

# SMOOTHED ANALYSIS FOR THE CONDITION NUMBER OF STRUCTURED REAL POLYNOMIAL SYSTEMS

ALPEREN A. ERGÜR, GRIGORIS PAOURIS, AND J. MAURICE ROJAS

ABSTRACT. We consider the sensitivity of real zeros of structured polynomial systems to perturbations of their coefficients. In particular, we provide explicit estimates for condition numbers of structured random real polynomial systems and extend these to a smoothed analysis setting.

## 1. INTRODUCTION

Efficiently finding real roots of real polynomial systems is one of the main objectives of computational algebraic geometry. There are numerous algorithms for this task, but the core steps of these algorithms are easy to outline: They are some combination of algebraic manipulation, a discrete/polyhedral computation, and a numerical iterative scheme.

From a computational complexity point of view, the cost of numerical iteration is much less transparent than the cost of algebraic or discrete computation. This paper constitutes a step toward understanding the complexity of numerically solving structured real polynomial systems. Our main results are Theorems 1.14, 1.16, and 1.18 below but we will first need to give some context for our results.

**1.1. How to control accuracy and complexity of numerics in real algebraic geometry?** In the numerical linear algebra tradition, going back to von Neumann and Turing, condition numbers play a central role in the speed of algorithms and the control of accuracy (see, e.g., [3, 6] for further background). Shub and Smale initiated the use of condition numbers for polynomial system solving over the field of complex numbers [37, 38]. Subsequently, condition numbers played a central role in the solution of Smale’s 17th problem [11, 6, 26].

The numerics of solving polynomial systems over the real numbers is more subtle than complex case: small perturbations can cause the solution set to change cardinality. One can even go from having no real zero to many real zeros by an arbitrarily small change in the coefficients. This behaviour doesn’t appear over the complex numbers as one has theorems (such as the Fundamental Theorem of Algebra) proving that root counts are “generically” constant. Luckily, a condition number theory that captures these subtleties was developed by Cucker [12]. Now we set up the notation and present Cucker’s definition.

**Definition 1.1** (Bombieri-Weyl Norm). *We set  $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$  where  $\alpha := (\alpha_1, \dots, \alpha_n)$ , and let  $P = (p_1, \dots, p_{n-1})$  be a system of homogenous polynomials with degree pattern  $d_1, \dots, d_{n-1}$ . Let  $c_{i,\alpha}$  denote the coefficient of  $x^\alpha$  in a  $p_i$ . We define the Weyl-Bombieri norms of  $p_i$  and  $P$  to be, respectively,*

$$\|p_i\|_W := \sqrt{\sum_{\alpha_1 + \dots + \alpha_n = d_i} \frac{|c_{i,\alpha}|^2}{\binom{d_i}{\alpha}}}$$

---

A.E. was partially supported by Einstein Foundation, Berlin and by the Pravesh Kothari of CMU. G.P. was partially supported by Simons Foundation Collaboration grant 527498 and NSF grant DMS-1812240. J.M.R. was partially supported by NSF grants CCF-1409020, DMS-1460766, and CCF-1900881.

and

$$\|P\|_W := \sqrt{\sum_{i=1}^{n-1} \|p_i\|_W^2}.$$

The following is Cucker’s condition number definition [12].

**Definition 1.2** (Real Condition Number). *For a system of homogenous polynomials  $P = (p_1, \dots, p_{n-1})$  with degree pattern  $(d_1, \dots, d_{n-1})$ , let  $\Delta_{n-1}$  be the diagonal matrix with entries  $\sqrt{d_1}, \dots, \sqrt{d_{n-1}}$  and let*

$$DP(x)|_{T_x S^{n-1}} : T_x S^{n-1} \longrightarrow \mathbb{R}^m$$

denote the linear map between tangent spaces induced by the Jacobian matrix of the polynomial system  $P$  evaluated at the point  $x \in S^{n-1}$ .

The local condition number of  $P$  at a point  $x \in S^{n-1}$  is

$$\tilde{\kappa}(P, x) := \frac{\|P\|_W}{\sqrt{\|DP(x)|_{T_x S^{n-1}}^{-1} \Delta_{n-1}\|^{-2} + \|P(x)\|_2^2}}$$

and the global condition number is

$$\tilde{\kappa}(P) := \sup_{x \in S^{n-1}} \tilde{\kappa}(P, x).$$

An important feature of Cucker’s real condition number is the following geometric fact[14].

**Theorem 1.3** (Real Condition Number Theorem). *We use  $H_D$  to denote the vector space of homogenous polynomial systems with degree pattern  $(d_1, \dots, d_{n-1})$ , and equip this space with the metric  $\rho(., .)$  induced by the Bombieri-Weyl norm. We define the set of ill-posed problems to be:  $\Sigma := \{P \in H_D : P \text{ has a singular zero in } S^{n-1}\}$ . Then we have  $\tilde{\kappa}(P) = \frac{\|P\|_W}{\rho(P, \Sigma)}$ .*

Cucker’s condition number is used in the design and analysis of a numerical algorithm for real zero counting [13, 14, 15], in the series of papers for computing homology groups of semialgebraic sets [16, 7, 8], and more recently in the analysis of a well-known algorithm for meshing curves and surfaces (the Plantinga-Vegter algorithm) [10].

One important observation is that the complexity of a numerical algorithm over the real numbers (imagine using bisection for finding real zeros of a given univariate polynomial) varies depending on the geometry of the input, and not just the bit complexity of its vector representation. Therefore it is more natural to go beyond worst-case analysis and seek quantitative bounds for “typical” inputs. We now explain the existing attempts toward mathematically modeling the intuitive phrase “typical input”.

**Random and adversarial random models.** Worst-case complexity theory, spearheaded by the P vs. NP question, has been a driving force behind many algorithmic breakthroughs in the last five decades. However, it has become clear that worst-case complexity theory fails to capture the practical performance of algorithms. The unreasonable effectiveness of everyday statistical methods are a case in point: the spotify app on cell phones solves instances of an NP-Hard problem all the time!

Two dominant paradigms for going beyond the worst-case analysis of algorithms are as follows: Assume an algorithm  $T$  operates on the input  $x \in \mathbb{R}^k$ , with the cost of output  $T(x)$  bounded from above by  $C(x)$ . One then equips the input space  $\mathbb{R}^k$  with a probability measure  $\mu$  and considers the average cost  $\mathbb{E}_{x \sim \mu} C(x)$ , or *smoothed analysis of the cost with*

parameter  $\delta > 0$ :  $\sup_{x \in \mathbb{R}^k} \mathbb{E}_{y \sim \mu} C(x + \delta \|x\| y)$ . Clearly, as  $\delta \rightarrow 0$ , smoothed complexity recovers worst-case complexity, and when  $\delta \rightarrow \infty$  we recover average-case complexity. It is also clear that to have a realistic complexity analysis, one should have a probability measure  $\mu$  that somehow reflects one's context, and use theorems that allow a broad class of measures  $\mu$ . The idea of smoothed analysis originated in work of Spielman and Teng [39].

**1.2. Existing results for average and smoothed analysis.** Existing results for the average analysis of real condition number from [15] can be roughly summarized as follows.

**Theorem 1.4** (Cucker, Krick, Malajovich, Wschebor). *Suppose*

$$p_i(x) := \sum_{\alpha_1 + \dots + \alpha_n = d_i} c_\alpha^{(i)} x^\alpha, \quad i = 1, \dots, n-1$$

are random polynomials where  $c_\alpha^{(i)}$  are centered Gaussian random variables with variances  $\binom{d_i}{\alpha}$ . Then, for the random polynomial system  $P = (p_1, \dots, p_{n-1})$  and for all  $t \geq 1$ , we have

$$\mathbb{P} \left\{ \tilde{\kappa}(P) \geq t \, 8d^{\frac{n+4}{2}} n^{5/2} N^{1/2} \right\} \leq \frac{(1 + \log(t))^{\frac{1}{2}}}{t}$$

where  $d = \max_i d_i$  and  $N = \sum_{i=1}^{n-1} \binom{n+d_i-1}{d_i}$ .

Recall the following smoothed analysis type result from [14]:

**Theorem 1.5** (Cucker, Krick, Malajovich, Wschebor). *Let  $Q$  be an arbitrary polynomial system with degree pattern  $(d_1, \dots, d_{n-1})$ , let  $P = (p_1, \dots, p_{n-1})$  be a random polynomial system as defined above. Now for a parameter  $0 < \delta < 1$  we define a random perturbation of  $Q$  with  $(P, \delta)$  as follows:  $G := Q + \delta \|Q\|_W P$ . Then we have*

$$\mathbb{P} \left\{ \tilde{\kappa}(G) \geq t \frac{13n^2 d^{2n+2} N}{\delta} \right\} \leq \frac{1}{t}.$$

**Remark 1.6.** *The randomness model considered in these seminal results has the following restriction: the induced probability measure is invariant under the action of the orthogonal group  $O(n)$  on the space of polynomials. The proof techniques used in the underlying papers seems to be only applicable when one has this group invariance property. This creates an obstruction against analysis on spaces of structured polynomials; spaces of structured polynomials are not necessarily closed under the action of  $O(n)$ , and hence do not support an  $O(n)$ -invariant probability measure.*

**1.3. What about structured polynomials?** Let  $H_{d_i}$  be the vector space of homogenous polynomials with  $n$  variables, and let  $H_D$  be the vector space of polynomial systems with degree pattern  $D = (d_1, \dots, d_{n-1})$ . Let  $E_i \subset H_{d_i}$  be linear subspaces for  $i = 1, \dots, n-1$ , and let  $E = (E_1, \dots, E_{n-1})$  be the corresponding vector space of polynomial systems.

For virtually any application of real root finding algorithms, the user has a polynomial system with a particular structure rather than a generic polynomial system with  $N = \sum_{i=1}^{n-1} \binom{n+d_i-1}{d_i}$  many coefficients. Suppose a user has identified the linear structure  $E$  that is present in the target equations, and would like to know about how much precision is expected for round-off errors in the space  $E$ . One could induce a probability measure  $\mu$  on  $E$  and use  $\mathbb{E}_{P \sim \mu} \log(\tilde{\kappa}(P))$  to determine the expected precision of numerical solutions. What could go wrong?

**Example 1.7.** Let  $u, v \in S^{n-1}$  be two vectors with  $u \perp v$ , and define the following subspaces:

$$E_i := \{p \in H_{d_i} : p(u) = \langle \nabla p(u), v \rangle = 0\}, \quad i = 1, \dots, n-1$$

where  $\nabla p(u)$  denotes the the gradient of  $p$  evaluated at  $u$ .  $E_i$  are codimension 2 linear subspaces of  $H_{d_i}$ . Now consider the space of polynomials  $E := (E_1, \dots, E_{n-1})$ ; any polynomial system in the space  $E$  has a singular real zero at  $u$ . Hence, for all  $P \in E$  the condition number  $\tilde{\kappa}(P)$  is infinite.

