF-1409020.

and the polynomial degrees and coefficient heights are bounded. More simply, a classical observation from the theory of resultants (see, e.g., [7]) is that, for any positive continuous probability measure on the coefficients, P having a root with Jacobian matrix possessing small determinant is a rare event. So, with high probability, small perturbations of a P with no degenerate roots should still have no degenerate roots. More precisely, we recall below a version of the condition number used in [25, 2, 18]. Let us also recall that the *singular values* of a matrix $T\in\mathbb{R}^{k\times(n-1)}$ are the (nonnegative) square roots of the eigenvalues of $T^\top T$, where $T^\top$ denotes the transpose of T.

Definition 1.1. *Given $n, d_1, \ldots, d_m \in \mathbb{N}$ and $i\in\{1,\ldots,m\}$, let $N_i := \binom{n+d_i-1}{d_i}$ and, for any homogenous polynomial $p_i\in\mathbb{R}[x_1,\ldots,x_n]$ with $\deg p_i \leq d_i$, note that the number of monomial terms of p_i is at most N_i. Letting $\alpha=(\alpha_1,\ldots,\alpha_n)$ and $x^\alpha = x_1^{\alpha_1}\cdots x_n^{\alpha_n}$, let $c_{i,\alpha}$ denote the coefficient of x^α in p_i, and set $P:=(p_1,\ldots,p_m)$. Also, let us define the* Weyl-Bombieri norms *of p_i and P to be, respectively,*

$$\|p_i\|_W := \sqrt{\sum_{\alpha_1+\cdots+\alpha_n=d_i} \frac{|c_{i,\alpha}|^2}{\binom{d}{\alpha}}} \text{ and } \|P\|_W := \sqrt{\sum_{i=1}^m \|p_i\|_W^2}.$$

Let $\Delta_m\in\mathbb{R}^{m\times m}$ be the diagonal matrix with diagonal entries $\sqrt{d_1},\ldots,\sqrt{d_m}$ and let $DP(x)|_{T_xS^{n-1}}$ denote the linear map defined by the Jacobian matrix of the polynomial system P, evaluated at a point x, and restricted to the tangent space of the (real) $(n-1)$-sphere S^{n-1} at x. Finally, when $m=n-1$, we define the (normalized) local condition number (for solving $P=\mathbf{O}$) *to be $\tilde{\mu}_{\text{norm}}(P,x) := \|P\|_W \sigma_{\max}\left(DP(x)|_{T_xS^{n-1}}^{-1}\Delta_{n-1}\right)$ or $\tilde{\mu}_{\text{norm}}(P,x) := \infty$, according as $DP(x)$ is invertible or not, where $\sigma_{\max}(A)$ is the largest singular value of a matrix A.* $\diamond$

Clearly, $\tilde{\mu}_{\text{norm}}(P,x)\to\infty$ as P approaches a system possessing a degenerate root $\zeta\in\mathbb{P}^{n-1}_{\mathbb{C}}$ and x approaches ζ. The intermediate normalizations in the definition are useful for geometric interpretations of $\tilde{\mu}_{\text{norm}}$ in terms of discriminant varieties (see, e.g., [10] and Theorem 2.1 below). In particular, the preceding condition number (in the special case $m=n-1$) was a central ingredient in the recent positive solution to *Smale's 17th Problem* [2, 18]: For the problem of numerically approximating a *single* complex root of a polynomial system, a particular randomization model (independent complex Gaussian coefficients with specially chosen variances) enables *polynomial-time* average-case complexity, in the face of exponential deterministic complexity.[1]

It is natural to seek similar speed-ups for the harder problem of numerically approximating real roots of real polynomial systems. However, an important subtlety one must consider is that the number of real roots of $n-1$ homogeneous polynomials in n variables (of fixed degree) is no longer constant with probability 1, even if the probability measure for the coefficients is continuous and positive. In particular, some systems can be perturbed infinitesimally from having no real roots at all to having many real roots, and vice-versa. A condition number for real solving that takes these subtleties into account was developed and applied in the seminal series of papers [9, 10, 11]. In these papers, the authors performed a probabilistic analysis assuming the coefficients were centered Gaussian distribution with very specially chosen variances and independent coordinates.

[1]Here, "complexity" simply means the total number of field operations over $\mathbb{C}$ needed to find an approximation guaranteed to be a start point for Newton iteration yielding quadratic convergence to a true root (see, e.g., [3, Ch. 8]).

Definition 1.2. [9] *Let* $\tilde{\kappa}(P,x) := \frac{\|P\|_W}{\sqrt{\|P\|_W^2 \tilde{\mu}_{\text{norm}}(P,x)^{-2} + \|P(x)\|_2^2}}$ *and* $\tilde{\kappa}(P) := \sup_{x\in S^{n-1}} \tilde{\kappa}(P,x)$. *We respectively call* $\tilde{\kappa}(P,x)$ *and* $\tilde{\kappa}(P)$ *the* local *and* global *condition numbers for real solving.* ⋄

Note in particular that a large condition number can be caused not only by a root with small Jacobian determinant but also by the existence of a critical point for P with small corresponding critical value. So a large condition number for real solving is meant to capture the spontaneous creation of real roots, as well as the bifurcation of a single degenerate root into multiple distinct real roots, arising from small perturbations of the coefficients.

Our main results, Theorems 3.9 and 3.10 in Section 3.4, show that useful condition number estimates can be derived for a broader class of probability measures than considered earlier. Our results thus allow dependence and non-Gaussian distributions and, unlike the existing literature, our methods do not use any additional algebraic structure, e.g., invariance under the unitary group acting linearly on the variables (as in [25, 9, 10, 11]). This aspect also allows us to begin to address sparse polynomials (in the sequel to this paper), where linear changes of variables would destroy sparsity.

To compare our results with earlier estimates, let us first recall a central estimate from [11].

Theorem 1.1. [11, Thm. 1.2] *Let* $P := (p_1, \dots, p_{n-1})$ *be a random system of homogeneous* n*-variate polynomials where* $n \geq 3$ *and* $p_i(x) := \sum_{\alpha_1+\cdots+\alpha_n=d_i} \sqrt{\binom{d_i}{\alpha}} c_{i,\alpha} x^\alpha$ *with the* $c_{i,\alpha}$ *independent real Gaussian random variables having mean* 0 *and variance* 1. *Then, letting* $N := \sum_{i=1}^{n-1} \binom{n+d_i-1}{d_i}$, $d := \max_i d_i$, $M' := 1 + 8d^2\sqrt{(n-1)^5 N \prod_{i=1}^{n-1} d_i}$, *and* $t \geq \sqrt{\frac{n-1}{4\prod_{i=1}^{n-1} d_i}}$, *we have:*

1. $\text{Prob}(\tilde{\kappa}(P) \geq tM') \leq \frac{\sqrt{1+\log(tM')}}{t}$
2. $\mathbb{E}(\log(\tilde{\kappa}(P))) \leq \log(M') + \sqrt{\log M'} + \frac{1}{\sqrt{\log M'}}$.

The expanded class of distributions we allow for the coefficients of P satisfy the following more flexible hypotheses:

Notation 1.2. *For any* $d_1, \dots, d_m \in \mathbb{N}$ *and* $i \in \{1, \dots, m\}$, *let* $d := \max_i d_i$, $N_i := \binom{n+d_i-1}{d_i}$, *and assume* $C_i = (c_{i,\alpha})_{\alpha_1+\cdots+\alpha_n=d_i}$ *is an independent random vector in* $\mathbb{R}^{N_i}$ *with probability distribution satisfying:*

1. *(Centering) For any* $\theta \in S^{N_i-1}$ *we have* $\mathbb{E}\langle C_i, \theta\rangle = 0$.
2. *(Sub-Gaussian) There is a* $K > 0$ *such that for every* $\theta \in S^{N_i-1}$ *we have*
$$\text{Prob}\left(|\langle C_i, \theta\rangle| \geq t\right) \leq 2e^{-t^2/K^2} \text{ for all } t > 0.$$
3. *(Small Ball) There is a* $c_0 > 0$ *such that for every vector* $a \in \mathbb{R}^{N_i}$ *we have*
$$\text{Prob}\left(|\langle a, C_i\rangle| \leq \varepsilon\|a\|_2\right) \leq c_0\varepsilon \text{ for all } \varepsilon > 0. \ \diamond$$

Any random vector $X \in \mathbb{R}^{N_i}$ with a bounded density function f supported on a compact set satisfies the above assumptions. In this case the constants c_0 and K will depend on the density f. We are mainly interested in distributions which satisfies the above assumptions with universal constants c_0 and K. Nevertheless, to maintain generality we kept track of the dependency on c_0 and K, hence our theorems hold for any distribution satisfying the above assumptions.

The standard Gaussian distribution is a typical example of a random vector satisfying our assumptions with universal constants. Note that it is easy to create one dimensional random variables satisfying the assumptions above. For example, any random variable X_i

with density function $f(t) = c_p e^{-|t|^p}$ where c_p is the appropriate normalization and $p > 2$ would do the assumptions with universal constants. Now for a random vector $X = (X_i)_{i=1}^{N_i}$ with independent coordinates, if X_i satisfy the above assumptions with universal constants then X satisfies the assumptions with universal constants as well. This is neither obvious nor trivial, it is a significant fact recently proved by Rudelson and Vershynin [23].

A non-trivial example of a random vector satisfying our assumptions is the uniform measure on $B_p^{N_i}$, $p > 2$, where $B_p^{N_i} := \{x \in R^{N_i} : \sum_{j=1}^{N_i} x_j^p \leq 1\}$. In this case the subgaussian assumption follows from ([1], Section 6) and the Small Ball Assumption is a direct consequence of the well-known fact that $B_p^{N_i}$ satisfies the Hyperplane Conjecture. Another important example is the uniform distribution on the unit sphere.

A simplified summary of our main results (Theorems 3.9 and 3.10 from Section 3.4), in the special case of square dense systems, is the following:

Corollary 1.3. *There is an absolute constant $A>0$ with the following property: Let $P := (p_1, \ldots, p_{n-1})$ be a random system of homogenous n-variate polynomials where $p_i(x) := \sum_{\alpha_1+\cdots+\alpha_n=d_i} \sqrt{\binom{d_i}{\alpha}} c_{i,\alpha} x^\alpha$ and $C_i = (c_{i,\alpha})_{\alpha_1+\cdots+\alpha_n=d_i}$ are random vectors satisfying the Centering, Sub-Gaussian and Small Ball assumptions, with underlying constants c_0 and K. Then, letting $d := \max_i d_i$, $N := \sum_{i=1}^{n-1} \binom{n+d_i-1}{d_i}$, and $M := K\sqrt{c_0 N}\,(c_0 K d^2 \log(ed))^{n-2}$, the following bounds hold:*

1. $\operatorname{Prob}(\tilde{\kappa}(P) \geq tM) \leq \begin{cases} \frac{3A}{t} & ; \text{ if } A \leq t \leq A(ed)^{n-1} \\ \frac{3A}{t}\left(\frac{t}{A(ed)^{n-1}}\right)^{\frac{1}{2\log(ed)}} & ; \text{ if } t \geq A(ed)^{n-1} \end{cases}$
2. $\mathbb{E}(\log \tilde{\kappa}(P)) \leq 1 + \log M$.

