THE EXPECTED NUMBER OF REAL ROOTS OF A MULTIHOMOGENEOUS SYSTEM OF POLYNOMIAL EQUATIONS

Andrew McLennan

Abstract. The methods of Shub and Smale [SS93] are extended to the class of multihomogeneous systems of polynomial equations, yielding Theorem 1, which is a formula expressing the mean (with respect to a particular distribution on the space of coefficient vectors) number of real roots as a multiple of the mean absolute value of the determinant of a random matrix. Theorem 2 derives closed form expressions for the mean in special cases that include: (a) Shub and Smale's result that the expected number of real roots of the general homogeneous system is the square root of the generic number of complex roots given by Bezout's theorem; (b) Rojas' [Roj96] characterization of the mean number of real roots of an "unmixed" multihomogeneous system. Theorem 3 gives upper and lower bounds for the mean number of roots, where the lower bound is the square root of the generic number of complex roots, as determined by Bernstein's [Ber75] theorem. These bounds are derived by induction from recursive inequalities given in Theorem 4.

Research at MSRI is supported in part by NSF grant DMS-9701755. I have benefited from numerous discussions with Maurice Rojas, and I am grateful for comments by seminar participants at the University of Minnesota, the AMS Meeting at Temple University in April 1998, and the conference on polynomial system solving at the Mathematical Sciences Research Institute in September 1998.

The Expected Number of Real Roots of a Multihomogeneous System of Polynomial Equations

1. Introduction

The study of the distribution of real roots of a polynomial with random coefficients, which traces back at least to [BP32], has recently been developed in the direction of multivariate systems. (This literature is ably surveyed, and extended, by Edelman and Kostlan [EK95].) Kostlan [Kos93] shows that, for a homogeneous polynomial equation of degree d in $n+1$ variables, a particular inner product on the space of coefficient vectors is distinguished by invariance under the natural action of $O(n+1)$ and orthogonality of monomials. He goes on to show that, for the system of n such equations, when the coefficient vectors for the various equations are independent random variables, with each one distributed according to the central normal distribution associated with this inner product, the mean number of projective roots in n-dimensional real projective space is $d^{n/2}$, which is the square root of the generic number of complex roots given by Bezout's theorem. Shub and Smale [SS93] extend this result to the general homogeneous system of n homogeneous polynomial equations of degrees $d_1, \ldots, d_n$, showing that the mean is $\sqrt{\prod_i d_i}$, which is again the square root of the Bezout number. Rojas [Roj96] studies unmixed[(1)] systems of multihomogeneous equations, arriving at a closed form formula for the mean number of roots in the cartesian product of projective spaces that is the natural root space for such systems.

This paper studies the more general case of mixed multihomogeneous systems. Theorem 1 is a formula expressing the mean number of real roots of a random multihomogeneous system as the product of the mean absolute value of the determinant of a random matrix times an expression composed of evaluations of Euler's function Γ at multiples of $1/2$. Theorem 2, which is a corollary, gives a closed form formula for this mean, for a smaller class of systems that includes both the general homogeneous system and the unmixed systems as special cases, so that the results of [SS93] and [Roj96] described above are corollaries. Theorem 3 generalizes the "square root of the Bezout number" result by giving upper and lower bounds on the mean

[(1)] Sparse systems of polynomial equations are described in Section 2. Roughly, such a system is unmixed if all polynomials have the same collection of monomials with nonzero coefficients, and otherwise it is mixed.

number of roots, where the lower bound is the square root of the maximal number of roots for the associated "demultihomogenized" system, as given by Bernshtein's [Ber75] extension of Bezout's theorem to sparse systems of polynomial equations. These bounds follow from recursive inequalities given in Theorem 4.