The preceding example illustrates that, for certain linear spaces  $E$ , the probabilistic analysis of condition numbers is meaningless: It is possible for *certain* spaces  $E$  that all inhabitants have infinite condition number. We will rule out these degenerate cases as follows.

**Definition 1.8** (Non-degenerate linear space). We call a linear space  $E_i \subset H_{d_i}$  non-degenerate if for all  $v \in S^{n-1}$ , there exists an element  $p_i \in E_i$  with  $p_i(v) \neq 0$ . In other words,  $E_i$  is non-degenerate if there is no base point  $v \in S^{n-1}$  where all the elements of  $E_i$  vanish all together. We call a space of polynomial systems  $E = (E_1, \dots, E_{n-1})$  non-degenerate if all  $E_i$  are non-degenerate for  $i = 1, \dots, n-1$ .

An easy corollary of Theorem 1.14 shows that the expected precision is finite for any non-degenerate space  $E$ .

**Corollary 1.9.** Let  $E \subset H_D$  be a non-degenerate linear space of polynomials. Let  $\mu$  be a probability measure supported on the space  $E$  that satisfies the assumptions listed in section 1.5. Then  $\mathbb{E}_{P \sim \mu} \log(\tilde{\kappa}(P))$  is finite.

This is clearly not the end of the story: a non-degenerate linear structure  $E$  may still be close to being degenerate, and this would make every element in the space  $E$  ill-conditioned. So we need to somehow quantify the numerical conditioning of a linear structure  $E$ . Next, we introduce the notion of *dispersion* as a rough measure of conditioning of a linear structure.

**1.4. The dispersion constant of a linear space.** Suppose a linear subspace  $F \subset H_d$  is given for some  $d > 1$  together with an orthonormal basis  $u_j(x)$ ,  $j = 1, \dots, m$  with respect to Bombieri-Weyl norm. Now suppose for a particular point  $v_0 \in S^{n-1}$  all the basis elements satisfy  $\text{abs}u_j(v_0) < \varepsilon$  where  $\varepsilon > 0$  is small. What kind of behavior one would expect from elements of  $F$  at the point  $v_0$ ? This point  $v_0$  would behave like a base point (like if all elements of  $F$  vanishes at  $v_0$ ) unless one employs rather high precision. This motivates the following definition.

**Definition 1.10** (Dispersion constant of a linear space of polynomials). Let  $F \subset H_d$  is given for some  $d > 1$ , and let  $u_j(x)$ ,  $j = 1, \dots, m$  be an orthonormal basis of  $F$  with respect to Bombieri-Weyl norm. We define the following two quantities

$$\sigma_{\min}(F) := \min_{v \in S^{n-1}} \left( \sum_j u_j(v)^2 \right)^{\frac{1}{2}}, \quad \sigma_{\max}(F) := \max_{v \in S^{n-1}} \left( \sum_j u_j(v)^2 \right)^{\frac{1}{2}}$$

and the dispersion constant  $\sigma(F)$  is their ratio:

$$\sigma(F) := \frac{\sigma_{\max}(F)}{\sigma_{\min}(F)}$$

The quantity  $\sigma_{\max}$  is introduced to make things scale invariant. We generalize the definition to polynomial systems in a straight-forward manner.

**Definition 1.11** (Dispersion constant of a linear space of polynomial systems). *Let  $E_i \subset H_{d_i}$  be linear spaces for  $i = 1, \dots, n-1$ , and let  $E = (E_1, \dots, E_{n-1})$ . We define the dispersion constant  $\sigma(E)$  as follows:  $\sigma(E) := \max_i \sigma(E_i)$ .*

Our estimates replace the dimension  $N$  in earlier results with the (potentially much smaller) dimension of  $E$ , at the expense of involving the new quantity  $\sigma(E)$ . So if a user has a fixed structure  $E$  with small dimension and tame dispersion constant, then the expected conditioning on  $E$  admits a much better bound than what earlier results suggest. On the other hand, if one has a sparse but highly sensitive structure, the resulting average-case conditioning could be a lot worse than the average over the entire space  $H_D$ .

**How big is the dispersion constant?** To better understand the dispersion constant, let us consider two examples at opposite extremes.

**Example 1.12** (A subspace with minimal dispersion constant). *Consider subspaces of polynomials  $F_i \subset P_{n,2d_i}$  defined as the span of*

$$u_{kl}^{(i)} = (x_1^2 + \dots + x_n^2)^{d_i-1} x_k x_l \text{ for } 1 \leq k, l \leq n$$

and let  $F = (F_1, \dots, F_n)$ . It is easy to show that  $\sigma(F) = 1$ .

**Example 1.13** (A sparse but highly sensitive structure). *Let  $E \subset P_{n,d}$  be the subspace of polynomials spanned by the monomials  $x_1^d, \dots, x_n^d$ . Then, we have  $\sigma(E) = n^{\frac{d-1}{2}}$ .*

One may wonder how big the dispersion constant for a “typical” linear space  $E$  is, for say,  $E$  of dimension around  $n^2 \log d$ . Would a typical low-dimensional space look like the second example or the first example? We address this question in Appendix A. For our main theorems, we will allow  $E$  to be arbitrary and give bounds depending explicitly on the dispersion constant  $\sigma(E)$ .

**1.5. A general model of randomness for structured polynomial systems.** In our precursor paper [19] we obtained probabilistic condition number estimates for general measures (without any group invariance assumption). In this paper we present probabilistic results for the same general family of measures, but this time supported on a structured space  $E$  instead of  $H_D$ . Note that here the structured space  $E$  will be fixed by the user, and our results will give estimates for a random element from  $E$ . First, we introduce our general model of randomness.

We say a random vector  $X \in \mathbb{R}^n$  satisfies the *Centering*, *sub-Gaussian*, and *Small Ball* properties, with constants  $K$  and  $c_0$ , if the following hold true:

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

We note that these three assumptions directly yield a relation between  $K$  and  $c_0$ : We in fact have  $Kc_0 \geq \frac{1}{4}$  (see [19] just before Section 3.2). Moreover, for a random variable  $X$  that satisfies above assumptions with constants  $K$  and  $c_0$ , and a scalar  $\lambda > 0$ , the random variable  $\lambda X$  satisfies the above assumptions with constants  $\lambda K$  and  $\lambda^{-1}c_0$ . In other words  $Kc_0$  is invariant under scaling, hence one can hope for a universal lower bound of  $\frac{1}{4}$ .

<sup>1</sup>Equivalently,  $\mathbb{E}X = \mathbf{0}$ .

Random vectors that satisfy these three properties form a large family of distributions, including standard Gaussian vectors and uniform measures on a large family of convex bodies called  $\Psi_2$ -bodies (such as uniform measures on  $l_p$ -balls for all  $p \geq 2$ ). We refer the reader to the book of Vershynin [45] for more details. Discrete sub-Gaussian distributions, such as the Bernoulli distribution, also satisfy an inequality similar to the small-ball inequality in our assumptions. However, the small-ball type inequality satisfied by such discrete distributions depends not only on the norm of the deterministic vector  $a$  but also on the arithmetic structure of  $a$ . It is possible that our methods, combined with the work of Rudelson and Vershynin on the Littlewood-Offord problem [33], can extend our main results to discrete distributions such as the Bernoulli distribution. In this work, we will content ourselves with continuous distributions.

The preceding examples of random vectors do not necessarily have independent coordinates. This provides important extra flexibility. There are also interesting examples of random vectors with independent coordinates. In particular, if  $X_1, \dots, X_m$  are independent centered random variables that each satisfy both the sub-Gaussian inequality with constant  $K$  and the Small Ball condition with  $c_0$ , then the random vector  $X = (X_1, \dots, X_m)$  also satisfies the sub-Gaussian and Small Ball inequalities with constants  $C_1 K$  and  $C_2 c_0$ , where  $C_1$  and  $C_2$  are universal constants. This is a relatively new result of Rudelson and Vershynin [35]. The best possible universal constant  $C_2$  is discussed in [29, 32]. To create a random variable satisfying the Small Ball and sub-Gaussian properties one can, for instance, start by fixing any  $p \geq 2$  and then considering a random variable with density function  $f(t) := c_p e^{-|t|^p}$ , for suitably chosen positive  $c_p$ .

**1.6. Our Results.** We present estimates for random structured polynomial systems, where the randomness model is the one introduced in the preceding section.

### Average-case condition number estimates for structured polynomial systems

**Theorem 1.14.** *Let  $E_i \subseteq H_{d_i}$  be non-degenerate linear subspaces, and let  $E = (E_1, \dots, E_{n-1})$ . Assume  $\dim(E) \geq n \log(ed)$  and  $n \geq 3$ . Let  $p_i \in E_i$  be independent random elements of  $E_i$  that satisfy the Centering property, the sub-Gaussian property with constant  $K$ , and the Small Ball property with constant  $c_0$ , each with respect to the Bombieri-Weyl inner product. We set  $d := \max_i d_i$  and  $M := nK \sqrt{\dim(E)} (c_0 C K d^2 \log(ed) \sigma(E))^{2n-2}$ , where  $C \geq 4$  is a universal constant. Then for the random polynomial system  $P = (p_1, \dots, p_{n-1})$ , we have*

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \begin{cases} 3t^{-\frac{1}{2}} & ; \text{if } 1 \leq t \leq e^{2n \log(ed)} \\ (e^2 + 1)t^{-\frac{1}{2} + \frac{1}{4 \log(ed)}} & ; \text{if } e^{2n \log(ed)} \leq t \end{cases}$$

Moreover, for  $0 < q < \frac{1}{2} - \frac{1}{2 \log(ed)}$ , we have  $\mathbb{E}(\tilde{\kappa}(P)^q) \leq M^q (1 + 4q \log(ed))$ . In particular,  $\mathbb{E} \log(\tilde{\kappa}(P)) \leq 1 + \log M$ .

**Smoothed analysis for structured polynomial systems** For smoothed analysis we need to introduce a slightly stronger assumption on the random input. This slightly stronger property is called the *Anti-Concentration Property* and it replaces the Small Ball assumption in our model of randomness. We will need a bit of terminology to define anti-concentration.

**Definition 1.15** (Anti-Concentration Property). *For any real-valued random variable  $Z$  and  $t \geq 0$ , the concentration function,  $F(Z, t)$ , is defined as  $F(Z, t) := \max_{u \in \mathbb{R}} \text{Prob}\{|Z - u| \leq t\}$ . Let  $\langle \cdot, \cdot \rangle$  denote the standard inner product on  $\mathbb{R}^n$ . We then say a random vector  $X \in \mathbb{R}^n$  satisfies the Anti-Concentration Property with constant  $c_0$  if we have  $F(\langle X, \theta \rangle, \varepsilon) \leq c_0 \varepsilon$  for all  $\theta \in S^{n-1}$ .  $\diamond$*

It is easy to check that if the random variable  $Z$  has bounded density  $f$  then  $F(Z, t) \leq \|f\|_\infty t$ . Moreover, the Lebesgue Differentiation theorem states that upper bounds for the function  $t^{-1}F(Z, t)$  for all  $t$  imply upper bounds for  $\|f\|_\infty$ . See [34] for the details.