Corollary 1.3 is proved in Section 3.4. Our main theorem in Section 3.4 includes stronger estimates, we preferred simplicity in corollary 1.3 with the expense of loosening our bounds.

Theorem 1.1 is restricted to a particular family of Gaussian distributions, and assumes $m=n-1$. For the case $m=n-1$, our estimates in Section 3.4 are comparable to Theorem 1.1: for $t \leq (ed)^{n-1}$ our bounds are slightly better than the bounds of Theorem 1.1, and for the other phase $t \geq (ed)^{n-1}$ bounds of Theorem 1.1 are slightly better than ours. Our estimates on $\mathbb{E}(\log \tilde{\kappa}(P))$ are quite similar to the Theorem 1.1 with the exception that we loose a factor 2.

Our techniques enable better bounds for arbitrary $m \times n$ systems when m is larger than $n-1$: See the next section for the definition of a condition number enabling $m > n-1$, and the statements of Theorems 3.9 and 3.10 for our most general condition number bounds. There appear to have been no probabilistic condition number estimates for the case $m>n-1$ until now.

To the best of our knowledge, the only other result toward estimating condition numbers of non-Gaussian random polynomial systems is due to Nguyen [20]. However, in [20] the degrees of the polynomials are assumed to be bounded by a small fraction of the number of variables, and the quantity analyzed in [20] is not the condition number considered in [25] or [9, 10, 11].

It would certainly be nice to know the correct decay rate for the probability that the condition number is large: This is not even known in the restricted Gaussian case of Cucker, Malajovich, Krick, and Wschebor. So we also prove *lower* bounds for the condition number of a random polynomial system. To establish these bounds, we need one more assumption on the randomness.

Notation 1.4. *For any* $d_1,\ldots,d_m\in\mathbb{N}$ *and* $i\in\{1,\ldots,m\}$*, let* $d:=\max_i d_i$*,* $N_i:=\binom{n+d_i-1}{d_i}$*, and assume* $C_i=(c_{i,\alpha})_{\alpha_1+\cdots+\alpha_n=d_i}$ *is an independent random vector in* $\mathbb{R}^{N_i}$ *with probability distribution satisfying:*
4. (Euclidean Small Ball) There is a constant $\tilde{c}_0>0$ *such that for every* $\varepsilon>0$ *we have*
$$\mathrm{Prob}\left(\|C_i\|_2\leq\varepsilon\sqrt{N_i}\right)\leq(\tilde{c}_0\varepsilon)^{N_i}.\ \diamond$$

Remark 1.5. *If the vectors* C_i *have independent coordinates satisfying the Centering and Small Ball Assumptions, then Lemma 3.4 from Section 3.3 implies that the Euclidean Small Ball Assumption holds as well. Moreover, if the* C_i *are uniformly distributed on a convex body* X *and satisfy our Centering and Sub-Gaussian assumptions, then a result of Jean Bourgain* [4] *(see also* [12] *or* [17] *for alternative proofs) implies that both the Small Ball and Euclidean Small Ball Assumptions hold, and with* $\tilde{c}_0$ *depending only on the Sub-Gaussian constant* K *(not the convex body* X*).* $\diamond$

Corollary 1.6. *Suppose* $n,d\geq 3$*,* $m=n-1$*, and* $d_j=d$ *for all* $j\in\{1,\ldots,n-1\}$*. Also let* $P:=(p_1,\ldots,p_m)$ *be a random polynomial system satisfying our Centering, Sub-Gaussian, Small Ball, and Tiny Ball assumptions, with respective underlying constants* K *and* $\tilde{c}_0$*. Then there are constants* $A_2\geq A_1>0$ *such that*
$$A_1(n\log(d)+d\log(n))\leq\mathbb{E}(\log\tilde{\kappa}(P))\leq A_2(n\log(d)+d\log(n)).\quad\blacksquare$$

Corollary 1.6 follows immediately from a more general estimate: Lemma 3.13 from Section 3.3. It would certainly be more desirable to know bounds within a constant multiple of $\tilde{\kappa}(P)$ instead. We discuss more refined estimates of the latter kind in Section 3.5, after the proof of Lemma 3.13.

As we close our introduction, we point out that one of the tools we developed for our main theorems may be of independent interest: Theorem 2.3 of the next section extends, to polynomial systems, an earlier estimate of Kellog [16] on the norm of the derivative of a single multivariate polynomial.

2. Technical Background

We start by defining an inner product structure on spaces of polynomial systems. For n-variate degree d homogenous polynomials $f(x):=\sum_{|\alpha|=d}b_\alpha x^\alpha$ and $g(x):=\sum_{|\alpha|=d}c_\alpha x^\alpha$, their *Weyl-Bombieri inner product* is defined as
$$\langle f,g\rangle_W:=\sum_{|\alpha|=d}\frac{b_\alpha c_\alpha}{\binom{d}{\alpha}}.$$

It is known (see, e.g., [19, Thm. 4.1]) that for $U\in O(n)$ we have
$$\langle f\circ U,g\circ U\rangle_W=\langle f,g\rangle_W.$$
Let $D:=(d_1,\ldots,d_m)$ and let H_D denote the space of (real) $m\times n$ systems of homogenous n-variate polynomials with degrees respectively bounded from above by $d_1,\ldots,d_m$. Then for $F:=(f_1,\ldots,f_m)\in H_D$ and $G:=(g_1,\ldots,g_m)\in H_D$ the Weyl-Bombieri inner product *for two polynomial systems* is defined as follows:
$$\langle F,G\rangle_W=\sum_{i=1}^m\langle f_i,g_i\rangle_W.$$

A geometric justification for the definition of the condition number $\tilde{\kappa}$ can then be derived as follows: For $x\in S^{n-1}$ we define the set of polynomial systems with singularity at x as

$$\Sigma_{\mathbb{R}}(x) := \{P \in H_D \mid P \text{ has a multiple root at } x\}$$

We then define $\Sigma_{\mathbb{R}}$ (the real part of the disciminant variety) to be:

$$\Sigma_{\mathbb{R}} := \{P \in H_D \mid P \text{ has a multiple root in } S^n\} = \bigcup_{x \in S^n} \Sigma_{\mathbb{R}}(x).$$

Using the Weyl-Bombieri inner-product to define the underlying distance, we point out the following important geometric characterization of $\tilde{\kappa}$:

Theorem 2.1. [10, Prop. 3.1] *When $m=n-1$ we have $\tilde{\kappa}(P) = \frac{\|P\|_W}{\text{Dist}(P,\Sigma_{\mathbb{R}})}$ for all $P \in H_D$.* ■

We call a polynomial system $P=(p_1,\ldots,p_m)$ with $m=n-1$ (resp. $m \geq n$) *square* (resp. *over-determined*). Newton's method for over-determined systems was studied in [13]. So now that we have a geometric characterization of the condition number for square systems it will be useful to also have one for over-determined systems.

Definition 2.1. *For any system of homogeneous polynomials $P \in (\mathbb{R}[x_1,\ldots,x_n])^m$ let us set*

$$L(P,x) = \sqrt{\sigma_{\min}\left(\Delta_m^{-1} DP(x)|_{T_x S^{n-1}}\right)^2 + \|P(x)\|_2^2},$$

where $\sigma_{\min}(A)$ is the smallest singular value of a matrix A. Let us then define $\tilde{\kappa}(P,x) = \frac{\|P\|_W}{L(P,x)}$ and $\tilde{\kappa}(P) = \sup_{x \in S^{n-1}} \kappa(P,x)$. ⋄

The quantity $\min_{x \in S^{n-1}} L(P,x)$ thus plays the role of $\text{Dist}(P,\Sigma_{\mathbb{R}})$ in the more general setting of $m \geq n-1$. We now recall an important observation from Section 2 of [10]: Setting $D_x(P) := DP(x)|_{T_x S^{n-1}}$ we have

$$\sigma_{\min}(\Delta_{n-1}^{-1} D_x(P)) = \sigma_{\max}\left(D_x(P)^{-1}\Delta_{n-1}\right)^{-1}$$

when $m=n-1$ and $D_x(P)$ is invertible. So by the definition of $\tilde{\mu}_{\text{norm}}(P,x)$ we have

$$L(P,x) = \sqrt{\sigma_{\max}\left(D_x(P)^{-1}\Delta_{n-1}\right)^{-2} + \|P(x)\|_2^2} = \sqrt{\|P\|_W^2 \tilde{\mu}_{\text{norm}}(P,x)^{-2} + \|P(x)\|_2^2}$$

and thus our more general definition agrees with the classical definition in the square case.

Since the W-norm of a random polynomial system has strong concentration properties for a broad variety of distributions (see, e.g., [26]), we will be interested in the behavior of $L(P,x)$. So let us define $\mathcal{L}(x,y) := \sqrt{\|\Delta_m^{-1} D^{(1)} P(x)(y)\|_2^2 + \|P(x)\|_2^2}$. It follows directly that $L(P,x) = \inf_{\substack{y \perp x \\ y \in S^{n-1}}} \mathcal{L}(x,y)$.

We now recall a classical result of Kellog:

Theorem 2.2. [16] *Let $p \in \mathbb{R}[x_1,\ldots,x_n]$ have degree d and set $\|p\|_\infty := \sup_{x \in S^{n-1}} |p(x)|$ and $\|D^{(1)}p\|_\infty := \max_{x,u \in S^{n-1}} |D^{(1)}p(x)(u)|$. Then:*

1. $\|D^{(1)}p\|_\infty \leq d^2 \|p\|_\infty$.
2. *If p is homogenous then we also have $\|D^{(1)}p\|_\infty \leq d\|p\|_\infty$.* ■

For any system of homogeneous polynomials $P := (p_1,\ldots,p_m) \in (\mathbb{R}[x_1,\ldots,x_n])^m$ define $\|P\|_\infty := \sup_{x \in S^{n-1}} \sqrt{\sum_{i=1}^m p_i(x)^2}$. Let $DP(x)(u)$ denote the image of the vector u under the linear operator $DP(x)$, and set

$$\left\|D^{(1)}P\right\|_\infty := \sup_{x,u \in S^{n-1}} \|DP(x)(u)\|_2 = \sup_{x,u \in S^{n-1}} \sqrt{\sum_{i=1}^m \langle \nabla p_i(x), u\rangle^2}.$$

Theorem 2.3. *Let $P := (p_1, \ldots, p_m) \in (\mathbb{R}[x_1, \ldots, x_n])^m$ be a polynomial system with p_i homogeneous of degree d_i for each i and set $d := \max_i d_i$. Then:*

(1) $\|D^{(1)}P\|_\infty \leq d^2 \|P\|_\infty$.
(2) *If* $\deg(p_i) = d$ *for all* $i \in \{1, \ldots, m\}$ *then we also have* $\|D^{(1)}P\|_\infty \leq d\|P\|_\infty$.