The author's interest in this topic is motivated in part by concepts of noncooperative game theory(2). The concept of a totally mixed Nash equilibrium for a normal form game amounts to a root, all of whose components must be positive, of particular sort of multihomogeneous system. McLennan and McKelvey [MM97] give a method for constructing normal form games that have as many regular (real) totally mixed Nash equilibria as are permitted by Bernshtein's theorem. The conceptual import of this result is that the maximal number of Nash equilibria is large, at least compared to most game theorists' prior intuition. Games that have the maximal number of equilibria are thought to be very atypical, and there arises the question of whether the set of equilibria is not only potentially large, but also large on average. McLennan [McL97] investigates the application, to this problem, of the results developed here, using Theorem 3 to show that the mean number of Nash equilibria can grow exponentially with various measures of the size of the game. Among other things, this analysis involves the extension of our work here to systems consisting of a multihomogeneous system of the sort studied here to which additional multihomogeneous polynomial inequalities have been appended, with the generalized formula being the one given here times a factor that may be regarded as the "probability" that a root of the system of equations also satisfies the inequalities.

In connection with speculation concerning whether analogues of Theorem 3 might hold for more general classes of sparse systems than the multihomogeneous ones, we recommend [Roj], which gives an extension to general sparse systems of the model of a random system studied here, and which presents results and conjectures along these lines. It is interesting to note that multihomogeneous systems are potentially special insofar as they can have as many real regular roots as are permitted by Bernshtein's theorem. (This is proved in [McL98] by pointing out that the argument

(2) This is not the place to give a general introduction to noncooperative game theory; Fudenberg and Tirole (1991) is a standard text. For the internal logic of this paper the description of *quasiequilibrium* (Section 2) is sufficient. For the connection between this notion and the standard concepts of *Nash equilibrium* and *totally mixed Nash equilibrium* see [MM97, McL97].

in [MM97], which establishes this claim for the systems arising in game theory, is actually valid for any multihomogeneous system.)

The proof of Theorem 1 parallels the analysis in [SS93] and [BCS98] rather closely, and is thus a descendant of the methods of [Kac43]. The *incidence variety* is the set of coefficient vector-root pairs. It is a submanifold of the cartesian product of the space of coefficient vectors and the root space, and the projection of it onto the root space is a fibration. The roots of the system at a particular coefficient vector are the preimages of the projection of the incidence variety onto the space of coefficient vectors, and an integral formula [SS93, p. 273; BCS98, p. 240] is used to reexpress the mean number of roots as a double integral, where the outer integral is over the root space and the inner integral is over the fibre of the projection onto the roots space at the root in question. Invariance is used to show that the inner integral does not depend on this root, so that the double integral is the volume of the root space times the inner integral, evaluated at a point in the root space which may be chosen at whim. For a particular choice it is possible to simplify the inner integral by transforming variables in a way that eliminates variables that do not enter the integrand, and from this Theorem 1 emerges.

The algorithms used by [MM97] to compute maximal numbers of Nash equilibria are based on recursive formulas for the Bernshtein number that extend directly to general multihomogeneous systems. Below (see also [McL98]) we describe how these formulas can be seen as the consequence of expressing the Bernshtein number for such a system as the permanent (e.g. [Ego96]) of a matrix, after which the recursions are obtained by expanding along a row or column. In investigating whether the mean number of real roots is greater than the square root of the Bernshtein number, as asserted by Theorem 3, it is natural to guess that the squares of the mean numbers of real roots obey the corresponding recursive inequalities, which is the assertion of Theorem 4, since then Theorem 3 follows from induction. Using Theorem 1, Proposition 7.1 restates these inequalities as recursive inequalities for the mean absoute values of the determinants of certain random matrices. The proof of Proposition 7.1 is, perhaps, rather surprising insofar as it depends on properties of normal random variables that seem quite distant from the geometric starting point of these investigations.