**Theorem 1.16.** *Let  $E \subseteq H_D$  be a non-degenerate linear subspace for  $D = (d_1, \dots, d_{n-1})$ . Assume  $\dim(E) \geq n \log^2(ed)$  and  $n \geq 3$ . Let  $Q \in E$  be a fixed (deterministic) polynomial system let  $G \in E$  be a random polynomial system given by the same model of randomness as in Theorem 1.14, but with the Small Ball Property replaced by the Anti-Concentration Property. Set  $d := \max_i d_i$ , and*

$$M := nK \sqrt{\dim(E)} (c_0 d^2 C K \log(ed) \sigma(E))^{2n-2} \left(1 + \frac{\|Q\|_W}{\sqrt{n} K \log(ed)}\right)^{2n-1}$$

where  $C \geq 4$  is a universal constant. Then for the randomly perturbed polynomial system  $P = Q + G$ , we have

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \begin{cases} 3t^{-\frac{1}{2}} & ; \text{if } 1 \leq t \leq e^{2n \log(ed)} \\ (e^2 + 1)t^{-\frac{1}{2} + \frac{1}{4 \log(ed)}} & ; \text{if } e^{2n \log(ed)} \leq t \end{cases}$$

Moreover, for  $0 < q < \frac{1}{2} - \frac{1}{2 \log(ed)}$ , we have

$$\mathbb{E}(\tilde{\kappa}(P)^q) \leq M^q (1 + 4q \log(ed)).$$

In particular,  $\mathbb{E} \log(\tilde{\kappa}(P)) \leq 1 + \log M$ .

We would like to consider a corollary to make the result easier to parse.

**Corollary 1.17.** *Let  $E \subseteq H_D$  be a non-degenerate linear subspace for  $D = (d_1, \dots, d_{n-1})$ . Assume  $\dim(E) \geq n \log(ed)^2$  and  $n \geq 3$ . Let  $Q \in E$  be a fixed (deterministic) polynomial system, and let  $G \in E$  be a random polynomial system given by the model of randomness as in Theorem 1.16, but with fixed  $K = 1$ . Now let  $0 < \delta < 1$  be a parameter and consider the polynomial system*

$$P := Q + \delta \|Q\|_W G$$

We set  $d := \max_i d_i$ , and

$$M := n \sqrt{\dim(E)} (c_0 C d^2 \log(ed) \sigma(E))^{2n-2} \delta \|Q\|_W \left(1 + \frac{1}{\delta \sqrt{n} \log(ed)}\right)^{2n-1}$$

where  $C \geq 4$  is a universal constant. Then, we have

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \begin{cases} 3t^{-\frac{1}{2}} & ; \text{if } 1 \leq t \leq e^{2n \log(ed)} \\ (e^2 + 1)t^{-\frac{1}{2} + \frac{1}{4 \log(ed)}} & ; \text{if } e^{2n \log(ed)} \leq t \end{cases}$$

**An interesting consequence** As a corollary of the smoothed analysis estimate in Theorem 1.16, we derive the following structural result.

**Theorem 1.18.** *Let  $E_i \subseteq H_{d_i}$  be full linear subspaces, let  $E = (E_1, \dots, E_{n-1})$ , and let  $Q \in E$ . Then, for every  $0 < \varepsilon < 1$ , there is a polynomial system  $P_\varepsilon \in E$  with the following properties:*

$$\|P_\varepsilon - Q\|_W \leq \varepsilon \|Q\|_W \left( \frac{\sqrt{\dim(E)}}{\log(ed) \sqrt{n}} \right)$$

and

$$\tilde{\kappa}(P_\varepsilon) \leq \sqrt{n} \sqrt{\dim(E)} \left( \frac{d^2 C \log(ed) \sigma(E)}{\varepsilon} \right)^{2n-2}$$

for a universal constant  $C$ .

One can view this result as a metric entropy statement as follows: Suppose we are given a bounded set  $\mathbb{T} \subset E$  with  $\sup_{P \in \mathbb{T}} \|P\|_W \leq 1$ , and we would like to cover  $\mathbb{T}$  with balls of radius  $\delta$ , i.e.,  $\mathbb{T} = \bigcup_i B(p_i, \delta)$ . Moreover, suppose we want the ball-centers  $p_i$  to have a controlled condition number. We can start with an arbitrary  $\frac{\delta}{2}$  covering  $\mathbb{T} = \bigcup_i B(p_i, \frac{\delta}{2})$ , and use Theorem 1.18 with  $\varepsilon = \frac{\delta\sqrt{n}}{2\sqrt{\dim(E)}}$  to find a  $p_i$  with controlled condition number in each one of the balls  $B(p_i, \frac{\delta}{2})$ . Then  $\mathbb{T} = \bigcup_i B(p_i, \delta)$  gives a  $\delta$ -covering of  $\mathbb{T}$  where  $p_i$  has controlled condition number.

## 2. BACKGROUND AND BASIC ESTIMATES

We first present a simple lemma for a single random polynomial.

**Lemma 2.1.** *Let  $F \subset H_d$  be non-degenerate linear subspace of degree  $d$  homogenous polynomials. We equip  $F$  with Bombieri-Weyl norm. Suppose  $p \in F$  is a random element that satisfies centering property, sub-Gaussian property with constant  $K$ , and small probability with constant  $c_0$  each with respect to Bombieri inner product. Then for all  $w \in S^{n-1}$  the following estimates hold:*

$$\text{Prob}(|p(w)| \geq t\sigma_{\max}(F)) \leq \exp\left(1 - \frac{t^2}{K^2}\right)$$

$$\text{Prob}(|p(w)| \leq \varepsilon\sigma_{\min}(F)) \leq c_0\varepsilon.$$

*Proof.* Suppose  $u_1, \dots, u_m$  is an orthonormal basis of  $F$  with respect to Bombieri-Weyl inner product. Let  $f \in F$  be a polynomial with  $f(x) = \sum_i f_i u_i(x)$ , then for any  $v \in S^{n-1}$ , clearly  $f(v) = \sum_i a_i u_i(v)$ . In other words, if we set  $q_v := \sum_i u_i(v) u_i(x)$  then we have  $f(v) = \langle f, q_v \rangle_W$ . Also note since  $u_i$  is an orthonormal basis with respect to Bombieri norm, we have  $\|q_v\|_W = (\sum_i u_i(v)^2)^{\frac{1}{2}}$ .

Now let  $p \in E'$  be the random element described above. The reasoning in the preceding paragraph gives us the following estimates for any fixed point  $v \in S^{n-1}$ :

$$\text{Prob}(|p(v)| \geq t\|q_v\|_W) \leq \exp\left(1 - \frac{t^2}{K^2}\right)$$

$$\text{Prob}(|p(v)| \leq \varepsilon\|q_v\|_W) \leq c_0\varepsilon.$$

By the definition of  $\sigma_{\max}(F)$  and  $\sigma_{\min}(F)$  these pointwise estimates yield the desired result.  $\blacksquare$

The following is the generalization of Lemma 2.1 to systems of polynomials.

**Lemma 2.2.** *Let  $D = (d_1, \dots, d_{n-1}) \in \mathbb{N}^{n-1}$ . For all  $i \in \{1, \dots, n-1\}$  let  $E_i \subseteq H_{d_i}$  be non-degenerate linear subspaces, and let  $E := (E_1, \dots, E_{n-1})$ . For each  $i$ , let  $p_i$  be chosen from  $E_i$  via a distribution satisfying the Centering Property, the Sub-Gaussian Property with constant  $K$ , and the Small Ball Property with constant  $c_0$  (each with respect to the Bombieri-Weyl inner product). Then, for the random polynomial system  $P = (p_1, \dots, p_{n-1})$ , and all  $v \in S^{n-1}$ , the following estimates hold:*

$$\text{Prob}(\|P(v)\|_2 \geq t\sigma_{\max}(E)\sqrt{n-1}) \leq \exp\left(1 - \frac{a_1 t^2 (n-1)}{K^2}\right)$$



$$\text{and } \text{Prob} \left( \|P(v)\|_2 \leq \varepsilon \sigma_{\min}(E) \sqrt{n-1} \right) \leq (a_2 c_0 \varepsilon)^{n-1},$$

where  $a_1$  and  $a_2$  are absolute constants.

For the proof of Lemma 2.2 we need to recall some theorems from probability theory and some basic tools developed in our earlier work [19]. These basic lemmata will also be used throughout the paper. We start with a theorem which is reminiscent of Hoeffding's classical inequality [20].

**Theorem 2.3.** [47, Prop. 5.10] *There is an absolute constant  $\tilde{c}_1 > 0$  with the following property: If  $X_1, \dots, X_n$  are centered, sub-Gaussian random variables with constant  $K$ ,  $a = (a_1, \dots, a_n) \in \mathbb{R}^n$  and  $t \geq 0$ , then*

$$\text{Prob} \left( \left| \sum_i a_i X_i \right| \geq t \right) \leq 2 \exp \left( \frac{-\tilde{c}_1 t^2}{K^2 \|a\|_2^2} \right). \quad \blacksquare$$

We will also need the following standard lemma (see, e.g., [33, Lemma 2.2]).

**Lemma 2.4.** *Assume  $Z_1, \dots, Z_n$  are independent random variables that have the property that  $F(Z_i, t) \leq c_0 t$  for all  $t > 0$ . Then for  $t > 0$  we have  $F(W, t\sqrt{n}) \leq (cc_0 t)^n$ , where  $W := \|(Z_1, \dots, Z_n)\|_2$ . Moreover, if  $\xi_1, \dots, \xi_k$  are independent random variables such that, for every  $\varepsilon > 0$ , we have  $\text{Prob}(|\xi_i| \leq \varepsilon) \leq c_0 \varepsilon$ . Then there is a universal constant  $\tilde{c} > 0$  such that for every  $\varepsilon > 0$  we have  $\text{Prob} \left( \sqrt{\xi_1^2 + \dots + \xi_k^2} \leq \varepsilon \sqrt{k} \right) \leq (\tilde{c} c_0 \varepsilon)^k$ .  $\blacksquare$*

Now that we have our basic probabilistic tools we proceed to deriving some deterministic inequalities.

The lemma below was proved in our earlier paper [19], generalizing a classical Theorem of Kellog [21]. To state the lemma we need a bit of terminology: For any system of homogeneous polynomials  $P := (p_1, \dots, p_{n-1}) \in (\mathbb{R}[x_1, \dots, x_n])^{n-1}$  define  $\|P\|_\infty := \sup_{x \in S^{n-1}} \sqrt{\sum_{i=1}^{n-1} p_i(x)^2}$ . Let  $DP(x)$  denote the Jacobian matrix of the polynomial system at point  $x$ , let  $DP(x)(u)$  denote the image of the vector  $u$  under the linear operator  $DP(x)$ , and set  $\|D^{(1)}P\|_\infty := \sup_{x, u \in S^{n-1}} \|DP(x)(u)\|_2$ . (Alternatively, the last quantity can be written  $\sup_{x, u \in S^{n-1}} \sqrt{\sum_{i=1}^{n-1} \langle \nabla p_i(x), u \rangle^2}$ .)