Proof. Let (x_0, u_0) be such that $\|D^{(1)}P\|_\infty = \|DP(x_0)(u_0)\|_2$ and let $\alpha := (\alpha_1, \ldots, \alpha_m)$ where $\alpha_i := \frac{\langle \nabla p_i(x_0), u_0\rangle}{\|D^{(1)}P\|_\infty}$. Note that $\|\alpha\|_2 = 1$. Now define a polynomial $q \in \mathbb{R}[x_1, \ldots, x_n]$ of degree d via $q(x) := \alpha_1 p_1(x) + \alpha_2 p_2(x) + \cdots + \alpha_m p_m(x)$ and observe that

$$\nabla q(x) = \left(\alpha_1 \frac{\partial p_1}{\partial x_1} + \alpha_2 \frac{\partial p_2}{\partial x_1} + \cdots + \alpha_m \frac{\partial p_m}{\partial x_1}, \ldots, \alpha_1 \frac{\partial p_1}{\partial x_n} + \alpha_2 \frac{\partial p_2}{\partial x_n} + \cdots + \alpha_m \frac{\partial p_m}{\partial x_n}\right),$$

$$\langle \nabla q, u\rangle = u_1 \left(\alpha_1 \frac{\partial p_1}{\partial x_1} + \alpha_2 \frac{\partial p_2}{\partial x_1} + \cdots + \alpha_m \frac{\partial p_m}{\partial x_1}\right) + \cdots + u_n \left(\alpha_1 \frac{\partial p_1}{\partial x_n} + \alpha_2 \frac{\partial p_2}{\partial x_n} + \cdots + \alpha_m \frac{\partial p_m}{\partial x_n}\right),$$

and $\langle \nabla q(x), u\rangle = \sum_{i=1}^m \alpha_i \langle \nabla p_i(x), u\rangle$. In particular, for our chosen x_0 and u_0, we have

$$\langle \nabla q(x_0), u_0\rangle = \sum_{i=1}^m \alpha_i \langle \nabla p_i(x_0), u_0\rangle = \sum_{i=1}^m \frac{\langle \nabla p_i(x_0), u_0\rangle^2}{\|D^{(1)}P\|_\infty} = \left\|D^{(1)}P_\infty\right\|.$$

Using the first part of Kellog's Theorem we have

$$\|D^{(1)}P\|_\infty \leq \sup_{x,u \in S^{n-1}} |\langle \nabla q(x), u\rangle| \leq d^2 \|q\|_\infty.$$

Now we observe by the Cauchy-Schwarz Inequality that

$$\|q\|_\infty = \sup_{x \in S^{n-1}} \left|\sum_{i=1}^m \alpha_i p_i(x)\right| \leq \sup_{x \in S^{n-1}} \sqrt{\sum_{i=1}^m p_i(x)^2}.$$

So we conclude that $\|D^{(1)}P\|_\infty \leq d^2 \|q\|_\infty \leq d^2 \sup_{x \in S^{n-1}} \sqrt{\sum_{i=1}^m p_i(x)^2} = d^2 \|P\|_\infty$. We also note that in the case $\deg(p_i) = d$ for all i, $q(x)$ is a homogenous polynomial of degree d. So for this special case, the second part of Kellog's Theorem directly implies $\|D^{(1)}P\|_\infty \leq d\|P\|_\infty$. ■

Using our extension of Kellog's Theorem to polynomial systems, we develop useful estimates for $\|P\|_\infty$ and $\|D^{(i)}P\|_\infty$. In what follows, we call a subset $\mathcal{N}$ of a metric space X a *δ-net on X* if and only if the maximal distance between a point of X and a point of $\mathcal{N}$ is δ. A basic fact we'll use repeatedly is that, for any $\delta > 0$ and *compact* X, one can always find a finite δ-net for X.

Lemma 2.4. *Let $P := (p_1, \ldots, p_m) \in (\mathbb{C}[x_1, \ldots, x_n])^m$ be a system of homogenous polynomials, $\mathcal{N}$ a δ-net on S^{n-1}, and set $d := \max_i d_i$. Let $\max_{\mathcal{N}}(P) := \sup_{y \in \mathcal{N}} \|P(y)\|_2$. Similarly let us define $\max_{\mathcal{N}^{k+1}}(D^{(k)}P) := \sup_{x,u_1,\ldots,u_k \in \mathcal{N}} \|D^{(k)}P(x)(u_1, \ldots, u_k)\|_2$, and set $\left\|D^{(k)}P\right\|_\infty := \sup_{x,u_1,\ldots,u_k \in S^{n-1}} \left\|D^{(k)}P(x)(u_1, \ldots, u_k)\right\|_2$. Then:*

(1) $\|P\|_\infty \leq \frac{\max_{\mathcal{N}}(P)}{1-\delta d^2}$ *and* $\|D^{(k)}P\|_\infty \leq \frac{\max_{\mathcal{N}^{k+1}}(D^{(k)}P)}{1-\delta d^2\sqrt{k+1}}$.
(2) *If* $\deg(p_i) = d$ *for each* $i \in \{1, \ldots, m\}$ *then we have*

$$\|P\|_\infty \leq \frac{\max_{\mathcal{N}}(P)}{1 - \delta d} \text{ and } \|D^{(k)}P\|_\infty \leq \frac{\max_{\mathcal{N}^{k+1}}(D^{(k)}P)}{1 - \delta d\sqrt{k+1}}.$$

Proof. We first prove Assertion (2). Toward this end, first observe that the Lipschitz constant of P on S^{n-1} is bounded from above by $\|D^{(1)}p\|_\infty$: This can be seen by taking $x, y \in S^{n-1}$ and considering the integral

$$P(x) - P(y) = \int_0^1 DP(y + t(x-y))(x-y)\, dt.$$

Since $\|y + t \cdot (x-y)\|_2 \leq 1$ for all $t \in [0,1]$, the homogeneity of the system P implies

$$\|DP(y + t(x-y))(x-y)\|_2 \leq \|D^{(1)}P\|_\infty \|x-y\|_2$$

Using the integral formula above we conclude

$$\|P(x) - P(y)\|_2 \leq \|D^{(1)}P\|_\infty \|x-y\|_2$$

Now, when the degrees of the p_i are identical, let the Lipschitz constant of P be M. By Assertion (2) of Theorem 2.3 we have

$$M \leq \|D^{(1)}P\|_\infty \leq d\|P\|_\infty$$

Now let $x_0 \in S^{n-1}$ be such that $\|P(x_0)\|_2 = \|P\|_\infty$ and let $y \in \mathcal{N}$ satisfy $|x_0 - y| \leq \delta$. Then $\|P\|_\infty = \|P(x_0)\|_2 \leq \|P(y)\|_2 + \|x_0 - y\|_2 M \leq \max_{\mathcal{N}}(P) + \delta d \|P\|_\infty$, and thus

$$\|P\|_\infty (1 - d\delta) \leq \max_{\mathcal{N}}(P).$$

To bound the norm of $D^{(k)}P(x)(u_1, \ldots, u_k)$ let us consider the net $\mathcal{N} \times \cdots \times \mathcal{N} = \mathcal{N}^{k+1}$ on $S^{n-1} \times \cdots \times S^{n-1}$. Let $x := (x_1, \ldots, x_{k+1}) \in S^{n-1} \times \cdots \times S^{n-1}$ and $y := (y_1, \ldots, y_{k+1}) \in \mathcal{N}^{k+1}$ be such that $\|x_i - y_i\|_2 \leq \varepsilon$ for all i. Clearly, $\|x - y\|_2 \leq \varepsilon\sqrt{k+1}$. Since x was arbitrary, this argument proves $\mathcal{N}^{k+1}$ is an $\varepsilon\sqrt{k+1}$-net. Note also that $D^{(k)}P(x)(u_1, \ldots, u_k)$ is a homogenous polynomial system with $(k+1)n$ variables and degree d. The desired bound then follows from the inequality obtained above.

To prove Assertion (1) of our current theorem, the preceding proof carries over verbatim, simply employing Assertion (1) (instead of Assertion (2)) from Theorem 2.3. ■

3. Condition Number of Random Polynomial Systems

3.1. **Introducing Randomness.** Now let $P := (p_1, \ldots, p_m)$ be a *random* polynomial system where $p_j(x) := \sum_{|\alpha|=d_j} c_{j,\alpha} \sqrt{\binom{d_j}{\alpha}} x^\alpha$. In particular, recall that $N_j = \binom{n+d_j-1}{d_j}$ and we let $C_j = (c_{j,\alpha})_{|\alpha|=d_j}$ be a random vector in $\mathbb{R}^{N_j}$ satisfying the Centering, Sub-Gaussian, and Small Ball assumptions from the introduction. Letting $\mathcal{X}_j := \left(\sqrt{\binom{d_j}{\alpha}} x^\alpha\right)_{|\alpha|=d_j}$ we then have $p_j(x) = \langle C_j, \mathcal{X}_j \rangle$. In particular, recall that the Sub-Gaussian assumption is that there is a $K > 0$ such that for each $\theta \in S^{N_j - 1}$ we have $\text{Prob}\left(|\langle C_j, \theta \rangle| \geq t\right) \leq 2e^{-t^2/K^2}$ for all $t > 0$. Recall also that the Small Ball assumption is that there is a $c_0 > 0$ such that for every vector $a \in \mathbb{R}^{N_i}$ we have $\text{Prob}\left(|\langle a, C_j \rangle| \leq \varepsilon\|a\|_2\right) \leq c_0\varepsilon$ for all $\varepsilon > 0$. In what follows, several of our bounds will depend on the parameters K and c_0 underlying the random variable being Sub-Gaussian and having the Small Ball property.

Let ξ be a random variable on $\mathbb{R}$. We denote the median of ξ by $\text{Med}(\xi)$. Now, if $\xi := |\langle C_j, \theta \rangle|$, then setting $t := 2K$ in the Sub-Gaussian assumption for C_j yields $\text{Prob}(\xi \geq 2K) \leq \frac{1}{2}$, i.e., $\text{Med}(\xi) \leq 2K$. On the other hand, setting $\varepsilon := \frac{1}{2c_0}$ in the Small Ball assumption for C_j yields $\text{Prob}(\xi \le \frac{1}{2c_0}) \le \frac{1}{2}$, i.e., $\text{Med}(\xi) \ge \frac{1}{2c_0}$. Writing $1 = \text{Med}(\xi) \cdot \frac{1}{\text{Med}(\xi)}$ we then easily obtain

$$Kc_0 \;\ge\; \frac{1}{4}. \tag{1}$$

In what follows we will use Inequality (1) several times.

3.2. **The Sub-Gaussian Assumption and Bounds Related to Operator Norms.** . We will need the following inequality, reminiscent of Hoeffding's classical inequality [15].