The remainder has the following organization. Section 2 describes multihomogeneous systems as a certain type of sparse system. Section 3 specifies an inner product on the space of coefficient vectors of a multihomogeneous equation that is uniquely characterized by invariance and orthogonality of monomials. The central normal distribution with respect to this inner product is our model of a random equation, and our random systems have the coefficient vectors of the various equations distributed independently according to these distributions. Section 4 states Theorem 1, and in Section 5 we discuss those systems for which it is possible to reduce the formula in Theorem 1 either to closed form or to an expression involving the formula applied to smaller systems. Section 6 defines mixed volume, states Bernshtein's theorem precisely, and shows how the generic number of complex roots of a multihomogeneous system may be computed recursively. Section 7 proves Theorems 3 and 4, and presents a result giving upper and lower bounds for the mean absolute value of the determinant of a random matrix. Sections 8–11 present the proof of Theorem 1.

2. Multihomogeneous Systems

In stating Bernshtein's theorem we will need to consider general sparse systems, so we describe multihomogeneous systems as a specialization of this concept. A *sparse system* of n polynomial equations in ℓ variables is

$$f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x})) = 0,$$

where $\mathbf{x} = (x_1, \dots, x_\ell)$ and, for each $i = 1, \dots, n$, there is a nonempty finite $\mathcal{A}_i \subset \mathbf{N}^\ell$ such that $f_i(\mathbf{x}) = \sum_{a \in \mathcal{A}_i} f_{ia} \mathbf{x}^a$ for some system of coefficients f_{ia}. (Here $\mathbf{x}^a$ denotes the monomial $x_1^{a_1} x_2^{a_2} \cdots x_\ell^{a_\ell}$.) The general approach of the theory of sparse systems is to hold the n-tuple of *supports* $(\mathcal{A}_1, \dots, \mathcal{A}_n)$ fixed while treating the coefficients f_{ia} as variables, for instance in the sense of studying properties that are generic in the space of vectors of coefficients. Such a system is said to be *unmixed* if $\mathcal{A}_1 = \dots = \mathcal{A}_n$; otherwise it is *mixed*. Identifying a polynomial with its vector of coefficients, we regard $\mathcal{H}_i := I\!R^{\mathcal{A}_i} \setminus \{0\}$ as the space of polynomials with real coefficients whose supports are nonempty subsets of $\mathcal{A}_i$. Let

$$\mathcal{H} := \mathcal{H}_1 \times \dots \times \mathcal{H}_n.$$

The system is *multihomogeneous* if the variables in $\mathbf{x}$ are divided into k groups, so that $\mathbf{x} = (\mathbf{y}_1, \dots, \mathbf{y}_k)$ where $\mathbf{y}_j = (y_{j0}, y_{j1}, \dots, y_{jn_j})$, and each equation is homogeneous of degree δ_{ij} as a function of $\mathbf{y}_j$, for any given values of the other variables $(\mathbf{y}_1, \dots, \mathbf{y}_{j-1}, \mathbf{y}_{j+1}, \dots, \mathbf{y}_k)$. More precisely, we require that there are nonnegative integers δ_{ij} ($i = 1, \dots, n$, $j = 1, \dots, k$) such that

$$\mathcal{A}_i = \mathcal{A}_{i1} \times \dots \times \mathcal{A}_{ik}, \quad \text{where} \quad \mathcal{A}_{ij} = \{\, \alpha \in \mathbf{N}^{n_j+1} : \alpha_0 + \alpha_1 + \dots + \alpha_{n_j} = \delta_{ij} \,\}.$$

When f_i is multihomogeneous, the truth value of the proposition '$f_i(\mathbf{x}) = 0$' is unaffected if each block of variables is multiplied by a nonzero scalar, so that, in effect, there are $\ell - k$ degrees of freedom. We work only with systems that are, in this sense, exactly determined: $\ell = n + k$, so that $n_1 + \ldots + n_k = n$. An instance of the type of system studied here is specified by the vector $\mathbf{n}$ and the $n \times k$ matrix $\delta := (\delta_{ij})$.