**Lemma 2.5.** *Let  $P := (p_1, \dots, p_{n-1}) \in (\mathbb{R}[x_1, \dots, x_n])^{n-1}$  be a polynomial system with  $p_i$  homogeneous of degree  $d_i$  for each  $i$  and set  $d := \max_i d_i$ . Then:*

- (1) *We have  $\|D^{(1)}P\|_\infty \leq d^2 \|P\|_\infty$  and, for any mutually orthogonal  $x, y \in S^{n-1}$ , we also have  $\|DP(x)(y)\|_2 \leq d \|P\|_\infty$ .*
- (2) *If  $\deg(p_i) = d$  for all  $i \in \{1, \dots, n-1\}$  then we also have  $\|D^{(1)}P\|_\infty \leq d \|P\|_\infty$ .  $\blacksquare$*

The final lemma we need is a discretization tool for homogenous polynomial systems that was developed in [19] based on Lemma 2.5. We need a bit of terminology to state the lemma.

**Definition 2.6.** *Let  $K$  be a compact set in a metric space  $(X, d)$ , then a set  $A \subseteq K$  with finitely many elements is called a  $\delta$ -net if for every  $x \in K$  there exists  $y \in A$  with  $d(x, y) \leq \delta$ .*

For the unit sphere in  $\mathbb{R}^n$ , equipped with the standard Euclidean metric, there are known bounds for the size of a  $\delta$ -net. We recall one such bound below.

**Lemma 2.7.** *Let  $S^{n-1}$  be the unit sphere in  $\mathbb{R}^n$  with respect to standard euclidean metric. Then for every  $\delta > 0$ , there exist a  $\delta$ -net  $\mathcal{N} \subset S^{n-1}$  with size at most  $2n(1 + \frac{2}{\delta})^{n-1}$ .*

Lemma 2.7 is almost folklore: a proof appears in Proposition 2.1 of [34].

**Lemma 2.8.** *Let  $P = (p_1, \dots, p_{n-1})$  be a system of homogenous polynomials  $p_i$  with  $n$  variables and  $\deg(p_i) = d_i$ . Let  $\mathcal{N}$  be a  $\delta$ -net on  $S^{n-1}$ . Let  $\max_{\mathcal{N}}(P) = \sup_{y \in \mathcal{N}} \|P(y)\|_2$  and  $\|P\|_\infty = \sup_{x \in S^{n-1}} \|P(x)\|_2$ . Similarly let us define,*

$$\max_{\mathcal{N}^{k+1}}(D^{(k)}P) = \sup_{x, u_1, \dots, u_k \in \mathcal{N}} \|D^{(k)}P(x)(u_1, \dots, u_k)\|_2$$

$$\text{and } \|D^{(k)}P\|_\infty = \sup_{x, u_1, \dots, u_k \in S^{n-1}} \|D^{(k)}P(x)(u_1, \dots, u_k)\|_2.$$

Then

(1) When  $\deg(p_i) = d$  for all  $i \in \{1, \dots, m\}$  we have  $\|P\|_\infty \leq \frac{\max_{\mathcal{N}}(P)}{1-d\delta}$  and

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

(2) When  $\max_i \{\deg(p_i)\} \leq d$  we have  $\|P\|_\infty \leq \frac{\max_{\mathcal{N}}(P)}{1-d^2\delta}$  and  $\|D^{(k)}P\|_\infty \leq \frac{\max_{\mathcal{N}^{k+1}}(D^{(k)}P)}{1-\delta d^2 \sqrt{k+1}}$ . ■

**Proof of Lemma 2.2:** We begin with the first claim. Using Lemma 2.1 and the fact that  $\sigma_{\max}(E) \geq \sigma_{\max}(E_i)$  for all  $i$ , we get the following estimate for any  $p_i \in E_i$  and  $w \in S^{n-1}$ :

$$\text{Prob}(|p_i(w)| \geq s\sigma_{\max}(E)) \leq \exp\left(1 - \frac{s^2}{K^2}\right).$$

Now let  $a = (a_1, \dots, a_{n-1}) \in \mathbb{R}^{n-1}$  with  $\|a\|_2 = 1$ , and apply Lemma 2.3 to the sub-Gaussian random variables  $\frac{p_i(w)}{\sigma_{\max}(E)}$  and the vector  $a$ . We then get

$$\text{Prob}\left(\left|\sum_i a_i p_i(w)\right| \geq s\sigma_{\max}(E)\right) \leq \exp\left(1 - \frac{\tilde{c}_1 s^2}{K^2}\right).$$

Observe that  $\|P(w)\|_2 = \max_{a \in S^{n-2}} |\langle a, P(w) \rangle|$ . For any fixed point  $w \in S^{n-1}$  and a free variable  $a \in \mathbb{R}^n$ , we have that  $\langle a, P(w) \rangle$  is a linear polynomial on  $a$ . We then use Lemma 2.8 on this linear polynomial, which gives us the following estimate:

$$\text{Prob}\left(\|P(w)\|_2 \geq \frac{s\sigma_{\max}(E)}{1-\delta}\right) \leq |\mathcal{N}| \exp\left(1 - \frac{\tilde{c}_1 s^2}{K^2}\right).$$

We then use Lemma 2.7 to control the cardinality of the  $\delta$ -net and get

$$|\mathcal{N}| \leq 2n\left(1 + \frac{2}{\delta}\right)^{n-1} \leq e^{(n-1)\tilde{c}\log(\frac{1}{\delta})},$$

for some absolute constant  $\tilde{c}$ . So we set  $t = 2s\sqrt{n-1}$ ,  $\delta = \frac{1}{2}$ , and obtain the following estimate for some universal constant  $a_1$ .

$$\text{Prob}\left(\|P(w)\|_2 \geq t\sigma_{\max}(E)\sqrt{n-1}\right) \leq \exp\left(1 - \frac{a_1 t^2(n-1)}{K^2}\right).$$

We continue with the proof of the second claim. Using Lemma 2.1 and the fact that  $\sigma_{\min}(E) \leq \sigma_{\min}(E_i)$  for all  $i$ , we deduce the following estimate for all  $p_i$  and for any  $\varepsilon > 0$ :

$$\text{Prob}\left(\left|\frac{p_i(w)}{\sigma_{\min}(E)}\right| \leq \varepsilon\right) \leq c_0\varepsilon.$$

Using Lemma 2.4 on the random variables  $\left|\frac{p_i(w)}{\sigma_{\min}(E)}\right|$  gives the following estimate:

$$\text{Prob}\left(\|P(w)\|_2 \leq \varepsilon\sigma_{\min}(E)\sqrt{n-1}\right) \leq (\tilde{c}_2 c_0 \varepsilon)^{n-1}. \quad \blacksquare$$

## 3. OPERATOR NORM TYPE ESTIMATES

In this section we will estimate the absolute maximum norm of a random polynomial system on the sphere. Recall that for a homogenous polynomial system  $P = (p_1, \dots, p_{n-1})$  the sup-norm is defined as  $\|P\|_\infty = \sup_{x \in S^{n-1}} \|P(x)\|_2$ . The following lemma is our sup-norm estimate for a random polynomial system  $P$ .

**Lemma 3.1.** *Let  $D = (d_1, \dots, d_{n-1})$  be a vector with positive integer coordinates, let  $E_i \subseteq H_{d_i}$  be full linear subspaces, and let  $E = (E_1, \dots, E_{n-1})$ . Let  $p_i \in E_i$  be independent random elements of  $E_i$  that satisfy the Centering Property, the Sub-Gaussian Property with constant  $K$ , and the Small Ball Property with constant  $c_0$ , each with respect to Bombieri-Weyl inner product. Let  $\mathcal{N}$  be a  $\delta$ -net on  $S^{n-1}$ . Then for  $P = (p_1, \dots, p_{n-1})$  we have*

$$\text{Prob} \left( \max_{x \in \mathcal{N}} \|P(x)\|_2 \geq t \sigma_{\max}(E) \sqrt{n} \right) \leq |\mathcal{N}| \exp \left( 1 - \frac{a_1 t^2 n}{K^2} \right),$$

where  $a_1$  is a universal constant. In particular, for  $d = \max_i \deg(p_i)$ ,  $\delta = \frac{1}{3d^2}$ , and  $t = s \log(ed)$  with  $s \geq 1$  this gives us the following estimate

$$\text{Prob} \left( \|P\|_\infty \geq s \sigma_{\max}(E) \sqrt{n} \log(ed) \right) \leq \exp \left( 1 - \frac{a_3 s^2 n \log(ed)^2}{K^2} \right)$$

where  $a_3$  is a universal constant.

*Proof.* The first statement is proven by just taking a union bound over  $\mathcal{N}$  and using Lemma 2.2. The second part of the statement immediately follows by using the first part and Lemma 2.8. ■

## 4. SMALL BALL TYPE ESTIMATES

We define the following quantity for later convenience.

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

It follows directly that

$$\frac{\|P\|_W}{\tilde{\kappa}(P, x)} = \sqrt{\|P\|_W^2 \tilde{\mu}_{\text{norm}}(P, x)^{-2} + \|P(x)\|_2^2} = \inf_{\substack{y \perp x \\ y \in S^{n-1}}} \mathcal{L}(x, y)$$

So we set  $L(P, x) = \frac{\|P\|_W}{\tilde{\kappa}(P, x)}$  and  $L(P) = \min_{x \in S^{n-1}} L(P, x)$ . We then have the following equalities:

$$L(P, x) = \inf_{\substack{y \perp x \\ y \in S^{n-1}}} \mathcal{L}(x, y), \quad \tilde{\kappa}(P, x) = \frac{\|P\|_W}{L(P, x)}$$

and, finally

$$\tilde{\kappa}(P) = \frac{\|P\|_W}{L(P)}.$$

In this section, we prove a small-ball type estimate to control behavior of the denominator  $L(p)$ . We first need to recall a technical lemma from our earlier paper [19], which builds on an idea of Nguyen [31].

**Lemma 4.1.** *Let  $n \geq 2$ , let  $P := (p_1, \dots, p_{n-1})$  be a system of  $n$ -variate homogenous polynomials, and assume  $\|P\|_\infty \leq \gamma$ . Let  $x, y \in S^{n-1}$  be mutually orthogonal vectors with  $\mathcal{L}(x, y) \leq \alpha$ , and let  $r \in [-1, 1]$ . Then for every  $w$  with  $w = x + \beta ry + \beta^2 z$  for some  $z \in B_2^n$ , we have the following inequalities:*

- (1) *If  $d := \max_i d_i$  and  $0 < \beta \leq d^{-4}$  then  $\|P(w)\|_2^2 \leq 8(\alpha^2 + (2 + e^4)\beta^4 d^4 \gamma^2)$ .*
- (2) *If  $\deg(p_i) = d$  for all  $i \in [n-1]$  and  $0 < \beta \leq d^{-2}$  then  $\|P(w)\|_2^2 \leq 8(\alpha^2 + (2 + e^4)\beta^4 d^4 \gamma^2)$ . ■*

We also need to state and prove the following simple Lemma for the clarity of succeeding proofs.