Theorem 3.1. [26, Prop. 5.10] *There is an absolute constant $c>0$ with the following property: If $X_1,\ldots,X_n$ are sub-Gaussian random variables with mean zero and underlying constant K, and $a=(a_1,\ldots,a_n)\in\mathbb{R}^n$ and $t\ge 0$, then*

$$\text{Prob}\left(\left|\sum_i a_iX_i\right| \ge t\right) \le 2\exp\left(\frac{-ct^2}{K^2\|a\|_2^2}\right). \qquad \blacksquare$$

Lemma 3.2. *Let $P := (p_1,\ldots,p_m)$ be a random polynomial system where, as before, $p_j(x) = \sum_{|\alpha|=d_j} c_{j,\alpha}\sqrt{\binom{d_j}{\alpha}}x^\alpha$ and the the coefficient vectors C_j are independent random vectors satisfying the Centering, Sub-Gaussian, and Small Ball assumptions from the introduction, with underlying constants K and c_0. Then, for $\mathcal{N}$ a δ-net over S^{n-1} and $t\ge 2$, we have the following inequalities:*

1. *If $\deg(p_j)=d$ for all $j\in\{1,\ldots,m\}$ then*
$$\text{Prob}\left(\|P\|_\infty \le \frac{2tK\sqrt{m}}{1-d\delta}\right) \ge 1-2|\mathcal{N}|e^{-O(t^2m)}$$
In particular, there is a constant $c_1\ge 1$ such that for $\delta=\frac{1}{3d}$ and $t=s\log(ed)$ with $s\ge 1$ we have $\text{Prob}\left(\|P\|_\infty \le 3sK\sqrt{m}\log(ed)\right)\ge 1-e^{-c_1s^2m\log(ed)}$.
2. *If $d:=\max_j \deg p_j$ then*
$$\text{Prob}\left(\|P\|_\infty \le \frac{2tK\sqrt{m}}{1-d^2\delta}\right) \ge 1-2|\mathcal{N}|e^{-O(tm)}$$
In particular, there is a constant $c_2\ge 1$ such that for $\delta=\frac{1}{3d^2}$, $t=s\log(ed)$ with $s\ge 1$, we have $\text{Prob}\left(\|P\|_\infty \le 3sK\sqrt{m}\log(ed)\right)\ge 1-e^{-c_2s^2m\log(ed)}$.

Proof. We prove Assertion (2) since the proofs of the two assertions are virtually identical. First observe that the identity $(x_1^2+\cdots+x_n^2)^d=\sum_{|\alpha|=d}\binom{d}{\alpha}x^{2\alpha}$ implies $\|\mathcal{X}_j\|_2=1$ for all $j\le m$. Using our sub-Gaussian assumption on the random vectors C_j, and the fact that $p_j(x)=\langle C_j,\mathcal{X}_j\rangle$, we obtain that $\text{Prob}\left(|p_j(x)|\ge t\right)\le 2e^{-t^2/K}$ for every $x\in S^{n-1}$.

Now we need to tensorize the preceding inequality. By Theorem 3.1, we have for all $a\in S^{m-1}$ that $\text{Prob}\left(|\langle a,P(x)\rangle|\ge t\right)\le 2e^{-ct^2/K^2}$. Letting $\mathcal{M}$ be a δ-net on S^{m-1} we then have $\text{Prob}\left(\max_{a\in\mathcal{M}}|\langle a,P(x)\rangle|\ge t\right)\le 2|\mathcal{M}|e^{-ct^2/K^2}$, where we have used the classical union bound for the multiple events defined by the (finite) δ-net $\mathcal{M}$. Since $\|P(x)\|_2=\max_{\theta\in S^{m-1}}|\langle\theta,P(x)\rangle|$, an application of Lemma 2.4 for the linear functional $\langle\,.\,,P(x)\rangle$ gives us $\text{Prob}\left(\|P(x)\|_2\ge\frac{t\sqrt{m}K}{1-\delta}\right)\le 2|\mathcal{M}|e^{-ct^2m}$.

It is known that for any $\delta>0$, S^{m-1} admits a δ-net $\mathcal{M}$ such that $|\mathcal{M}|\le\left(\frac{3}{\delta}\right)^m$ (see, e.g, [26, Lemma 5.2]). So for $t\ge 1$ and $\delta=\frac{1}{2}$ we have $\text{Prob}\left(\|P(x)\|_2\ge 2t\sqrt{m}K\right)\le 2e^{-c_2t^2m}$ for some suitable constant $c_2 \geq c$. We have thus arrived at a point-wise estimate on $\|P(x)\|_2$. Doing a union bound on a δ-net $\mathcal{N}$ now on S^{n-1} we then obtain:

$$\mathrm{Prob}\left(\max_{x\in\mathcal{N}}\|P(x)\|_2 \geq 2t\sqrt{m}K\right) \leq 2|\mathcal{N}|e^{-c_1 t^2 m}.$$

Using Lemma 2.4 once again completes our proof. ■

Theorem 2.3 and Lemma 3.2 then directly imply the following:

Corollary 3.3. *Let P be a random polynomial system as in Lemma 3.2. Then there are constants $c_1, c_2 \geq 1$ such that the following inequalities hold for $s \geq 1$:*

1. *If $\deg(p_j) = d$ for all $j \in \{1, \dots, m\}$ then both* $\mathrm{Prob}\left(\|D^{(1)}P\|_\infty \leq 3sK\sqrt{m}d\log(ed)\right)$ *and* $\mathrm{Prob}\left(\|D^{(2)}P\|_\infty \leq 3sK\sqrt{m}d^2\log(ed)\right)$ *are bounded from below by $1-2e^{-c_1 s^2 m\log(ed)}$.*
2. *If $d := \max_j \deg p_j$ then both* $\mathrm{Prob}\left(\|D^{(1)}P\|_\infty \leq 3sK\sqrt{m}d^2\log(ed)\right)$ *and* $\mathrm{Prob}\left(\|D^{(2)}P\|_\infty \leq 3sK\sqrt{m}d^4\log(ed)\right)$ *are bounded from below by $1-2e^{-c_2 s^2 m\log(ed)}$.* ■

3.3. **The Small Ball Assumption and Bounds for $L(P,x)$.** We will need the following standard lemma (see, e.g., [21, Lemma 2.2] or [27]).

Lemma 3.4. *Let $\xi_1, \dots, \xi_m$ be independent random variables such that, for every $\varepsilon > 0$, we have* $\mathrm{Prob}(|\xi_i| \leq \varepsilon) \leq c_0\varepsilon$. *Then there is a constant $\tilde{c} > 0$ such that for every $\varepsilon > 0$ we have* $\mathrm{Prob}\left(\sqrt{\xi_1^2 + \cdots + \xi_m^2} \leq \varepsilon\sqrt{m}\right) \leq (\tilde{c}c_0\varepsilon)^m$. ■

We can then derive the following result:

Lemma 3.5. *Let $P = (p_1, \dots, p_m)$ be a random polynomial system, satisfying the Small Ball assumption with underlying constant c_0. Then there is a constant $\tilde{c} > 0$ such that for every $\varepsilon > 0$ and $x \in S^{n-1}$ we have* $\mathrm{Prob}(\|P(x)\|_2 \leq \varepsilon\sqrt{m}) \leq (\tilde{c}c_0\varepsilon)^m$.

Proof. By the Small Ball assumption on the random vectors C_i, and observing that $p_i(x) = \langle C_i, \mathcal{X}_i\rangle$ and $\|\mathcal{X}_i\|_2 = 1$ for all $x \in S^{n-1}$, we have $\mathrm{Prob}(|p_i(x)| \leq \varepsilon) \leq c_0\varepsilon$. By Lemma 3.4 we are done. ■

The next lemma is a variant of [20, Claim 2.4].

Lemma 3.6. *Let $n \geq 3$, let $P := (p_1, \dots, p_m)$ be a system of n-variate homogenous polynomials, and assume $\|P\|_\infty \leq \gamma$ and $d := \max_i d_i$. Let $x, y \in S^{n-1}$ be such that $x \perp y$ and $\mathcal{L}(x,y) \leq \alpha$, and let $r \in [-1,1]$. Then for every w with $w = x + \beta r y + \beta z$ for some $z \in S^{n-1}$ with $z \perp x$ and $z \perp y$, we have the following inequalities:*

1. *If $0 \leq \beta \leq \frac{1}{d^2}$ then $\|P(w)\|_2^2 \leq 5(\alpha^2 + 19\beta^2\gamma^2 d^4)$.*
2. *If $\deg(p_i) = d$ for all $i \in \{1, \dots, m\}$ and $\beta \leq \frac{1}{d}$ then $\|P(w)\|_2^2 \leq 5(\alpha^2 + 19\beta^2\gamma^2 d^2)$.*

Proof. We will prove just Assertion (2) since the proof of Assertion (1) is almost the same. We start with some auxiliary observations on $\|P\|_\infty$: First note that Theorem 2.3 tells us that $\|P\|_\infty \leq \gamma$ implies $\|D^{(1)}P\|_\infty \leq d\gamma$ and, similarly, $\|D^{(k)}P\|_\infty \leq d^k\gamma$ for every $k \geq 1$. Also, for any w and $u_i \in S^{n-1}$ with $i \in \{1, \dots, k\}$, $\|P\|_\infty \leq \gamma$ and the homogeneity of the p_i implies $\sup_{u_1,\dots,u_k}\|D^{(k)}P(w)(u_1, \dots, u_k)\|_2 \leq \|w\|_2^{d-k}d^k\gamma$. These observations then yield the following inequalities for $w = x + \beta r y + \beta z$ with $z \in S^{n-1}$, $|r| \leq 1$, $\beta \leq d^{-1}$, $k = 3$, and $u_1, u_2, u_3 \in S^{n-1}$:

$$\|D^{(3)}P(w)(u_1, u_2, u_3)\|_2 \leq \|w\|_2^{d-3}d^3\gamma \leq \left(1 + \tfrac{2}{d}\right)^{d-3} d^3\gamma \leq ed^3\gamma$$

and
$p_j(w) = p_j(x) + \langle \nabla p_j(x), \beta ry + \beta z\rangle + \frac{1}{2}(\beta ry + \beta z)^T D^{(2)} p_j(x)(\beta ry + \beta z) + \frac{1}{6}(2\beta)^3 A_j(x)$,
where $A_j(x) := \int_0^1 D^{(3)} p_j(x + t\|v\|_2 v)(v, v, v)dt$.

Setting $v := \frac{\beta ry+\beta z}{\|\beta ry+\beta z\|_2}$ we then have

$$|p_j(w)| \leq |p_j(x)| + \beta|\langle \nabla p_j(x), y\rangle| + \beta|\langle \nabla p_j(x), z\rangle| + \frac{1}{2}\|\beta ry+\beta z\|_2^2 |D^{(2)} p_j(x)(v, v)| + \left|\frac{4}{3}\beta^3 A_j(x)\right|.$$

Note that $\|\beta ry + \beta z\|_2 \leq 2\beta$. Applying the Cauchy-Schwarz Inequality to the vectors $(1, \beta d_j^{\frac{1}{2}}, 1, 1, 1)$ and $(|p_j(x)|, d_j^{-\frac{1}{2}}|\langle \nabla p_j(x), y\rangle|, \ldots, \left|\frac{4}{3}\beta^3 A_j(x)\right|)$ then implies the following inequality:
$p_j(w)^2 \leq (4 + \beta^2 d_j)(p_j(x)^2 + d_j^{-1}\langle p_j(x), y\rangle^2 + \beta^2\langle \nabla p_j(x), z\rangle^2 + \frac{1}{4}(2\beta)^4 (D_j^{(2)} p_j(x)(v, v))^2 + \frac{1}{36}(2\beta)^6 A_j(x)^2)$.

Summing over all $j \in \{1, \ldots, m\}$, using the assumption $\|P\|_\infty \leq \gamma$ and our earlier observations, we have:
$\|P(w)\|_2^2 \leq (4 + \beta^2 d)(\|P(x)\|_2^2 + \|M^{-1} D^{(1)} P(x)(y)\|_2^2 + \beta^2 d^2\gamma^2 + 4\beta^4 d^4\gamma^2 + \frac{64}{36}\beta^6 \sum_j A_j(x)^2)$.