Four particular types of multihomogeneous system figure in our discussion:

(a) When $k = 1$ we have the general homogeneous system, for which the problem studied here was analyzed in [SS93]. In inductive constructions it will be convenient to allow the numbers of variables in some blocks to be zero, and we will use the phrase 'general homogeneous system' to describe any multihomogeneous system with $n_j = n$ for some j, in which case we must have $n_h = 0$ for all $h \neq j$.

(b) The *unmixed* multihomogeneous systems studied in [Roj96] are described by the condition that all equations have the same support: there are integers $e_1, \ldots, e_k$ such that

$$\delta_{1j} = \ldots = \delta_{nj} = e_j \quad (j = 1, \ldots, k).$$

(c) Generalizing (a) and (b) are the systems for which there are numbers $d_1, \ldots, d_n$ and $e_1, \ldots, e_k$ such that $\delta_{ij} = d_i e_j$ for all i and j.

(d) The systems arising, in game theory, from the concept of quasiequilibrium ([MM97]) of a finite normal form game, have, for each $j = 1, \ldots, k$, n_j equations that are homogeneous of degree one in $\mathbf{y}_h$ for all $h \neq j$, and are homogeneous of degree zero in $\mathbf{y}_j$. Formally these systems can be characterized as follows:

$$\delta_{ij} = \begin{cases} 0 & \text{if } q(i) = j, \\ 1 & \text{otherwise,} \end{cases} \tag{1}$$

where $q : \{1, \ldots, n\} \to \{1, \ldots, k\}$ is the function defined implicitly by the inequality

$$n_1 + \ldots + n_{q(i)-1} < i \leq n_1 + \ldots + n_{q(i)}.$$

3. An Invariant Inner Product

Fix a pair $(\mathbf{n}, \delta)$. Since $n_1 + \ldots + n_k = n$, we may index the components of an exponent vector $a \in \mathbf{N}^{n+k}$ by the pairs (j, h) for $j = 1, \ldots, k$ and $h = 0, \ldots, n_j$. For such an a let

$$\eta(a) := \frac{a_{10}! \cdot \ldots \cdot a_{1n_1}!}{(a_{10} + \ldots + a_{1n_1})!} \cdot \ldots \cdot \frac{a_{k0}! \cdot \ldots \cdot a_{kn_k}!}{(a_{k0} + \ldots + a_{kn_k})!}$$

$$= \binom{a_{10} + \ldots + a_{1n_1}}{a_{10}, \ldots, a_{1n_1}}^{-1} \cdot \ldots \cdot \binom{a_{k0} + \ldots + a_{kn_k}}{a_{k0}, \ldots, a_{kn_k}}^{-1}.$$

We endow each $\mathcal{H}_i$ with the inner product

$$\langle f_i, f_i' \rangle_i := \sum_{a \in \mathcal{A}_i} \eta(a) f_{ia} f_{ia}'.$$

Let $\| \cdot \|_i$ be the norm derived from $\langle \cdot, \cdot \rangle_i$.

Consider the product group

$$G := O(n_1 + 1) \times \ldots \times O(n_k + 1).$$

There is the obvious component-wise action of G on $I\!R^{n_1+1} \times \ldots \times I\!R^{n_k+1}$, and for $f_i \in \mathcal{H}_i$ and $O \in G$, $f_i \circ O^{-1}$ is easily seen to be a polynomial function that is multihomogeneous for the same numbers δ_{ij}, so $f_i \circ O^{-1}$ is an element of $\mathcal{H}_i$. Thus the formula $Of_i := f_i \circ O^{-1}$ defines an action from the left of G on $\mathcal{H}_i$. The following generalizes [Kos93, Th. 4.2], which is the case $k = 1$.