**Lemma 4.2.** *Let  $n \geq 1$  be an integer. Then for  $0 \leq x \leq \frac{1}{n}$  we have  $(1+x)^n \leq 1+3nx$ .*

*Proof.* For every  $0 \leq y \leq 1$  we have  $1+3y \geq e^y$ . This can be seen by setting  $f(y) = 1+3y-e^y$ , observing  $f'(y) > 0$  for all  $0 \leq y \leq 1$  and  $f(1) > 0, f(0) = 0$ . With a similar reasoning one can prove  $e^x \geq 1+x$ , and hence  $e^{nx} \geq (1+x)^n$  for all  $0 \leq x \leq 1$ . Using  $y = nx$  completes the proof. ■

**Theorem 4.3.** *Let  $D = (d_1, \dots, d_{n-1})$  be a vector with positive integer coordinates, let  $E_i \subseteq H_{d_i}$  be full linear subspaces, and let  $E = (E_1, \dots, E_{n-1})$ . Let  $p_i \in E_i$  be independent random elements of  $E_i$  that satisfy the Centering Property, the Sub-Gaussian Property with constant  $K$ , and the Small Ball Property with constant  $c_0$ , each with respect to Bombieri-Weyl inner product. Let  $\gamma \geq 1$ ,  $d := \max_i d_i$ , and assume  $\alpha \leq \min\{d^{-8}, n^{-1}\}$ . Then for  $P = (p_1, \dots, p_{n-1})$  we have*

$$\text{Prob}(L(P) \leq \alpha) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + c\alpha^{\frac{1}{2}}\sqrt{n} \left( \frac{c_0 d^2 \gamma C}{\sigma_{\min}(E)\sqrt{n}} \right)^{n-1}$$

where  $C$  is a universal constant.

The proof of Theorem 4.3 is similar to a proof in our earlier paper [19]. We reproduce the proof here due to the importance of Theorem 4.3 in the flow of our current paper.

*Proof.* We assume the hypotheses of Assertion (1) in Lemma 4.1: Let  $\alpha, \gamma > 0$  and  $\beta \leq d^{-4}$ . Let  $\mathbf{B} := \{P \mid \|P\|_\infty \leq \gamma\}$  and let

$\mathbf{L} := \{P \mid L(P) \leq \alpha\} = \{P \mid \text{There exist } x, y \in S^{n-1} \text{ with } x \perp y \text{ and } \mathcal{L}(x, y) \leq \alpha\}$ .  
Let  $\Gamma := 8(\alpha^2 + (2 + e^4)\beta^4 d^4 \gamma^2)$  and let  $B_2^n$  denote the unit  $\ell_2$ -ball in  $\mathbb{R}^n$ .

Lemma 4.1 implies that if the event  $\mathbf{B} \cap \mathbf{L}$  occurs then there exists a non-empty set

$$V_{x,y} := \{w \in \mathbb{R}^n : w = x + \beta ry + \beta^2 z, x \perp y, |r| \leq 1, z \perp y, z \in B_2^n\} \setminus B_2^n$$

such that  $\|P(w)\|_2^2 \leq \Gamma$  for every  $w$  in this set. Let  $V := \text{Vol}(V_{x,y})$ . Note that for  $w \in V_{x,y}$  we have  $\|w\|_2^2 = \|x + \beta^2 z\|_2^2 + \|\beta y\|_2^2 \leq 1 + 4\beta^2$ . Hence we have  $\|w\|_2 \leq 1 + 2\beta^2$ . Since  $V_{x,y} \subseteq (1 + 2\beta^2)B_2^n \setminus B_2^n$ , we have showed that

$$\mathbf{B} \cap \mathbf{L} \subseteq \{P \mid \text{Vol}(\{x \in (1 + 2\beta^2)B_2^n \setminus B_2^n \mid \|P(x)\|_2^2 \leq \Gamma\}) \geq V\}.$$

Using Markov's Inequality, Fubini's Theorem, and Lemma 2.2, we can estimate the probability of this event. Indeed,

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

Now recall that  $\text{Vol}(B_2^n) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}$ . Then  $\frac{\text{Vol}(B_2^n)}{\text{Vol}(B_2^{n-1})} \leq \frac{c'}{\sqrt{n}}$  for some constant  $c' > 0$ . If we assume that  $\beta^2 \leq \frac{1}{2n}$ , then Lemma 4.2 implies  $(1 + 2\beta^2)^n \leq 1 + 6n\beta^2$ , and we obtain

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

for some absolute constant  $c > 0$ . Note that here, for a lower bound on  $V$ , we used the fact that  $V_{x,y}$  contains more than half of a cylinder with base having radius  $\beta^2$  and height  $2\beta$ .

Writing  $\tilde{x} := \frac{x}{\|x\|_2}$  for any  $x \neq 0$  we then obtain, for  $z \notin B_2^n$ , that

$$\|P(z)\|_2^2 = \sum_{j=1}^m |p_j(z)|^2 = \sum_{j=1}^m |p_j(\tilde{z})|^2 \|z\|_2^{2d_j} \geq \sum_{j=1}^m |p_j(\tilde{z})|^2 = \|P(\tilde{z})\|_2^2.$$

This implies, via Lemma 2.2, that for every  $w \in (1 + 2\beta^2)B_2^n \setminus B_2^n$  we have

$$\text{Prob}(\|P(w)\|_2^2 \leq \Gamma) \leq \text{Prob}(\|P(\tilde{w})\|_2^2 \leq \Gamma) \leq \left( cc_0 \sqrt{\frac{\Gamma}{n\sigma_{\min}(E)^2}} \right)^{n-1}.$$

So we conclude that  $\text{Prob}(\mathbf{B} \cap \mathbf{L}) \leq c\sqrt{n}\beta^{3-2n} \left( cc_0 \sqrt{\frac{\Gamma}{n\sigma_{\min}(E)^2}} \right)^{n-1}$ . Since  $\text{Prob}(L(P) \leq \alpha) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + \text{Prob}(\mathbf{B} \cap \mathbf{L})$  we then have

$$\text{Prob}(L(P) \leq \alpha) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + c\sqrt{n}\beta^{3-2n} \left( cc_0 \sqrt{\frac{\Gamma}{n\sigma_{\min}(E)^2}} \right)^{n-1}$$

Recall that  $\Gamma = 8(\alpha^2 + (5 + e^4)\beta^4 d^4 \gamma^2)$ . We set  $\beta^2 := \alpha$ . Our choice of  $\beta$  and the assumption that  $\gamma \geq 1$  then imply that  $\Gamma \leq C\alpha^2 \gamma^2 d^4$  for some constant  $C$ . So we obtain

$$\text{Prob}(L(P) \leq \alpha) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + c\sqrt{n}(\alpha)^{\frac{3}{2}-n} \left( \frac{c_0 C \alpha d^2 \gamma}{\sigma_{\min}(E) \sqrt{n}} \right)^{n-1}$$

$$\text{Prob}(L(P) \leq \alpha) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + c\sqrt{n}(\alpha)^{\frac{1}{2}} \left( \frac{c_0 d^2 \gamma C}{\sigma_{\min}(E) \sqrt{n}} \right)^{n-1}$$

and our proof is complete. ■

## 5. PROOF OF THEOREM 1.14

We first need to estimate Bombieri norm of a random polynomial system. The following lemma is more or less standard, and it follows from Lemma 2.3.

**Lemma 5.1.** *Let  $D = (d_1, \dots, d_{n-1})$  be a vector with positive integer coordinates, let  $E_i \subseteq H_{d_i}$  be full linear subspaces, and let  $E = (E_1, \dots, E_{n-1})$ . Let  $p_i \in E_i$  be random elements of  $E_i$  that satisfy the Centering Property and the Sub-Gaussian Property with constant  $K$ , each with respect to Bombieri-Weyl inner product. Then for all  $t \geq 1$ , we have*

$$\text{Prob} \left( \|p_i\|_W \geq t\sqrt{\dim(E_i)} \right) \leq \exp \left( 1 - \frac{t^2 \dim(E_i)}{K^2} \right)$$

and for the random polynomial system  $P = (p_1, \dots, p_{n-1})$  we have

$$\text{Prob} \left( \|P\|_W \geq t\sqrt{\dim(E)} \right) \leq \exp \left( 1 - \frac{t^2 \dim(E)}{K^2} \right). \blacksquare$$

Now we have all the necessary tools to prove our probabilistic condition number theorem. We will prove the following statement:

**Theorem 5.2.** *Let  $D = (d_1, \dots, d_{n-1})$  be a vector with positive integer coordinates, let  $E_i \subseteq H_{d_i}$  be non-degenerate linear subspaces, and let  $E = (E_1, \dots, E_{n-1})$ . We assume that  $\dim(E) \geq n \log(ed)$  and  $n \geq 3$ . Let  $p_i \in E_i$  be independent random elements of  $E_i$  that satisfy the Centering Property, the Sub-Gaussian Property with constant  $K$ , and the Small Ball Property with constant  $c_0$ , each with respect to the Bombieri-Weyl inner product. We set  $d := \max_i d_i$ , and*

$$M := nK\sqrt{\dim(E)}(c_0 d^2 C K \log(ed)^2 \sigma(E))^{2n-2}$$

where  $C \geq 4$  is a universal constant. Then for  $P = (p_1, \dots, p_{n-1})$ , we have

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

For notational simplicity we set  $m = \dim(E)$ . To start the proof we observe the following:

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \text{Prob}(\|P\|_W \geq sK\sqrt{m}) + \text{Prob}\left(L(P) \leq \frac{sK\sqrt{m}}{tM}\right)$$

The first probability on the right hand side will be controlled by Lemma 5.1, and the second will be controlled by Theorem 4.3. Theorem 4.3 states that for any  $\gamma \geq 1$  and for  $\frac{sK\sqrt{m}}{tM} \leq \min\{d^{-8}, n^{-1}\}$ , we have

$$\text{Prob}\left(L(P) \leq \frac{sK\sqrt{m}}{tM}\right) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + \left(\frac{sK\sqrt{m}}{tM}\right)^{\frac{1}{2}} \sqrt{n} \left(\frac{c_0 C \gamma d^2}{\sigma_{\min}(E)\sqrt{n}}\right)^{n-1}$$