Clearly $\sum_{j\leq m} A_j(x)^2 \leq \max_{w\in V_{x,y}} \|D^{(3)} P(w)(u_1, u_2, u_3)\|_2^2 \leq e^2 d^6\gamma^2$. Hence we have:

$$\|P(w)\|_2^2 \leq (4 + \beta^2 d)(\alpha^2 + \beta^2 d^2\gamma^2 + 4\beta^4 d^4\gamma^2 + \frac{64e^2}{36}\beta^6 d^6\gamma^2).$$

Since $\beta \leq d^{-1}$ and $\frac{64e^2}{36} \leq 14$, we then have

$$\|P(w)\|_2^2 \leq (4 + \beta^2 d)(\alpha^2 + 19\beta^2 d^4\gamma^2) \leq 5(\alpha^2 + 19\beta^2 d^4\gamma^2).$$

■

Theorem 3.7. *Let $n \geq 3$ and let $P := (p_1, \ldots, p_m)$ be a system of random homogenous n-variate polynomials such that $p_j(x) = \sum_{|a|=d_j} c_{j,a}\sqrt{\binom{d_i}{a}} x^a$ where $C_j = (c_{j,a})_{|a|=d_j}$ are random vectors satisfying the Small Ball assumption with underlying constant c_0. Let $\alpha, \gamma > 0$. Then we have the following inequalities:*

1. *If $d := \max_j \deg(p_j)$ and $\alpha \leq \gamma \min\{1, d/\sqrt{n}\}$ then*
$$\text{Prob}\,(L(P) \leq \alpha) \;\leq\; \text{Prob}\,(\|P\|_\infty \geq \gamma) + \alpha\sqrt{\frac{n}{m}}(d^2\gamma)^{n-2}\alpha^{m-n+1} O\left(\frac{c_0}{\sqrt{m}}\right)^{m-1}.$$
2. *If $\deg(p_j) = d$ for all $j \in \{1, \ldots, m\}$ and $\alpha \leq \gamma \min\{1, d/\sqrt{n}\}$ then*
$$\text{Prob}\,(L(P) \leq \alpha) \;\leq\; \text{Prob}\,(\|P\|_\infty \geq \gamma) + \alpha\sqrt{\frac{n}{m}}(d\gamma)^{n-2}\alpha^{m-n+1} O\left(\frac{c_0}{\sqrt{m}}\right)^{m-1}.$$

Proof. We first assume the hypotheses of Assertion (1): Let $\alpha, \gamma > 0$ and $\beta \leq \frac{1}{d^2}$. Let $\mathbf{B} := \{P \mid \|P\|_\infty \leq \gamma\}$ and let $\mathbf{L} := \{P \mid L(P) \leq \alpha\} := \{\exists x \perp y : \mathcal{L}(x, y) \leq \alpha\}$. Let $\Gamma := 5(\alpha^2 + 19\beta^2 d^4\gamma^2)$ and let B_2^n denote the ℓ_2 unit ball in $\mathbb{R}^n$. Lemma 3.6 implies that, if the event $\mathbf{B} \cap \mathbf{L}$ occurs, then there exists a set

$$V_{x,y} := \{w \in \mathbb{R}^n : w = x + \beta ry + \beta z, z \in S^{n-1}, |r| \leq 1, x \perp z, y \perp z\} \setminus B_2^n$$

such that $\|P(w)\|_2^2 \leq \Gamma$ for every w in this set. Let $V := |V_{x,y}|$. Note that for $w \in V_{x,y}$, $1 \leq \|w\|_2 \leq 1 + 2\beta^2$. Since $V_{x,y} \subseteq (1 + 2\beta^2)B_2^n \setminus B_2^n$, we have showed that the event $\mathbf{B} \cap \mathbf{L}$ is included in the event $\{P \mid \text{Vol}(\{x \in (1 + 2\beta^2)B_2^n \setminus B_2^n \mid \|P(x)\|_2 \leq \Gamma\}) \geq V\}$. Using Markov's Inequality, Fubini's Theorem, and Lemma 3.5, we can estimate the probability of

this event. Indeed,
$\text{Prob}\left(\left|\{x \in (1+2\beta^2)B_2^n \setminus B_2^n : \|P(x)\|_2 \le \Gamma\}\right| \ge V\right)$

$$\begin{aligned}
&\le \frac{1}{V}\mathbb{E}\left|\{x \in (1+2\beta^2)B_2^n \setminus B_2^n : \|P(x)\|_2^2 \le \Gamma\}\right| \\
&\le \frac{1}{V}\int_{(1+2\beta^2)B_2^n\setminus B_2^n} \text{Prob}\left(\|P(x)\|_2^2 \le \Gamma\right) dx \\
&\le \frac{|(1+2\beta^2)B_2^n \setminus B_2^n|}{V} \max_{x\in(1+2\beta^2)B_2^n\setminus B_2^n} \text{Prob}\left(\|P(x)\|_2^2 \le \Gamma\right).
\end{aligned}$$

Now recall that $\text{Vol}(B_2^n) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}$. Then $\frac{\text{Vol}(B_2^n)}{\text{Vol}(B_2^{n-1})} \le \frac{c'}{\sqrt{n}}$ for some constant $c' > 0$. We also assume that $\beta^2 \le \frac{1}{n}$ which yields $(1+2\beta^2)^n \le e^2$, and we compute that

$$\frac{\text{Vol}((1+2\beta^2)B_2^n \setminus B_2^n)}{V} \le \frac{\text{Vol}(B_2^n)\left((1+2\beta^2)^n - 1\right)}{\beta\beta^{n-1}\text{Vol}(B_2^{n-1})} \le c\sqrt{n}\beta^{2-n},$$

for some absolute constant $c > 0$. Write $\tilde{x} := \frac{x}{\|x\|_2}$ for any $x \ne 0$. Then for $z \notin B_2^n$ we have that $\|P(z)\|_2^2 = \sum_{j=1}^m |p_j(z)|^2 = \sum_{j=1}^m |p_j(\tilde{z})|^2\|z\|_2^{2d_j} \ge \sum_{j=1}^m |p_j(\tilde{z})|^2 = \|P(\tilde{z})\|_2^2$, which implies that for every $w \in (1+2\beta^2)B_2^n \setminus B_2^n$ we have

$$\text{Prob}\left(\|P(w)\|_2^2 \le \Gamma\right) \le \text{Prob}\left(\|P(\tilde{w})\|_2^2 \le \Gamma\right) \le \left(cc_0\sqrt{\frac{\Gamma}{m}}\right)^m,$$

where we have used Lemma 3.5. So we conclude that
$\text{Prob}(L(P) \le \alpha) \le \text{Prob}\left(\|P\|_\infty \ge \gamma\right) + \text{Prob}\left(\mathbf{B} \cap \mathbf{L}\right) \le \text{Prob}\left(\|P\|_\infty \ge \gamma\right) + c\sqrt{n}\beta^{2-n}\left(cc_0\sqrt{\frac{\Gamma}{m}}\right)^m.$
Recall that $\Gamma = 5(\alpha^2 + 19\beta^2 d^4\gamma^2)$. We choose $\beta := \frac{\alpha}{\gamma d^2}$ and observe that, under our assumption that $\alpha \le \gamma$, our choice of β implies that $\Gamma = 100\alpha^2$. So we obtain

$$\text{Prob}(L(P) \le \alpha) \le \text{Prob}\left(\|P\|_\infty \ge \gamma\right) + c\sqrt{n}\left(\frac{\alpha}{\gamma d^2}\right)^{2-n}\left(\frac{10cc_0\alpha}{\sqrt{m}}\right)^m$$

and thus we have proved Assertion (1).

Applying our preceding argument under the assumptions of Assertion (2), we obtain $\Gamma := 5(\alpha^2 + 19\beta^2 d^2\gamma^2)$. In this case, choosing $\beta = \frac{\alpha}{\gamma d}$ implies

$$\text{Prob}(L(P) \le \alpha) \le \text{Prob}\left(\|P\|_\infty \ge \gamma\right) + c\sqrt{n}\left(\frac{\alpha}{\gamma d}\right)^{2-n}\left(\frac{10cc_0\alpha}{\sqrt{m}}\right)^m.$$

By adjusting constants we are done. ■

3.4. **The Condition Number Theorem and its Consequences.** We will now need bounds for the Weil-Bombieri norm of polynomial systems. Note that, with

$$p_j(x) = \sum_{\alpha_1+\cdots+\alpha_n=d_j} \sqrt{\binom{d_j}{\alpha}}c_{j,\alpha}x^\alpha,$$

we have $\|p_j\|_W := \|(c_{j,\alpha})_\alpha\|_2$ for $j \in \{1,\ldots,m\}$. The following lemma providing large deviation estimates for the Euclidean norm is standard and follows, for instance, from Theorem 3.1.

Lemma 3.8. *There is a universal constant $c'>0$ such that for any a random n-variate polynomial system P satisfying the Centering and sub-Gaussian assumptions with underlying constant K, $j\in\{1,\ldots,m\}$, $N_j:=\binom{n+d_j-1}{d_j}$, $N:=\sum_{j=1}^m N_j$, and $t\geq 1$, we have:*

(1) $\mathrm{Prob}\left(\|p_j\|_W\geq c'tK\sqrt{N_j}\right)\leq e^{-t^2N_j}$.
(2) $\mathrm{Prob}\left(\|P\|_W\geq c'tK\sqrt{N}\right)\leq e^{-t^2N}$. ■

We are now ready to prove our main theorem on condition numbers of random polynomial systems.