Lemma 3.1: *The inner product (4) is the unique (up to multiplication by a scalar) inner product on $\mathcal{H}_i$ that is invariant under the action of G and with respect to which the monomials are pairwise orthogonal.*

Proof: Let $\langle f_i, f_i' \rangle_i^* = \sum_{a \in \mathcal{A}_i} \eta^*(a) f_{ia} f_{ia}'$ be an invariant inner product with all monomials orthogonal. We wish to show that $\eta^*(a)/\eta^*(a') = \eta(a)/\eta(a')$ for all $a, a' \in \mathcal{A}_i$. Fixing arbitrary $a \in \mathcal{A}_i$ and $j = 1, \ldots, k$, it suffices to establish that this formula holds for those $a' \in \mathcal{A}_i$ with $a_{\ell h} = a'_{\ell h}$ whenever $\ell \neq j$, and this follows from [Kos93, Th. 4.2] applied to the subspace of $\mathcal{H}_i$ spanned by such a'.

To see that $\langle \cdot, \cdot \rangle_i$ is invariant under the action of G observe that, by [Kos93, Th. 4.2], it is invariant under the action of any group element g

with only one component g_j different from the identity in $O(n_j + 1)$, and that such group elements generate G. ■

Following [Kos93, EK95, Roj96], in our model of a random multi-homogeneous system the coefficient vectors of the various equations are statistically independent, with the coefficient vector of the i^{th} equation centrally normally distributed in $\mathcal{H}_i$ relative to $\langle \cdot, \cdot \rangle_i$. Concretely this means that the coefficients $\tilde{f}_{ia}$ are independent Gaussian random variables with mean 0 and variance $\eta(a)^{-1}$. In the setting of arbitrary sparse systems [Roj96] presents a definition and motivation of these variances that is geometric and general, in the sense that it pertains to any sparse system. Let μ_i be the probability measure on $\mathcal{H}_i$ that is the distribution of $\tilde{f}_i$, and let

$$\mu := \mu_1 \times \ldots \times \mu_n$$

be the distribution of $\tilde{f} := (\tilde{f}_1, \ldots, \tilde{f}_n)$.

In the calculations used to prove Theorem 1 we also consider the model in which the coefficient vectors $f_1, \ldots, f_n$ are statistically independent, with each f_i uniformly distributed in the unit sphere (relative to $\| \cdot \|_i$) of $\mathcal{H}_i$. The distribution of roots depends only on the distribution of the normalized coefficient vectors $f_i / \|f_i\|_i$, so standard facts concerning the multivariate normal distribution imply that, from our point of view, the two models are equivalent.

4. The Central Formula

We count roots in the k-fold product of projective spaces

$$P := P_1 \times \ldots \times P_k$$

where, for $j = 1, \ldots, k$, $P_j := \mathbf{P}^{n_j}(\mathbb{R})$ is n_j-dimensional real projective space. In the usual way, the equation $f_i(\zeta) = 0$ is meaningful for $f_i \in \mathcal{H}_i$ and $\zeta \in P$ even though f_i is not a function defined on P. Our central concern is the expected number of roots

$$E(\mathbf{n}, \delta) := \mathbf{E}\big(\#(\{\, \zeta \in P : \tilde{f}(\zeta) = 0 \,\})\big),$$

but in fact we completely characterize the distribution of roots.

Let $\tilde{Z}$ be a random $n \times n$ matrix with rows indexed by the integers $i = 1, \ldots, n$, columns indexed by the pairs jh for $j = 1, \ldots, k$ and

$h = 1, \dots, n_j$, and entries $\tilde{z}_i^{jh}$ that are independently distributed normal random variables with mean zero and variance δ_{ij}. Let $\Gamma(s) := \int_0^\infty \exp(-t)t^{s-1}\,dt$ be Euler's function.