To have  $\frac{sK\sqrt{m}}{tM} \leq \min\{d^{-8}, n^{-1}\}$  is equivalent to  $tM \min\{d^{-8}, n^{-1}\} \geq sK\sqrt{m}$ . We will check this condition at the end of the proof. Now, for  $\gamma = u\sigma_{\max}(E)\sqrt{n} \log(ed)K$  with  $u \geq 1$ , from Lemma 3.1 we have  $\text{Prob}(\|P\|_\infty \geq u\sigma_{\max}(E)\sqrt{n} \log(ed)K) \leq \exp(1 - a_3 u^2 n \log(ed)^2)$ . That is, for  $\gamma = u\sigma_{\max}(E)\sqrt{n} \log(ed)K$ , we have the following estimate:

$$\text{Prob}\left(L(P) \leq \frac{sK\sqrt{m}}{tM}\right) \leq \exp(1 - a_3 u^2 n \log(ed)^2) + \left(\frac{sK\sqrt{m}}{tM}\right)^{\frac{1}{2}} \sqrt{n} \left(\frac{c_0 C u \sigma_{\max}(E) \log(ed) d^2 K}{\sigma_{\min}(E)}\right)^{n-1}.$$

Since  $\sigma(E) = \frac{\sigma_{\max}(E)}{\sigma_{\min}(E)}$  and  $M = n\sqrt{n}K(c_0C \log(ed)d^2K\sigma(E))^{2n-2}$ , we have

$$\text{Prob}\left(L(P) \leq \frac{sK\sqrt{m}}{tM}\right) \leq \exp(1 - a_3u^2n \log(ed)) + \left(\frac{s}{t}\right)^{\frac{1}{2}} u^{n-1}.$$

Using Lemma 5.1 and the assumption that  $m \geq n \log(ed)$  we then obtain

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \exp(1 - s^2n \log(ed)^2) + \exp(1 - a_3u^2n \log(ed)) + \left(\frac{s}{t}\right)^{\frac{1}{2}} u^{n-1}.$$

If  $t \leq e^{2n \log(ed)}$  then setting  $s = u = 1$  gives the desired inequality. If  $t \geq e^{2n \log(ed)}$  then we set  $s = u^2 = \frac{\log(t)}{2\tilde{a}n \log(ed)}$ , where  $\tilde{a} > a_3 > 0$  is a constant greater than 1. We then obtain

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \exp\left(2 - \frac{1}{2} \log(t)\right) + \left(\frac{\log(t)}{2n \log(ed)}\right)^{\frac{n}{2}} \frac{1}{\sqrt{t}}.$$

Observe that  $\exp\left(2 - \frac{1}{2} \log t\right) = \frac{e^2}{\sqrt{t}}$ . So we have  $\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \left(\frac{\log(t)}{2n \log(ed)}\right)^{\frac{n}{2}} \frac{e^2+1}{\sqrt{t}}$ . To finalize our proof we need to check if  $tM \min\{d^{-8}, n^{-1}\} \geq sK\sqrt{m}$ . So we check the following:

$$tKn\sqrt{m}(c_0C \log(ed)d^2K\sigma(E))^{2n-2} \min\{d^{-8}, n^{-1}\} \stackrel{?}{\geq} \frac{\log(t)}{2n \log(ed)} K\sqrt{m}.$$

For  $n \geq 3$  we have  $(d^2 \log(ed))^{2n-2} > d^8$ . Since  $Kc_0 \geq \frac{1}{4}$ ,  $C \geq 4$ , and  $\sigma(E) \geq 1$ , we have

$$(c_0C \log(ed)d^2K\sigma(E))^{2n-2} > d^8.$$

Hence, it suffices to check if  $t \geq \frac{\log(t)}{2n \log(ed)}$ , which is clear.

We would like to complete the proof of Theorem 1.14 as it was stated in the introduction, for which the following easy observation suffices.

**Lemma 5.3.** *For  $t \geq e^{2n \log(ed)}$ , we have  $\left(\frac{\log(t)}{2n \log(ed)}\right)^{\frac{n}{2}} \leq t^{\frac{1}{4 \log(ed)}}$ .*

*Proof.* Let  $t = xe^{2n \log(ed)}$  where  $x \geq 1$ . Then

$$\left(\frac{\log(t)}{2n \log(ed)}\right)^{\frac{n}{2}} = \left(1 + \frac{\log(x)}{2n \log(ed)}\right)^{\frac{n}{2}} \leq e^{\frac{\log(x)}{4 \log(ed)}} = x^{\frac{1}{4 \log(ed)}}$$

Since  $x \leq t$ , we are done. ■

We now state the resulting bounds on the expectation of the condition number.

**Corollary 5.4.** *Under the assumptions of Theorem 5.2,  $0 < q < \frac{1}{2} - \frac{1}{2 \log(ed)}$  implies that  $\mathbb{E}(\tilde{\kappa}(P)^q) \leq M^q(1 + 4q \log(ed))$ . In particular,  $\mathbb{E} \log(\tilde{\kappa}(P)) \leq 1 + \log M$ .*

*Proof.* Observe that

$$\mathbb{E}(\tilde{\kappa}(P)^q) = M^q + qM^q \int_{t=1}^{\infty} \mathbb{P}\{\tilde{\kappa}(P) \geq tM\} t^{q-1} dt.$$

For  $t \geq e^{2n \log(ed)}$ , we have

$$\mathbb{P}\{\tilde{\kappa}(P) \geq tM\} t^{q-1} \leq t^{q + \frac{1}{4 \log(ed)} - \frac{3}{2}} \leq t^{-1 - \frac{1}{4 \log(ed)}}.$$

For  $t \leq e^{2n \log(ed)}$  we have even stronger tail bounds:

$$\mathbb{E}(\tilde{\kappa}(P)^q) \leq M^q \left( 1 + q \int_{t=1}^{\infty} t^{-1 - \frac{1}{4 \log(ed)}} dt \right).$$

This proves the first claim. The second claim follows by sending  $q \rightarrow 0$  and using Jensen's inequality. ■

## 6. PROOF OF THEOREM 1.16

Let  $E_i \subseteq H_{d_i}$  be non-degenerate linear spaces, and let  $E = (E_1, \dots, E_{n-1})$ . Suppose  $Q \in E$  is a fixed polynomial system. Let  $g_i \in E_i$  be independent random elements of  $E_i$  that satisfy the Centering Property, the Sub-Gaussian Property with constant  $K$ , and the Anti-Concentration Property with constant  $c_0$ , each with respect to the Bombieri-Weyl inner product. Let  $G := (g_1, \dots, g_{n-1})$  be the corresponding polynomial system. We define random perturbation of  $Q$  as follows:  $P := Q + G$ . We will use this notation for  $P$ ,  $Q$  and  $G$  for the rest of this section.

**Lemma 6.1.** *Let  $Q \in E$  be a polynomial system, let  $G$  be a random polynomial system in  $E$  that satisfies the Centering, sub-Gaussian, and Anti-Concentration hypotheses, and let  $P = Q + G$ . Then we have*

$$\text{Prob}(\|P\|_{\infty} \geq s\sigma_{\max}(E)\sqrt{n} \log(ed) + \|Q\|_{\infty}) \leq \exp\left(1 - \frac{a_3 s^2 n \log(ed)}{K^2}\right)$$

where  $a_3$  is an absolute constant.

*Proof.* The triangle inequality implies  $\|P\|_{\infty} \leq \|Q\|_{\infty} + \|G\|_{\infty}$ . We complete the proof by using Lemma 3.1 for the random system  $G$ . ■

**Lemma 6.2.** *Let  $Q \in E$  be a polynomial system, let  $G$  be a random polynomial system in  $E$  that satisfies the Centering, sub-Gaussian, and Anti-Concentration hypotheses, and let  $P = Q + G$ . Then, for all  $\varepsilon > 0$ , and for any  $w \in S^{n-1}$  we have*

$$\text{Prob}(\|P(w)\|_2 \leq \varepsilon \sigma_{\min}(E)\sqrt{n-1}) \leq (a_2 c_0 \varepsilon)^{n-1}$$

where  $a_2$  is an absolute constant.

*Proof.* By the Anti-Concentration Property, for all  $1 \leq i \leq n-1$ , we have

$$\text{Prob}\{|g_i(w) + q_i(w)| \leq c_0 \varepsilon \sigma_{\min}(E_i)\} \leq c_0 \varepsilon$$

We then use Lemma 2.4 with the random variables  $g_i(w) + q_i(w)$ . ■

**Lemma 6.3.** *Let  $Q \in E$  be a polynomial system, let  $G$  be a random polynomial system in  $E$  that satisfies the Centering, sub-Gaussian, and Anti-Concentration hypotheses, and let  $P = Q + G$ . Then for all  $t \geq 1$ , we have*

$$\text{Prob}\left(\|P\|_W \geq tK\sqrt{\dim(E)} + \|Q\|_W\right) \leq \exp(1 - t^2 m).$$

*Proof.* For all  $1 \leq i \leq n-1$ , by triangle inequality  $\|p_i\|_W \leq \|q_i\|_W + \|g_{C_i}\|_W$ . So using the first claim of Lemma 5.1 gives

$$\text{Prob}\left(\|p_i\|_W \geq t\sqrt{\dim(E_i)} + \|q_i\|_W\right) \leq \exp\left(1 - \frac{t^2 \dim(E_i)}{K^2}\right)$$

Note that  $\|P\|_W = \max_{\|w\|_2=1} \text{abs}(w, (p_1, \dots, p_{n-1}))$ . So proceeding as in the proof of Lemma 2.2 completes the proof. ■



**Theorem 6.4.** *Let  $Q \in E$  be a polynomial system, let  $G$  be a random polynomial system in  $E$  that satisfies the Centering, sub-Gaussian, and Anti-Concentration hypotheses, and let  $P = Q + G$ . Now let  $\gamma \geq 1$ ,  $d := \max_i d_i$ , and assume  $\alpha \leq \min\{d^{-8}, n^{-1}\}$ . Then*

$$\text{Prob}(L(P) \leq \alpha) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + c\alpha^{\frac{1}{2}}\sqrt{n} \left( \frac{c_0 d^2 \gamma C}{\sigma_{\min}(E)\sqrt{n}} \right)^{n-1}$$

where  $C$  is a universal constant.

The proof of Theorem 6.4 is identical to Theorem 4.3, so we skip it. Now we are ready to state main theorem of this section.