Theorem 3.9. *There are universal constants $A,c>0$ such that the following hold: Let $P=(p_1,\ldots,p_m)$ be a system of homogenous random polynomials with $p_j(x)=\sum_{|\alpha|=d_j}c_{j,\alpha}\sqrt{\binom{d_j}{\alpha}}x^\alpha$ and let each $C_j=(c_{j,\alpha})_{|\alpha|=d_j}$ be an independent random vector satisfying the Sub-Gaussian and Small Ball assumptions, with respective underlying constants K and c_0. Then:*

(1) *If $d=\deg(p_j)$ for all $j\in\{1,\ldots,m\}$ then, setting*
$$M:=cc_0A\left(\frac{n}{c_0}\right)^{\frac{1}{2(m-n+2)}}\sqrt{\frac{N}{m}}K\left(3c_0AKd\log(ed)\right)^{\frac{n-2}{m-n+2}}\max\left\{1,\frac{\sqrt{n}}{d}\right\},$$
we have that $\mathrm{Prob}(\tilde{\kappa}(P)\geq tM)$ is bounded from above by
$$\begin{cases}\frac{3}{t^{m-n+2}} & \text{if } 0\leq t\leq e^{\frac{m\log(ed)}{m-n+2}}\\ \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{m\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^{\frac{m\log(ed)}{m-n+2}}\leq t\leq e^{\frac{N}{m-n+2}}\\ \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{N}\right)^{\frac{m}{2}}\left(\frac{N}{m\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^{\frac{N}{m-n+2}}\leq t\end{cases}$$

(2) *If $d=\max_j\deg p_j$ then, setting*
$$M:=cc_0A\left(\frac{n}{c_0}\right)^{\frac{1}{2(m-n+2)}}\sqrt{\frac{N}{m}}K\left(3c_0AKd^2\log(ed)\right)^{\frac{n-2}{m-n+2}}\max\left\{1,\frac{\sqrt{n}}{d^2}\right\},$$
we have two cases:

(a) *If $N\geq m\log(ed)$ then $\mathrm{Prob}(\tilde{\kappa}(P)\geq tM)$ is bounded from above by*
$$\begin{cases}\frac{3}{t^{m-n+2}} & \text{if } 1\leq t\leq e^{\frac{m\log(ed)}{m-n+2}}\\ \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{m\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^{\frac{m\log(ed)}{m-n+2}}\leq t\leq e^{\frac{N}{m-n+2}}\\ \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{N}\right)^{\frac{m}{2}}\left(\frac{N}{m\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^{\frac{N}{m-n+2}}\leq t\end{cases}$$

(b) *If $N\leq m\log(ed)$ then $\mathrm{Prob}(\tilde{\kappa}(P)\geq tM)$ is bounded from above by*
$$\begin{cases}\frac{3}{t^{m-n+2}} & \text{if } 1\leq t\leq e^{\frac{N}{m-n+2}}\\ \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{N}\right)^{\frac{m}{2}} & \text{if } e^{\frac{N}{m-n+2}}\leq t\end{cases}$$

Proof. We first consider Assertion (2). Recall that $\tilde{\kappa}(P)=\frac{\|P\|_W}{L(P)}$. Note that if $u>0$, and the inequalities $\|P\|_W\geq ucK\sqrt{N}$ and $L(P)\leq\frac{ucK\sqrt{N}}{tM}$ hold, then we clearly have $\tilde{\kappa}(P)\geq tM$. In particular, $u>0$ implies that
$$\mathrm{Prob}\left(\tilde{\kappa}(P)\geq tM\right)\leq\mathrm{Prob}\left(\|P\|_W\geq ucK\sqrt{N}\right)+\mathrm{Prob}\left(L(P)\leq\frac{ucK\sqrt{N}}{tM}\right).$$

To estimate the last two probabilities, we apply Theorem 3.7 with $\alpha := \frac{ucK\sqrt{N}}{tM}$ and $\gamma := 3sK\sqrt{m}\log(ed)$, Lemma 3.2, and Lemma 3.8. Now let us note our restrictions: We have that $s \geq 1$, $u \geq 1$, and (since $\alpha \leq \min\left\{1, \frac{d^2}{\sqrt{n}}\right\}\gamma$) we have

$$(*) \quad \left(\frac{c_0 m}{n}\right)^{\frac{1}{2(m-n+2)}} \frac{u\sqrt{m^{\frac{m-n+1}{m-n+2}}}}{\max\{1, \frac{\sqrt{n}}{d^2}\} c_0 At(c_0 AKd^2\log(ed))^{\frac{n-2}{m-n+2}}} \leq 3Ks\sqrt{m}\log(ed)\min\left\{1, \frac{d^2}{\sqrt{n}}\right\}.$$

Note that $(*)$ is true if $u \leq s$ and $t \geq 1$. Under the above restrictions we have that if $Q := \mathrm{Prob}(\tilde{\kappa}(p) \geq tM)$ then

$$Q \leq A\sqrt{\frac{n}{m}}\frac{ucK\sqrt{N}}{tM}\left(\frac{3c_0AKd^2s\sqrt{m}\log(ed)}{\sqrt{m}}\right)^{n-2}\left(\frac{c_0cuAK\sqrt{N}}{tM\sqrt{m}}\right)^{m-n+1} + e^{-c_2s^2m\log(ed)} + e^{-u^2N}$$

or,

$$Q \leq \frac{u^{m-n+2}s^{n-2}}{t^{m-n+2}} + e^{-c_2s^2m\log(ed)} + e^{-u^2N}$$

for some suitable $c_2 > 0$. We consider first the case where $N \geq m\log(ed)$. If $1 \leq t \leq e^{\frac{c_1 m\log(ed)}{m-n+2}}$ then we take $u = s = 1$ (note that $(*)$ is satisfied and that $c_2 \geq 1$) and we then obtain $Q \leq \frac{1}{t^{m-n+2}} + e^{-c_2m\log(ed)} + e^{-N} \leq \frac{3}{t^{m-n+2}}$ for $t \geq 1$ and $u \geq s \geq 1$. In the case where $e^{\frac{m\log(ed)}{m-n+2}} \leq t \leq e^{\frac{N}{m-n+2}}$ we choose $u = 1$ and $s := \sqrt{\frac{(m-n+2)\log t}{m\log(ed)}} \geq 1$. (Note that $u \leq s$). These choices then yield

$$Q \leq \frac{1}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{m\log(ed)}\right)^{\frac{n-2}{2}} + \frac{1}{t^{c_2(m-n+2)}} + e^{-N} \leq \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{m\log(ed)}\right)^{\frac{n-2}{2}}.$$

In the case where $e^{\frac{N}{m-n+2}} \leq t$, we choose $s := \sqrt{\frac{(\log t)(m-n+2)}{m\log(ed)}}$ and $u := \sqrt{\frac{(m-n+2)\log t}{N}}$. (Note that $u \leq s$ also in this case). In this case we get that

$$\begin{aligned} Q &\leq \frac{1}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{N}\right)^{\frac{m}{2}}\left(\frac{N}{m\log(ed)}\right)^{\frac{n-2}{2}} + \frac{1}{t^{c_2(m-n+2)}} + \frac{1}{t^{m-n+2}} \\ &\leq \frac{3}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{N}\right)^{\frac{m}{2}}\left(\frac{N}{m\log(ed)}\right)^{\frac{n-2}{2}}. \end{aligned}$$

We consider now the case where $N \leq m\log(ed)$. In the case $1 \leq t \leq e^{\frac{N}{m-n+2}}$ we choose $s = 1$ and $u = 1$ and we get

$$Q \leq \frac{1}{t^{m-n+2}} + e^{-c_2m\log(ed)} + e^{-N} \leq \frac{3}{t^{m-n+2}}$$

as before. In the case $t \geq e^{\frac{N}{m-n+2}}$, we choose $s = u := \sqrt{\frac{(m-n+2)\log t}{m\log(ed)}}$. Note that again $(*)$ is satisfied and with these choices we get

$$\begin{aligned} Q &\leq \frac{1}{t^{m-n+2}}\left(\frac{(m-n+2)\log t}{N}\right)^{\frac{m}{2}} + \frac{1}{t^{c_2(m-n+2)m\log(ed)/N}} + \frac{1}{t^{m-n+2}} \\ &\leq \frac{3}{t^{m-n+2}}\left(\frac{(\log t)(m-n+2)}{N}\right)^{\frac{m}{2}}. \end{aligned}$$

When $d_j = d$ for all $j \in \{1, \ldots, m\}$ the proof is similar: One has just to observe that in this case we have $N \geq m\log(ed)$. ■

Theorem 3.10. *Let P be a random polynomial system as in Theorem 3.9 and let M be as defined in Theorem 3.9. Set*

$$\delta_1 := \frac{q\sqrt{\pi n}}{m-n+2}\left(\frac{n-2}{2em\log(ed)}\right)^{\frac{n}{2}-1}\frac{1}{\left(1-\frac{q}{m-n+2}\right)^{\frac{n}{2}}} \qquad \textit{and}$$

$$\delta_2 := \left(\frac{m}{N}\right)^{\frac{m-n+2}{2}}\frac{1}{\left(1-\frac{q}{m-n+2}\right)^{\frac{m}{2}}}\frac{\sqrt{\pi m}q}{m-n+2-q}e^{-\frac{m}{2}}\frac{1}{(\log(ed))^{\frac{n}{2}-1}}.$$

Then we have the following estimates:

(1) *If $d=d_j$ for all $j\in\{1,\ldots,m\}$ and $q\in(0,m-n+2)$ then we have*

$$(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le M\left(1+\frac{q}{m-n-q+2}+\delta_1+\delta_2\right)^{\frac{1}{q}}.$$

In particular, if $q\in\left(0,(m-n+2)\left(1-\frac{1}{2\log(ed)}\right)\right]$, then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le M\left(\frac{3m\log(ed)}{n}\right)^{\frac{1}{q}}$, and if $q\le\frac{m-n+2}{2}$ then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le 4^{1/q}M$. Furthermore, $\mathbb{E}(\log\tilde\kappa(P))\le 1+\log M$.

(2) *If $\max_j\deg(p_j)=d$ then consider the following two cases:*

(a) *If $N\ge m\log(ed)$ and $q\in(0,m-n+2)$ then*

$$(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le M\left(1+\frac{q}{m-n-q+2}+\delta_1+\delta_2\right)^{\frac{1}{q}}.$$

In particular, if $q\in\left(0,(m-n+2)\left(1-\frac{1}{2\log(ed)}\right)\right]$, then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le M\left(\frac{3m\log(ed)}{n}\right)^{\frac{1}{q}}$, and if $q\le\frac{m-n+2}{2}$ then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le 4^{1/q}M$. Furthermore, $\mathbb{E}(\log\tilde\kappa(P))\le 1+\log M$.

(b) *If $N\le m\log(ed)$ then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le M\left(1+\frac{q}{m-n-q+2}+\delta_2\right)^{\frac{1}{q}}$. In particular, if $q\le(m-n+2)\left(1-\frac{m}{eN}\right)$, then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le M\left(\frac{3m\log(ed)}{n}\right)^{\frac{1}{q}}$, and if $q\le\frac{m-n+2}{2}$ then $(\mathbb{E}(\tilde\kappa(P)^q))^{\frac{1}{q}} \le 4^{1/q}M$. Furthermore, $\mathbb{E}(\log\tilde\kappa(P))\le 1+\log M$.*

Proof. We first consider the case where $d=d_j$ for all $j\in\{1,\ldots,m\}$. Set

$$\Delta_1 := \left(\frac{m-n+2}{m\log ed}\right)^{\frac{n}{2}-1},\ \Delta_2 := \left(\frac{m-n+2}{N}\right)^{\frac{m}{2}}\left(\frac{N}{m\log ed}\right)^{\frac{n}{2}-1},$$

$$r := m-n-q+3,\ a_1 := \frac{m\log ed}{m-n+2},\ \text{and } a_2 := \frac{N}{m-n+2}.$$

Note that we have assumed that $r\ge 1$. Using the formula

$$\mathbb{E}((\tilde\kappa(P))^q) = q\int_0^\infty t^{q-1}\mathrm{Prob}\left(\tilde\kappa(p)\ge t\right)dt$$

(which follows easily from the definition of expectation), and Theorem 3.9, we have that

$$\mathbb{E}((\tilde\kappa(P))^q) \le M^q\left(1+q\int_1^\infty t^{q-1}\mathrm{Prob}\left(\tilde\kappa(p)\ge tM\right)dt\right)$$

or $\frac{\mathbb{E}((\tilde{\kappa}(P))^q)}{M^q} \leq 1 + q\int_1^{e^{a_1}} \frac{1}{t^r}dt + q\Delta_1 \int_{e^{a_1}}^{e^{a_2}} \frac{(\log t)^{\frac{n}{2}-1}}{t^r}dt + q\Delta_2 \int_{e^{a_2}}^{\infty} \frac{(\log t)^{\frac{m}{2}}}{t^r}dt.$ We will give upper bounds for the last three integrals. First note that

$$q\int_1^{e^{a_1}} \frac{1}{t^r}dt = \frac{q}{r-1}\left(1 - e^{(r-1)a_1}\right) \leq \frac{q}{r-1}.$$