Theorem 1:

(a)

$$E(\mathbf{n}, \delta) = 2^{-n/2} \cdot \Big(\prod_{j=1}^{k} \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n_j+1}{2})} \Big) \cdot \mathbf{E}\big(|\det \tilde{Z}|\big). \tag{2}$$

(b) The induced distribution of roots is uniform: for any open $W \subset P$,

$$\mathbf{E}\big(\#(\{\, \zeta \in W : \tilde{f}(\zeta) = 0 \,\})\big) = \frac{\mathrm{vol}(W)}{\mathrm{vol}(P)} E(\mathbf{n}, \delta).$$

This will be proved in Sections 8–11. The next three sections describe the consequences of this result.

5. Reduction to Closed Form

In certain circumstances the RHS of (2) can be reexpressed in closed form or in terms of the expressions derived from application of this formula to systems that are, in certain senses, smaller. Insofar as $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(s + 1) = s\Gamma(s)$ for all $s > 0$, the evaluations of Γ in (2) will be regarded as being in closed form already, so the problem is to reduce the term $\mathbf{E}\big(|\det \tilde{Z}|\big)$.

We begin by considering systems in which there is a subset of the variables that are determined by equations involving only those variables. Specifically, suppose there is some integer k' between 1 and k such that $\delta_{ij} = 0$ for all i, j such that $q(i) \le k'$ and $k' < j$, where $q(\cdot)$ is the function defined at the end of Section 2. Set $n' := n_1 + \ldots + n_{k'}$. Then

$$\delta = \begin{bmatrix} \delta^{11} & 0 \\ \delta^{21} & \delta^{22} \end{bmatrix}.$$

where δ^{11}, δ^{21}, and δ^{22} have dimensions $n' \times k'$, $(n-n') \times k'$, and $(n-n') \times (k - k')$ respectively. Then (with probability one) $\tilde{Z}$ has an $n' \times (n - n')$ block of zeros in its upper right corner, so its determinant is the product of the determinants of the $n' \times n'$ submatrix in the upper left and the $(n - n') \times (n - n')$ submatrix in the lower right. In particular, $\mathbf{E}\big(|\det \tilde{Z}|\big)$ does not depend on δ^{21}. Consequently (2) implies that $E(\mathbf{n}, \delta)$ is also independent of δ^{21}. When we set $\delta^{21} = 0$ we have a cartesian product of two independent systems, and our assumed distribution of coefficients for the combined system is the product measure of the assumed distributions for the subsystems. For any particular coefficient vector for the combined system, the number of roots is the product of the numbers of roots of the subsystems, so the following is a consequence of the fact that the mean of a product of independent random variables is the product of their means. Computationally, it follows immediately from the fact that the determinant of $\tilde{Z}$ is the product of the determinants of the submatrices.

Corollary 1: *Suppose there is some $1 \leq k' < k$ such that $\delta_{ij} = 0$ whenever $q(i) \leq k' < j$, and let δ^{11} and δ^{22} be as above. Then*

$$E(\mathbf{n}, \delta) = E((n_1, \ldots, n_{k'}), \delta^{11}) \cdot E((n_{k'+1}, \ldots, n_k), \delta^{22}).$$

A second general principle results from the effect on the determinant of multiplying a row or a column by a scalar.

Corollary 2: *If there are nonnegative integers $d_1, \ldots, d_n$ and $e_1, \ldots, e_k$ such that $\delta'_{ij} = d_i \cdot e_j \cdot \delta_{ij}$, then*

$$E(\mathbf{n}, \delta') = \sqrt{\prod_{i=1}^{n} d_i} \cdot \sqrt{\prod_{j=1}^{k} e_j^{n_j}} \cdot E(\mathbf{n}, \delta).$$

Consider now the particular case of $k = 1$ and $\delta_{11} = \ldots = \delta_{n1} = 1$. This corresponds to a system of n linear functionals in $n+1$ variables, and there is exactly one projective root for almost all coefficient vectors. In view of (2) we must have:

Proposition 5.1: *The mean absolute value of the determinant of a random $n \times n$ matrix whose entries are independently distributed normal random variables with mean zero and unit variance is*

$$2^{n/2} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})}.$$

Combining the last two results with Theorem 1 yields

Theorem 2: *If there are nonnegative integers $d_1, \ldots, d_n$ and $e_1, \ldots, e_k$ such that $\delta_{ij} = d_i \cdot e_j$, then*

$$E(\mathbf{n}, \delta) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})} \cdot \Big(\prod_{j=1}^{k} \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n_j+1}{2})} \Big) \cdot \sqrt{\prod_{i=1}^{n} d_i} \cdot \sqrt{\prod_{j=1}^{k} e_j^{n_j}}.$$

The Shub-Smale formula is the special case $k = 1$, and Rojas' formula for unmixed systems is obtained when $d_1 = \ldots = d_n = 1$.

There is a class of systems for which $E(\mathbf{n}, \delta)$ can be computed exactly by combining Corollaries 1 and 2 with Proposition 5.1. I know of no case outside this class in which the expectation $\mathbf{E}\big(|\det \tilde{Z}|\big)$ evaluates to a closed form expression. For the systems arising from normal form games we are able to evaluate in closed form only when $k = 2$, which corresponds to a game with two players. Applying ideas similar to those underlying Corollary 1 yields:

Corollary 3: *In the case of the game equilibrium system given by (1), if $k = 2$ then*

$$E(\mathbf{n}, \delta) = \begin{cases} 1 & \text{if } n_1 = n_2, \\ 0 & \text{otherwise.} \end{cases}$$

6. The BKK Bound for Multihomogeneous Systems

This section explains the consequences of Bernshtein's [Ber75] theorem for multihomogeneous systems. Let $f(\mathbf{z}) = (f_1(\mathbf{z}), \ldots, f_n(\mathbf{z}))$ be a general sparse system of n equations in the n variables $z_1, \ldots, z_n$, where f_i has support $\mathcal{A}_i \subset \mathbf{N}^n$. The *Newton polytope* of f_i is the convex polytope $Q_i = \mathrm{con}(\mathcal{A}_i)$. The *mixed volume* of $Q_1, \ldots, Q_n$, which was first defined and studied by Minkowski, and which we denote by $\mathcal{MV}(Q_1, \ldots, Q_n)$, may be defined to be the coefficient of the monomial $\lambda_1 \cdot \ldots \cdot \lambda_n$ in the polynomial[(3)] $\mathrm{vol}(Q_\lambda)$ where

$$Q_\lambda = \lambda_1 Q_1 + \ldots + \lambda_n Q_n.$$

Theorem: ([Ber75]) *Let $\mathbf{C}^* := \mathbf{C} \setminus \{0\}$. Let $\mathcal{H}^{\mathbf{C}} = \mathcal{H}_1^{\mathbf{C}} \times \ldots \times \mathcal{H}_n^{\mathbf{C}}$ where $\mathcal{H}_i^{\mathbf{C}} = \mathbf{C}^{\mathcal{A}_i}$ is the space of complex polynomials with support $\mathcal{A}_i$.*

(3) See [Ewa96] for a proof that $\mathrm{vol}(Q_\lambda)$ is, in fact, a polynomial function of λ.

For systems f in the complement, in $\mathcal{H}^{\mathbf{C}}$, of an algebraic set of positive (complex) codimension, there are $\mathcal{MV}(Q_1, \ldots, Q_n)$ roots in $(\mathbf{C}^*)^n$.

The maximal number $\mathcal{MV}(Q_1, \ldots, Q_n)$ of roots is often referred to as the "BKK bound" of the system in recognition of closely related work [Kus76, Kov78].