**Theorem 6.5.** *Let  $Q \in E$  be a polynomial system, let  $G$  be a random polynomial system in  $E$  that satisfies the Centering, sub-Gaussian, and Anti-Concentration hypotheses, and let  $P = Q + G$ . Also let  $d := \max_i d_i$ , and set*

$$M = nK \sqrt{\dim(E)} (c_0 d^2 C K \log(ed) \sigma(E))^{2n-2} \left( 1 + \frac{\|Q\|_W}{\sqrt{n} K \log(ed)} \right)^{2n-1}$$

where  $C \geq 4$  is a universal constant. Assume also that  $\dim(E) \geq n \log(ed)^2$  and  $n \geq 3$ . Then

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

*Proof.* We need a quick observation before we start our proof: For any  $Q \in E$  and  $w \in S^{n-1}$ , we have  $\|Q(w)\|_2^2 \leq \sum_{i=1}^{n-1} \|q_i\|_W^2 \sigma_{\max}(E_i)^2 \leq \|Q\|_W^2 \sigma_{\max}(E)^2$ . So we have

$$\|Q\|_\infty \leq \|Q\|_W \sigma_{\max}(E).$$

Using this upper bound on  $\|Q\|_\infty$  and the assumption that  $\dim(E) \geq n \log(ed)^2$ , we deduce  $M \geq nK \sqrt{\dim(E)} (c_0 d^2 C K \log(ed) \sigma(E))^{2n-2} \left( 1 + \frac{\|Q\|_W}{nK \sqrt{\dim(E)}} \right) \left( 1 + \frac{\|Q\|_\infty}{\sqrt{n} \log(ed) K \sigma_{\max}(E)} \right)^{2n-2}$ .

We will use this lower bound on  $M$  later in our proof. Now let  $m = \dim(E)$ . We start our proof with the following observation:

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \text{Prob}(\|P\|_W \geq sK\sqrt{m} + \|Q\|_W) + \text{Prob}\left(L(P) \leq \frac{sK\sqrt{m} + \|Q\|_W}{tM}\right).$$

Lemma 6.3 states that

$$\text{Prob}(\|P\|_W \geq sK\sqrt{m} + \|Q\|_W) \leq \exp(1 - s^2 m).$$

Theorem 6.4 states that for  $\frac{sK\sqrt{m} + \|Q\|_W}{tM} \leq \min\{d^{-8}, n^{-1}\}$  we have

$$\text{Prob}\left(L(P) \leq \frac{sK\sqrt{m} + \|Q\|_W}{tM}\right) \leq \text{Prob}(\|P\|_\infty \geq \gamma) + c \left( \frac{sK\sqrt{m} + \|Q\|_W}{tM} \right)^{\frac{1}{2}} \sqrt{n} \left( \frac{c_0 d^2 \gamma C}{\sigma_{\min}(E)\sqrt{n}} \right)^{n-1}.$$

We set  $\gamma = u\sigma_{\max}(E)\sqrt{n} \log(ed)K + \|Q\|_\infty$ . From Lemma 6.1, we have

$$\text{Prob}(\|P\|_\infty \geq u\sigma_{\max}(E)\sqrt{n} \log(ed)K + \|Q\|_\infty) \leq \exp(1 - a_3 u^2 n \log(ed)).$$

We also have

$$\left( \frac{c_0 d^2 \gamma C}{\sigma_{\min}(E)\sqrt{n}} \right)^{n-1} = (c_0 u d^2 C K \log(ed) \sigma(E))^{n-1} \left( 1 + \frac{\|Q\|_\infty}{u\sqrt{n} \log(ed) K \sigma_{\max}(E)} \right)^{n-1}.$$

Using  $u \geq 1$ ,  $s \geq 1$ ,  $m \geq n \log(ed)^2$ , and the lower obtained on  $M$ , we obtain

$$\text{Prob}(\tilde{\kappa}(P) \geq tM) \leq \exp(1 - s^2 n \log(ed)) + \exp(1 - a_3 u^2 n \log(ed)) + \left( \frac{s}{t} \right)^{\frac{1}{2}} u^{n-1}$$

The rest of the proof is identical to the proof of Theorem 5.2. ■

## 7. PROOF OF THEOREM 1.18

Define a random polynomial system  $F_\varepsilon = Q + G$  where  $G$  is Gaussian random polynomial system with  $K = \frac{\varepsilon \|Q\|_W}{\sqrt{n \log(ed)}}$  and  $c_0 K = \frac{1}{\sqrt{2\pi}}$ . Using Lemma 5.1 with  $t = 1$ , we have with probability at least  $1 - \exp(1 - \dim(E))$  that

$$\|F_\varepsilon - Q\|_W = \|G\|_W \leq \frac{\varepsilon \|Q\|_W \sqrt{\dim(E)}}{\sqrt{n \log(ed)}}.$$

For the condition estimate we will use Theorem 6.5: First note that with  $K = \frac{\varepsilon \|Q\|_W}{\sqrt{n \log(ed)}}$  and  $c_0 K = \frac{1}{\sqrt{2\pi}}$ , the quantity  $M$  in Theorem 6.5 is the following:

$$M = \frac{\varepsilon \sqrt{n} \sqrt{\dim(E)}}{\log(ed)} \left( \frac{d^2 C \log(ed) \sigma(E)}{\sqrt{2\pi}} \right)^{2n-2} \left( 1 + \frac{1}{\varepsilon} \right)^{2n-1}.$$

So we have  $M \leq 2\sqrt{n} \sqrt{\dim(E)} \left(\frac{2}{\varepsilon}\right)^{2n-2} \left(\frac{d^2 C \log(ed) \sigma(E)}{\sqrt{2\pi}}\right)^{2n-2}$ . Using Theorem 6.5 with  $t = 36$  we deduce that with probability greater than  $\frac{1}{2}$  we have

$$\tilde{\kappa}(F_\varepsilon) \leq 2\sqrt{n} \sqrt{\dim(E)} \left( \frac{d^2 C \log(ed) \sigma(E)}{\varepsilon} \right)^{2n-2}.$$

Since the union of the complement of these two events has measure less than  $\frac{1}{2} + \exp(1 - \dim(E))$ , their intersection has positive measure, and the proof is completed. ■

**Remark 7.1.** *The proof of Theorem 6.4 actually works for*

$$M = nK \sqrt{\dim(E)} (c_0 d^2 C K \log(ed) \sigma(E))^{2n-2} (1 + \|Q\|_W) \left( 1 + \frac{\|Q\|_\infty}{\sqrt{n \log(ed)} K \sigma_{\max}(E)} \right)^{2n-2},$$

*which is often much more smaller than the  $M$  used in the theorem statement.  $\diamond$*

## 8. APPENDIX A

In this part we address the question how big the dispersion constant is for a typical low-dimensional linear space. Imagine we fixed a dimension  $m \sim n \log(d)$  and considered subspaces of dimension  $m$  inside  $H_d$  (the vector space homogenous polynomials of degree  $d$ ). How does the dispersion constant vary among these subspaces? We know that some of these subspaces will be degenerate and will have infinite dispersion constant. Can we argue this high dispersion constant behaviour is rare?

What we do in this section is the following: we represent the space of  $m$ -dimensional linear subspaces of  $H_d$  by the *Grassmannian variety*,  $\text{Gr}(m, \dim(H_d))$ , and this space comes with a Haar measure. We will analyze the measure of the set of subspaces in  $\text{Gr}(m, \dim(H_d))$  that has high dispersion constant with respect to the Haar measure on  $\text{Gr}(m, \dim(H_d))$  (see Corollary 8.4 below).

We need to use the following notion from high-dimensional probability.

**Definition 8.1** (Gaussian Complexity). *Let  $X \subseteq \mathbb{R}^n$  be a set, then the Gaussian complexity of  $X$  denoted by  $\gamma(X)$  is defined as follows:*

$$\gamma(X) := \mathbb{E} \sup_{x \in X} |\langle G, x \rangle|$$

where  $G$  is distributed according to standard normal distribution  $\mathcal{N}(0, \mathbb{I})$  on  $\mathbb{R}^n$ .

The use of the term complexity in definition 8.1 might look unorthodox to readers with a computational complexity theory background. The rationale behind this standard terminology in high dimensional probability is that the Gaussian complexity of a set  $X$  is known to control the complexity of stochastic processes indexed on the set  $X$  (see for instance [45]).

A corollary of Lemma 2.1 and Lemma 2.8 is the following.

**Corollary 8.2** (Gaussian Complexity of the Veronese Embedding). *Let  $H_d$  be the vector space of degree  $d$  homogenous polynomials in  $n$  variables. Let  $u_i$   $i = 1, \dots, \binom{n+d-1}{d}$  be an orthonormal basis for the vector space  $H_d$  with respect to the Bombieri-Weyl norm. For every  $v \in S^{n-1}$ , we define the following polynomial  $q_v$ :*

$$q_v(x) := \sum_i u_i(v) u_i(x)$$

and the following set created out of  $q_v$ :

$$B_d := \{q_v : v \in S^{n-1}\}$$

Then, we have  $\gamma(B_d) \leq c\sqrt{n} \log(ed)$  for a universal constant  $c$ .

*Proof.* We need to consider a Gaussian element  $G$  in the vector space  $H_d$ . Note that for  $G \sim \mathcal{N}(0, \mathbb{I})$  in  $H_d$  we have  $\left\langle G, \sqrt{\binom{d}{\alpha}} x^\alpha \right\rangle_W \sim \mathcal{N}(0, 1)$  since  $\sqrt{\binom{d}{\alpha}} x^\alpha$  is an orthonormal basis with respect to the Weyl-Bombieri inner product. This means Gaussian elements of  $H_d$  are included in our model of randomness for the special case  $K = 1$ . Since  $\sigma_{\max}(H_d) = 1$ , Lemma 2.1 gives us the following estimate for pointwise evaluations of the Gaussian element  $G \sim \mathcal{N}(0, \mathbb{I})$  in  $H_d$ :

$$\text{Prob}\{|G(v)| \geq t\} \leq \exp\left(1 - \frac{t^2}{2}\right).$$

Note that  $\|G\|_\infty = \max_{v \in S^{n-1}} |G(v)| = \max_{q_v \in B_d} |\langle G, q_v \rangle|$ . So to estimate Gaussian complexity of the Veronese embedding  $B_d$ , we need to estimate  $\mathbb{E} \|G\|_\infty$ . Let  $\mathcal{N}$  be a  $\delta$ -net on the sphere  $S^{n-1}$ . Using a union bound, we then have

$$\text{Prob}\{\max_{v \in \mathcal{N}} |G(v)| \geq t\} \leq |\mathcal{N}| \exp\left(1 - \frac{t^2}{2}\right).$$

Setting  $\delta = \frac{1}{d}$  and using Lemma 2.8 for  $t \geq a_1 \sqrt{n} \log(ed)$  then gives the following:

$$\text{Prob}\{\|G\|_\infty \geq a_1 t \sqrt{n} \log(ed)\} \leq |\mathcal{N}| \exp\left(1 - \frac{a_1^2 t^2 n \log(ed)}{2}\right)$$

It is known that  $|\mathcal{N}| \leq \exp(a_0 n \log d)$ . So we have

$$|\mathcal{N}| \exp\left(1 - \frac{a_1^2 t^2 n \log(ed)}{2}\right) \leq \exp(1 - a_2 t^2 n \log(ed))$$

for some constant  $a_2$ . So  $\text{Prob}\{\|G\|_\infty \geq a_1 t \sqrt{n} \log(ed)\} \leq \exp(1 - a_2 t^2 n \log(ed))$ . Using this inequality one can routinely derive the estimate for  $\mathbb{E} \|G\|_\infty$ . ■

Since Talagrand proved his celebrated ‘‘majorizing measure theorem’’ (see [42]) it has been observed that for a set  $X$  and a random  $k \times n$  sub-Gaussian matrix  $A$  the deviation  $\sup_{x \in X} \|\|Ax\|_2 - \mathbb{E}\|Ax\|_2\|$  is controlled by  $\gamma(X)$ . There is a series of papers that established several variants of this fact, mainly in [41], [43] and more recently in [40] and [44]. Vershynin

has devoted the 9th chapter in his recent book [45] on these results and their applications. The version that we will use has been established in [43] but not stated explicitly. The authors devoted the last pages of their paper to present the proof. The following statement can be found in ([45] pages 221-222).