Also, we have that

$$q\Delta_1 \int_{e^{a_1}}^{e^{a_2}} \frac{(\log t)^{\frac{n}{2}-1}}{t^r}dt = q\Delta_1 \int_{a_1}^{a_2} t^{\frac{n}{2}-1}e^{(r-1)t}dt = \frac{q\Delta_1}{(r-1)^{\frac{n}{2}}} \int_{a_1(r-1)}^{a_2(r-1)} t^{\frac{n}{2}-1}e^{-t}dt$$

$$\leq \frac{q\Delta_1}{(r-1)^{\frac{n}{2}}}\Gamma\left(\frac{n}{2}\right) \leq \frac{q\sqrt{\pi n}}{m-n+2}\left(\frac{n-2}{2em\log(ed)}\right)^{\frac{n}{2}-1} \frac{1}{\left(1-\frac{q}{m-n+2}\right)^{\frac{n}{2}}}.$$

Finally, we check that

$$q\Delta_2 \int_{e^{a_2}}^{\infty} \frac{(\log t)^{\frac{m}{2}}}{t^r}dt = q\Delta_2 \int_{a_2}^{\infty} t^{\frac{m}{2}}e^{(r-1)t}dt = \frac{q\Delta_2}{(r-1)^{\frac{m}{2}+1}} \int_{a_2(r-1)}^{\infty} t^{\frac{m}{2}}e^{-t}dt$$

$$\leq \frac{q\Delta_2}{(r-1)^{\frac{m}{2}+1}}\Gamma\left(\frac{m}{2}+1\right) \leq \frac{\sqrt{\pi m}q}{(m-n-q+2)^{\frac{m}{2}+1}}\left(\frac{m(m-n+2)}{eN}\right)^{\frac{m}{2}}\left(\frac{N}{m\log ed}\right)^{\frac{n}{2}-1}$$

$$= \left(\frac{m}{N}\right)^{\frac{m-n+2}{2}} \frac{1}{\left(1-\frac{q}{m-n+2}\right)^{\frac{m}{2}}} \frac{\sqrt{\pi m}q}{m-n+2-q}e^{-\frac{m}{2}} \frac{1}{(\log(ed))^{\frac{n}{2}-1}}.$$

Note that if $q \leq (m-n+2)\left(1 - \frac{1}{2\log(ed)}\right)$ then $\delta_1, \delta_2 \leq 1$.

The proof for the case $\max_j d_j = d$ and $N \geq m\log(ed)$ is identical. For the case $\max_j d_j = d$ and $N \leq m\log(ed)$, working as before, we get that

$$\frac{\mathbb{E}((\tilde{\kappa}(P))^q)}{M^q} \leq 1 + q\int_1^{e^{a_2}} \frac{1}{t^r}dt + q\Delta_2 \int_{e^{a_2}}^{\infty} \frac{(\log t)^{\frac{m}{2}}}{t^r}dt \leq 1 + \frac{q}{r-1} + \delta_2.$$

In the case $N \leq m\log(ed)$ we have $\delta_2 \leq \frac{\sqrt{\pi m}q}{m-n+2}\left(\frac{m}{eN}\right)^{\frac{m}{2}} \frac{1}{\left(1-\frac{q}{m-n+2}\right)^{\frac{m}{2}+1}}$. In particular, for this case, it easily follows that $q \leq (m-n+2)\left(1 - \frac{m}{N}\right)$ implies $\delta_2 \leq 1$. ■

Note that if $m = n-1$ then $N \geq m\log(ed)$, and in then easily follows that (∗) still holds even if reduce M by deleting its factor of $\max\left\{1, \frac{\sqrt{n}}{d^2}\right\}$. So then, for the important case $m = n-1$, our main theorems immediately admit the following refined form:

Corollary 3.11. *There are universal constants $A, c > 0$ such that if P is any random polynomial system as in Theorem 3.9, but with $m = n-1$, then the following hold:*

(1) *If $d = \deg(p_i)$, $1 \leq i \leq m$ then, setting $M := cA\sqrt{2c_0N}K\left(3Ac_0Kd\log(ed)\right)^{n-2}$, we have*

$$\mathrm{Prob}(\tilde{\kappa}(P) \geq tM) \leq \begin{cases} \frac{3}{t} & \text{if } 1 \leq t \leq e^{(n-1)\log(ed)} \\ \frac{3}{t}\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^{(n-1)\log(ed)} \leq t \leq e^N \\ \frac{3}{t}\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-2}{2}}\left(\frac{\log t}{N}\right)^{\frac{1}{2}} & \text{if } e^N \leq t \end{cases}$$

and, for all $q \in \left(0, 1 - \frac{1}{2\log(ed)}\right]$, we have $\left(\mathbb{E}(\tilde{\kappa}(P)^q)\right)^{\frac{1}{q}} \leq Me^{\frac{1}{q}}$. Furthermore, $\mathbb{E}(\log\tilde{\kappa}(P)) \leq 1 + \log M$.

(2) *If* $\max_i \deg(p_i) = d$ *then, setting* $M := cA\sqrt{2c_0N}K\left(3Ac_0Kd^2\log(ed)\right)^{n-2}$, *we have*

$$\mathrm{Prob}(\tilde{\kappa}(P) \geq tM) \leq \begin{cases} \frac{3}{t} & \text{if } 1 \leq t \leq e^{(n-1)\log(ed)} \\ \frac{3}{t}\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^{(n-1)\log(ed)} \leq t \leq e^N \\ \frac{3}{t}\left(\frac{\log t}{N}\right)^{\frac{1}{2}}\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-2}{2}} & \text{if } e^N \leq t \end{cases},$$

and, for all $q \in \left(0, 1 - \frac{1}{2\log(ed)}\right]$, *we have* $(\mathbb{E}(\tilde{\kappa}(P)^q))^{\frac{1}{q}} \leq Me^{\frac{1}{q}}$.
Furthermore, $\mathbb{E}(\log\tilde{\kappa}(P)) \leq 1 + \log M$. ■

We are now ready to prove Corollary 1.3 from the introduction.

Proof of Corollary 1.3: From Corollary 3.11, Bound (2) follows immediately, and Bound (1) is clearly true for the smaller domain of t. So let us now consider $t = xe^{(n-1)\log(ed)}$ with $x \geq 1$. Clearly, $\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-1}{2}} = \left(1 + \frac{\log x}{(n-1)\log(ed)}\right)^{\frac{n-1}{2}}$, and thus $\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-1}{2}} < e^{\frac{\log x}{2\log(ed)}} = x^{\frac{1}{2\log(ed)}}$. Since $x = \frac{t}{e^{(n-1)\log(ed)}}$ we thus obtain

$$\frac{3}{t}\left(\frac{\log t}{(n-1)\log(ed)}\right)^{\frac{n-1}{2}} \leq \frac{3}{t}\left(\frac{t}{e^{(n-1)\log(ed)}}\right)^{\frac{1}{2\log(ed)}}.$$

Renormalizing the pair (M, t) (since the M from Corollary 3.11 is larger than the M from Corollary 1.3 by a factor of A), we are done. ■

3.5. **On the Optimality of Condition Number Estimates.** As mentioned in the introduction, to establish a lower bound we need one more assumption on the randomness. For the convenience of the reader, we recall the Euclidean small ball assumption.

(**Euclidean Small Ball**) *There is a constant* $\tilde{c}_0 > 0$ *such that for each* $j \in \{1, \ldots, m\}$ *and* $\varepsilon > 0$ *we have* $\mathrm{Prob}\left(\|C_j\|_2 \leq \varepsilon\sqrt{N_j}\right) \leq (\tilde{c}_0\varepsilon)^{N_j}$.

We will need an extension of Lemma 3.4: Lemma 3.12 below (see also [23, Thm. 1.5 & Cor. 8.6]). Toward this end, for any matrix $T := (t_{i,j})_{1 \leq i,j \leq m}$, write $\|T\|_{HS}$ for its *Hilbert-Schmidt norm* of T and $\|T\|_{op}$ for its *operator norm*, i.e.,

$$\|T\|_{HS} := \left(\sum_{i,j=1}^{m} t_{i,j}^2\right)^{\frac{1}{2}}, \ \|T\|_{op} := \max_{\theta \in S^{n-1}} \|T\theta\|_2.$$

Lemma 3.12. *Let* $\xi_1, \ldots, \xi_m$ *be independent random variables satisfying the following Small Ball assumption: For all* $i \in \{1, \ldots, m\}$ *and* $\varepsilon > 0$ *we have* $\mathrm{Prob}(\xi_i \leq \varepsilon) \leq c_0\varepsilon$. *Let* $\xi := (\xi_1, \ldots, \xi_m)$. *Then there is a constant* $c > 0$ *such that for any* $m \times m$ *matrix* T *and* $\varepsilon > 0$ *we have* $\mathrm{Prob}(\|T\xi\|_2 \leq \varepsilon\|T\|_{HS}) \leq (cc_0\varepsilon)^{c\frac{\|T\|_{HS}^2}{\|T\|_{op}^2}}$.

Our main lower bound for the condition number is then the following:

Lemma 3.13. *Let* $P = (p_1, \ldots, p_m)$ *be a homogeneous* n*-variate polynomial system with* $\deg(p_j) = d_j$ *for all* j. *Then* $\tilde{\kappa}(P) \geq \frac{\|P\|_W}{\|P\|_\infty\sqrt{m+1}}$. *Moreover if* $P := (p_1, \ldots, p_m)$ *is a random polynomial system satisfying our Sub-Gaussian and Tiny Ball assumptions, with respective* *underlying constants K and $\tilde{c}_0$, then we have*

$$\operatorname{Prob}\left(\tilde{\kappa}(P) \le \varepsilon \frac{\sqrt{N}}{Kmd\log(ed)}\right) \le (c\tilde{c_0}\varepsilon)^{c'\min\left\{N\frac{\min_j N_j}{\max_j N_j}, md\log(ed)\right\}} \quad \text{and}$$

$$\operatorname{Prob}\left(\tilde{\kappa}(P) \le \varepsilon \frac{\sqrt{N}}{Km\log(ed)}\right) \le (c\tilde{c_0}\varepsilon)^{c'm\log(ed)}, \text{ if } d = d_j \text{ for all } j \in \{1,\ldots,m\},$$

where $c, c' > 0$ are absolute constants. In particular we have $e^{\mathbb{E}(\log \tilde{\kappa}(P))} \ge c\frac{\sqrt{N}}{m\log(ed)}$ when $d = d_j$ for all $j \in \{1,\ldots,m\}$.