**Theorem 8.3.** *Let  $F$  be a random  $m$  dimensional subspace of  $\mathbb{R}^n$  drawn from Haar measure on  $\text{Gr}(m, n)$ , and let  $P_F$  be orthogonal projection map on  $F$ . Let  $X \subseteq \mathbb{R}^n$  be a set. Then there is a universal constant  $C$  such that*

$$\sup_{x \in X} \left| \sqrt{n} \|P_F(x)\| - \sqrt{m} \|x\| \right| \leq Ct\gamma(X)$$

with probability greater than  $1 - e^{-t^2}$ .

**Corollary 8.4.** *Let  $F$  be a random  $m$  dimensional subspace of  $H_d$  drawn from the Haar measure on  $\text{Gr}(m, \dim(H_d))$ , where  $m \geq 16Cn \log(ed)^2$ . Then*

$$\sigma(F) \leq \frac{\sqrt{m} + Ct\sqrt{n} \log(ed)}{\sqrt{m} - Ct\sqrt{n} \log(ed)}$$

with probability greater than  $1 - e^{-t^2}$ , where  $C$  is the absolute constant from Theorem 8.3.

*Proof.* Since  $\|q_v\|_W = 1$  for all  $v \in S^{n-1}$ , applying Theorem 8.3 to the set  $B_d$  implies that

$$\sup_{x \in B_d} \left| \binom{n+d-1}{d}^{\frac{1}{2}} \|\Pi_F(x)\| - \sqrt{m} \right| \leq Ct\sqrt{n} \log(ed)$$

with probability greater than  $1 - e^{-t^2}$  for all  $t \geq 1$ . Since  $\sigma_{\min}(F) = \min_{x \in B_d} \|\Pi_F(x)\|$  and  $\sigma_{\max}(F) = \max_{x \in B_d} \|\Pi_F(x)\|$ , we have

$$\frac{\sqrt{m} - Ct\sqrt{n} \log(ed)}{\binom{n+d-1}{d}^{\frac{1}{2}}} \leq \sigma_{\min}(F) \leq \sigma_{\max}(F) \leq \frac{\sqrt{m} + Ct\sqrt{n} \log(ed)}{\binom{n+d-1}{d}^{\frac{1}{2}}}$$

with probability greater than  $1 - e^{-t^2}$ . ■

## 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  $\psi_2$ -bodies,” in *Geometric Aspects of Functional Analysis*, Lecture Notes in Mathematics, vol. 1807, pp. 114–121, Springer Berlin Heidelberg, 2003.
- [5] Peter Bürgisser and Felipe Cucker, “On a problem posed by Steve Smale,” *Annals of Mathematics*, pp. 1785–1836, Vol. 174 (2011), no. 3.
- [6] Peter Bürgisser and Felipe Cucker, *Condition*, Grundlehren der mathematischen Wissenschaften, no. 349, Springer-Verlag, 2013.
- [7] Peter Bürgisser, Felipe Cucker, Pierre Lairez, *Computing the homology of basic semialgebraic sets in weakly exponential time*, *Journal of the ACM (JACM)* 66, 2018, pg 1–30
- [8] Peter Bürgisser, Felipe Cucker, Josué Tonelli-Cueto, *Computing the homology of semialgebraic sets. I: Lax formulas*, *Foundations of Computational Mathematics*, 2018
- [9] Peter Bürgisser, Felipe Cucker, Josué Tonelli-Cueto, *Computing the Homology of Semialgebraic Sets. II: General formulas*, arXiv preprint arXiv:1903.10710

- [10] Felipe Cucker, Alperen A. Ergür, Josué Tonelli-Cueto, *Plantinga-Vegter algorithm takes average polynomial time.*, In Proceedings of the 2019 on International Symposium on Symbolic and Algebraic Computation, pp. 114–121. 2019.
- [11] 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.
- [12] Felipe Cucker, “Approximate zeros and condition numbers,” *Journal of Complexity* 15 (1999), no. 2, pp. 214–226.
- [13] 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
- [14] 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.
- [15] 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.
- [16] Felipe Cucker; Teresa Krick; and Mike Shub, “Computing the Homology of Real Projective Sets,” *Foundations of Computational Mathematics*, Foundations of Computational Mathematics, 2018, pg 929–970
- [17] Jean-Piere Dedieu, Mike Shub, “Newton’s Method for Overdetermined Systems Of Equations,” *Math. Comp.* 69 (2000), no. 231, pp. 1099–1115.
- [18] James Demmel, Benjamin Diament, and Gregorio Malajovich, “On the Complexity of Computing Error Bounds,” *Found. Comput. Math.* pp. 101–125 (2001).
- [19] Alperen A. Ergür, Grigoris Paouris, J. Maurice Rojas, “Probabilistic Condition Number Estimates For Real Polynomial Systems I: A Broader Family Of Distributions”, *Foundations of Computational Mathematics*, to appear. DOI: 10.1007/s10208-018-9380-5.
- [20] Wassily Hoeffding, “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, 58 (301):13–30, 1963.
- [21] O. D. Kellog, “On bounded polynomials in several variables,” *Mathematische Zeitschrift*, December 1928, Volume 27, Issue 1, pp. 55–64.
- [22] Bo’az Klartag and Emanuel Milman, “Centroid bodies and the logarithmic Laplace Transform – a unified approach,” *J. Func. Anal.*, 262(1):10–34, 2012.
- [23] Bo’az Klartag and Shahar Mendelson, *Empirical processes and random projections*, *J. Functional Analysis*, Vol. 225, no. 1 (2005) 229–245.
- [24] Alexander Koldobsky and Alain Pajor, “A Remark on Measures of Sections of  $L_p$ -balls,” *Geometric Aspects of Functional Analysis*, *Lecture Notes in Mathematics* 2169, pp. 213–220, Springer-Verlag, 2017.
- [25] 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.
- [26] 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.
- [27] Michel Ledoux, *The Concentration of Measure Phenomemon*, *Mathematical Surveys & Monographs*, Book 89, AMS Press, 2005.
- [28] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin, “A simple tool for bounding the deviation of random matrices on geometric sets,” *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, *Lecture Notes in Mathematics* 2169, Springer, 2017, pp. 277–299.
- [29] G. Livshyts, Grigoris Paouris and P. Pivovarov, “Sharp bounds for marginal densities of product measures,” *Israel Journal of Mathematics*, Vol. 216, Issue 2, pp. 877–889.
- [30] Assaf Naor and Artem Zvavitch, “Isomorphic embedding of  $\ell_p^n$ ,  $1 < p < 2$ , into  $\ell_1^{(1+\varepsilon)n}$ ,” *Israel J. Math.* **122** (2001), pp. 371–380.
- [31] Hoi H. Nguyen, “On a condition number of general random polynomial systems,” *Mathematics of Computation* (2016) 85, pp. 737–757
- [32] Grigoris Paouris and Peter Pivovarov, “Randomized Isoperimetric Inequalities”, *IMA Volume ”Discrete Structures: Analysis and Applications”* (Springer)
- [33] Mark Rudelson and Roman Vershynin, “The Littlewood-Offord Problem and Invertibility of Random Matrices,” *Adv. Math.* 218 (2008), no. 2, pp. 600–633.

- [34] Mark Rudelson and Roman Vershynin, “*The Smallest Singular Value of Rectangular Matrix*,” *Communications on Pure and Applied Mathematics* 62 (2009), pp. 1707–1739.
- [35] 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.
- [36] Igor R. Shafarevich, *Basic Algebraic Geometry 1: Varieties in Projective Space*, 3rd edition, Springer-Verlag (2013).
- [37] Mike Shub and Steve Smale, “*Complexity of Bezout’s Theorem I. Geometric Aspects*,” *J. Amer. Math. Soc.* 6 (1993), no. 2, pp. 459–501.
- [38] Steve Smale, “*Mathematical Problems for the next Century*,” *The Mathematical Intelligencer*, 1998, 20.2, pp. 7–15.
- [39] D. A. Spielman and S.-H. Teng, “*Smoothed analysis of algorithms*,” *Proc. Int. Congress Math. (Beijing, 2002)*, Volume I, 2002, pp. 597–606.
- [40] S. Dirksen, *Tail bounds via generic chaining*, *Electron. J. Probab.* 20 (2015), no. 53, 29 pp.
- [41] G. Schechtman, *Two observations regarding embedding subsets of Euclidean spaces in normed spaces*, *Adv. Math.* 200 (2006), 125–135.
- [42] M. Talagrand, *The generic chaining. Upper and lower bounds of stochastic processes*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.
- [43] Bo’az Klartag and Shahar Mendelson, *Empirical processes and random projections*, *J. Funct. Anal.* Vol. 225, no. 1 (2005) 229–245.
- [44] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin, “*A simple tool for bounding the deviation of random matrices on geometric sets*,” *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016, Lecture Notes in Mathematics* 2169, Springer, 2017, pp. 277–299.
- [45] Roman Vershynin, “*High Dimensional Probability: An Introduction with Application in Data Science*”, *Cambridge Series in Statistical and Probabilistic Mathematics (Book 47)*, Cambridge University Press, 2018.
- [46] Roman Vershynin, “*Four Lectures on Probabilistic Methods for Data Science*”, available at <https://www.math.uci.edu/~rvershyn/papers/four-lectures-probability-data.pdf>
- [47] Roman Vershynin, “*Introduction to the Non-Asymptotic Analysis of Random Matrices*,” *Compressed sensing*, pp. 210–268, Cambridge Univ. Press, Cambridge, 2012.

SCHOOL OF COMPUTER SCIENCE, CARNEGIE MELLON UNIVERSITY, 5000 FORBES AVENUE, PITTSBURGH, PA, 15213

*Email address:* aergur@cs.cmu.edu

DEPARTMENT OF MATHEMATICS, TEXAS A&M UNIVERSITY TAMU 3368, COLLEGE STATION, TEXAS 77843-3368, USA.

*Email address:* grigoris@math.tamu.edu

DEPARTMENT OF MATHEMATICS, TEXAS A&M UNIVERSITY TAMU 3368, COLLEGE STATION, TEXAS 77843-3368, USA.

*Email address:* rojas@math.tamu.edu