Proof. First note that Theorem 2.2 implies that for every $x, y \in S^{n-1}$ we have

$$\|d_j^{-1}D^{(1)}p_j(x)y\|_2^2 \le \|p_j\|_\infty^2.$$

So we have $\|M^{-1}D^{(1)}P(x)(y)\|_2^2 \le \sum_{j=1}^m \|p_j\|_\infty^2 \le m\|P\|_\infty^2$. Now recall that

$$\mathcal{L}^2(x,y) := \|M^{-1}D^{(1)}P(x)(y)\|_2^2 + \|p(x)\|_2^2.$$

So we get $L^2(P) := \min_{x\perp y} \mathcal{L}^2(x,y) \le (m+1)\|P\|_\infty^2$, which in turn implies that

$$\tilde{\kappa}(P) \ge \frac{\|P\|_W}{L(P)} \ge \frac{\|P\|_W}{\|P\|_\infty\sqrt{m+1}}.$$

The proof for the case where $d_j = d$ for all $j \in \{1,\ldots,m\}$ is identical.

We now show that under our Tiny Ball Assumption the following holds: For every $\varepsilon > 0$

$$\operatorname{Prob}\left(\|P\|_W \le \varepsilon\sqrt{N}\right) \le (c\tilde{c_0}\varepsilon)^{cN\frac{\min_j N_j}{\max_j N_j}}.$$

Indeed, recall that $\|p_j\|_W = \|C_j\|_{\ell_2^{N_j}}$. Then for any fixed $\varepsilon > 0$,

$$\operatorname{Prob}\left(\|p_j\|_W \le \varepsilon\sqrt{N_j}\right) \le (\tilde{c_0}\varepsilon)^{N_j} \le (\tilde{c_0}\varepsilon)^{N_{j_0}}.$$

where $j_0 \in \{1,\ldots,m\}$ is such that $N_{j_0} := \min_j N_j$. Let $\xi_j := \frac{\|p_j\|_W}{\sqrt{N_j}}$ for any $j \in \{1,\ldots,m\}$. Set $\xi := (\xi_1,\cdots,\xi_m)$ and $T := \operatorname{diag}(\sqrt{N_1},\cdots,\sqrt{N_m})$. Note that $\|P\|_W = \|T\xi\|_2$, $\|T\|_{HS} = \sqrt{\sum_{j=1}^m N_j} = \sqrt{N}$, and $\|T\|_{op} := \max_{1\le j\le m}\sqrt{N_j}$. Then Lemma 3.12 implies

$$\operatorname{Prob}\left(\|P\|_W \le \varepsilon\sqrt{N}\right) \le (c\tilde{c_0}\varepsilon)^{cN\frac{\min_j N_j}{\max_j N_j}}.$$

Recall that Lemma 3.2 implies that for every $t \ge 1$,

$$\operatorname{Prob}\left(\|p\|_\infty \ge ctK\sqrt{m}\log(ed)\right) \le e^{-t^2 m\log(ed)}.$$

So, using our lower bound estimate for the condition number, we get

$$\operatorname{Prob}\left(\frac{\|P\|_W}{\|P\|_\infty} \ge \frac{c'\varepsilon\sqrt{N}}{tK\sqrt{m}\log(ed)}\right) \le \operatorname{Prob}\left(\tilde{\kappa}(P) \ge \frac{c\varepsilon\sqrt{N}}{tKmd\log(ed)}\right),$$

$$\operatorname{Prob}\left(\{\|P\|_W \ge c'\varepsilon\sqrt{N}\} \cap \{\|P\|_\infty \le ctK\sqrt{m}\log(ed)\}\right) \le \operatorname{Prob}\left(\tilde{\kappa}(P) \ge \frac{c\varepsilon\sqrt{N}}{tKmd\log(ed)}\right),$$

and

$$\text{Prob}\left(\|P\|_W \geq c'\varepsilon\sqrt{N}\right) + \text{Prob}\left(\|P\|_\infty \leq c\sqrt{m}\log(ed)\right) - 1 \geq 1 - (c\tilde{c_0}\varepsilon)^{-cN\frac{\min_j N_j}{\max_j N_j}} - e^{-t^2 m\log(ed)}.$$

We may choose $t := \sqrt{\log\frac{1}{\varepsilon}}$ and, by adjusting constants, we get our result. The case where $d_j = d$ for all $j \in \{1, \ldots, m\}$ is similar. The bounds for the expectation follow by integration. ■

Observe that the dominant factor in the very last estimate of Lemma 3.13 is $\sqrt{N}$, which is the normalization coming from the Weil-Bombieri norm of the polynomial system. So it makes sense to seek the asymptotic behavior of $\frac{\tilde{\kappa}(P)}{\sqrt{N}}$. When $m = n-1$, the upper bounds we get are exponential with respect to n, while the lower bounds are not. But when $m = 2n-4$ we have the following estimates when $d = d_j$ for all $j \in \{1, \ldots, m\}$:

$$\frac{A_1}{nd\log(ed)} \leq \frac{\mathbb{E}(\tilde{\kappa}(P))}{\sqrt{N}} \leq \frac{A_2 d\log ed}{\sqrt{n}},$$

where A_1, A_2 are constants depending on (K, c_0). This suggests that our estimates are closer to optimality when m is much larger than n.

Remark 3.14. *There are similarities between our probability tail estimates and the older estimates in the linear case studied in* [22]*. In particular our estimates in the quadratic case $d = 2$, when m is a constant multiple of n and $n \longrightarrow \infty$, are quite similar to the optimal result (for the linear case) appearing in* [22]*. This indicates that, in the proportional case (i.e., when m is a constant multiple of n and $n \longrightarrow \infty$), our tail bounds are close to optimal.* ⋄

References

[1] Franck Barthe and Alexander Koldobsky, *"Extremal slabs in the cube and the Laplace transform,"* Adv. Math. 174 (2003), pp. 89–114.

[2] Carlos Beltrán and Luis-Miguel Pardo, *"Smale's 17th problem: Average polynomial time to compute affine and projective solutions,"* Journal of the American Mathematical Society **22** (2009), pp. 363–385.

[3] Lenore Blum, Felipe Cucker, Mike Shub, and Steve Smale, *Complexity and Real Computation*, Springer-Verlag, 1998.

[4] Jean Bourgain, *"On the isotropy-constant problem for ψ_2-bodies,"* in *Geometric Aspects of Functional Analysis*, Lecture Notes in Mathematics, vol. 1807, pp. 114–121, Springer Berlin Heidielberg, 2003.

[5] Peter Burgisser and Felipe Cucker, *"On a problem posed by Steve Smale,"* Annals of Mathematics, pp. 1785–1836, Vol. 174 (2011), no. 3.

[6] Peter Burgisser and Felipe Cucker, *Condition,* Grundlehren der mathematischen Wissenschaften, no. 349, Springer-Verlag, 2013.

[7] John F. Canny, *The Complexity of Robot Motion Planning,* ACM Doctoral Dissertation Award Series, MIT Press, 1987.

[8] D. Castro, Juan San Martín, Luis M. Pardo, *"Systems of Rational Polynomial Equations have Polynomial Size Approximate Zeros on the Average,"* Journal of Complexity 19 (2003), pp. 161–209.

[9] Felipe Cucker, Teresa Krick, Gregorio Malajovich, and Mario Wschebor, *"A numerical algorithm for zero counting I. Complexity and accuracy,"* J. Complexity 24 (2008), no. 5–6, pp. 582–605

[10] Felipe Cucker, Teresa Krick, Gregorio Malajovich, and Mario Wschebor, *"A numerical algorithm for zero counting II. Distance to ill-posedness and smoothed analysis,"* J. Fixed Point Theory Appl. 6 (2009), no. 2, pp. 285–294.

[11] Felipe Cucker, Teresa Krick, Gregorio Malajovich, and Mario Wschebor, *"A numerical algorithm for zero counting III: Randomization and condition,"* Adv. in Appl. Math. 48 (2012), no. 1, pp. 215–248.

[12] Nikos Dafnis and Grigoris Paouris, *"Small ball probability estimates, Ψ_2-behavior and the hyperplane conjecture,"* Journal of Functional Analysis 258 (2010), pp. 1933–1964.

[13] Jean-Piere Dedieu, Mike Shub, *"Newton's Method for Overdetermined Systems Of Equations,"* Math. Comp. 69 (2000), no. 231, pp. 1099–1115.

[14] James Demmel, Benjamin Diament, and Gregorio Malajovich, *"On the Complexity of Computing Error Bounds,"* Found. Comput. Math. pp. 101–125 (2001).

[15] Wassily Hoeffding, *"Probability inequalities for sums of bounded random variables,"* Journal of the American Statistical Association, 58 (301):13–30, 1963.

[16] O. D. Kellog, *"On bounded polynomials in several variables,"* Mathematische Zeitschrift, December 1928, Volume 27, Issue 1, pp. 55–64.

[17] Bo'az Klartag and Emanuel Milman, *"Centroid bodies and the logarithmic Laplace Transform – a unified approach,"* J. Func. Anal., 262(1):10–34, 2012.

[18] Pierre Lairez, *"A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time,"* Foundations of Computational Mathematics, DOI 10.1007/s10208-016-9319-7

[19] Eric Kostlan, *"On the Distribution of Roots of Random Polynomials,"* Ch. 38 (pp. 419–431) of *From Topology to Computation: Proceedings of Smalefest (M. W. Hirsch, J. E. Marsden, and M. Shub, eds.)*, Springer-Verlag, New York, 1993.

[20] Hoi H. Nguyen, *"On a condition number of general random polynomial systems,"* Mathematics of Computation (2016) 85, pp. 737–757

[21] Mark Rudelson and Roman Vershynin, *"The Littlewood-Offord Problem and Invertibility of Random Matrices,"* Adv. Math. 218 (2008), no. 2, pp. 600–633.

[22] Mark Rudelson and Roman Vershynin, *"The Smallest Singular Value of Rectangular Matrix,"* Communications on Pure and Applied Mathematics 62 (2009), pp. 1707–1739.

[23] Mark Rudelson and Roman Vershynin, *"Small ball Probabilities for Linear Images of High-Dimensional Distributions,"* Int. Math. Res. Not. (2015), no. 19, pp. 9594–9617.

[24] Igor R. Shafarevich, *Basic Algebraic Geometry 1: Varieties in Projective Space*, 3rd edition, Springer-Verlag (2013).

[25] Mike Shub and Steve Smale, *"Complexity of Bezout's Theorem I. Geometric Aspects,"* J. Amer. Math. Soc. 6 (1993), no. 2, pp. 459–501.

[26] Roman Vershynin, *"Introduction to the Non-Asymptotic Analysis of Random Matrices,"* Compressed sensing, pp. 210–268, Cambridge Univ. Press, Cambridge, 2012.

[27] Assaf Naor and Artem Zvavitch, *"Isomorphic embedding of* ℓ_p^n, $1 < p < 2$, *into* $\ell_1^{(1+\varepsilon)n}$*,"* Israel J. Math. **122** (2001), pp. 371–380.

Department of Mathematics, 2108 SAS Hall Box 8205, NC State University, Raleigh, NC 27695-8205, USA.

E-mail address: aaergur@ncsu.edu

Department of Mathematics, Texas A&M University TAMU 3368, College Station, Texas 77843-3368, USA.

E-mail address: grigoris@math.tamu.edu

Department of Mathematics, Texas A&M University TAMU 3368, College Station, Texas 77843-3368, USA.

E-mail address: rojas@math.tamu.edu