We apply this result to the "demultihomogenized" system obtained, from the given multihomogeneous system, by setting $y_{10} = \ldots = y_{k0} = 1$. In comparing the roots of the latter system, in $(\mathbf{C}^*)^n$, with the roots, in P, of the given multihomogeneous system, there is the possibility of roots in one of the coordinate subspaces (in the projective sense) along which one of the variables vanishes, but invariance under the action of G quickly implies that generic systems do not have such roots, or roots at projective infinity. Thus, generically, there is a one-to-one correspondence between the roots of the given multihomogeneous system and of the demultihomogenized system. The Newton polytope of the i^{th} demultihomogenized equation is $Q_i = \prod_{j:n_j>0} \delta_{ij}\Delta(n_j)$, where

$$\Delta(n_j) := \{\, (z_{j1}, \ldots, z_{jn_j}) \in \mathbb{R}^{n_j}_{\geq 0} : z_{j1} + \ldots + z_{jn_j} \leq 1 \,\},$$

and the generic number of complex roots of the system is

$$BKK(\mathbf{n}, \delta) := \mathcal{MV}\Big(\prod_{j:n_j>0} \delta_{1j}\Delta(n_j), \ldots, \prod_{j:n_j>0} \delta_{nj}\Delta(n_j)\Big).$$

Our analysis of this quantity employs the following concept. The *permanent* (e.g., [Ego96]) of an $m \times n$ matrix D with entries d_{ij} is

$$\operatorname{per} D := \sum_{\sigma \in S_{m,n}} \left(\prod_{i=1}^{n} d_{i\sigma(i)} \right)$$

where $S_{m,n}$ is the set of one to one functions from $\{1, \ldots, m\}$ to $\{1, \ldots, n\}$. Since there are no such functions when $m > n$, in which case $\operatorname{per} D$ is automatically zero. Note that multiplying any row of D by a scalar has the effect of multiplying the permanent by that scalar, and that we may expand by minors along any row: for each $i = 1, \ldots, m$

$$\operatorname{per} D = \sum_{j=1}^{n} d_{ij} \cdot \operatorname{per} D^{ij}$$

where D^{ij} is the $(m-1)\times(n-1)$ matrix obtained from D by eliminating the i^{th} row and the j^{th} column. When $m=n$ the permanent of D agrees with the permanent of its transpose, and these comments hold with rows and columns reversed.

Let $\Delta(\mathbf{n},\delta)$ be the $n\times n$ matrix whose first n_1 columns are the first column of δ, whose next n_2 columns are the second column of δ, and so forth. The computation

$$\operatorname{vol}\Big(\sum_{i=1}^{n}\lambda_i\Big(\prod_{j:n_j>0}\delta_{ij}\Delta(n_j)\Big)\Big)=\operatorname{vol}\Big(\prod_{j:n_j>0}\Big(\sum_{i=1}^{n}\lambda_i\delta_{ij}\Big)\Delta(n_j)\Big)$$

$$=\frac{\prod_{j:n_j>0}\Big(\sum_{i=1}^{n}\lambda_i\delta_{ij}\Big)^{n_j}}{n_1!\cdot\ldots\cdot n_k!}$$

has the following immediate implication:

Proposition 6.1: ([McL98])

$$BKK(\mathbf{n},\delta)=\frac{\operatorname{per}\Delta(\mathbf{n},\delta)}{n_1!\cdot\ldots\cdot n_k!}.$$

The next result enumerates consequences of the elementary properties of the permanent, applied to this result. For $i=1,\ldots,n$ let δ^{-i} be the $(n-1)\times k$ matrix obtained by eliminating the i^{th} row of δ. For $j=1,\ldots,k$ let $\mathbf{e}_j$ be the j^{th} standard unit basis vector of $\mathbb{R}^k$. In the recursive formulas below we are adopting the convention that

$$BKK((0,\ldots,0),\delta_0)=E((0,\ldots,0),\delta_0)=1,$$

where δ_0 is the $0\times k$ matrix. This means that the “null system” with no variables and no equations has one root.

Proposition 6.2: ([McL98])

(a) For all $i=1,\ldots,n$,

$$BKK(\mathbf{n},\delta)=\sum_{j:n_j>0}\delta_{ij}\cdot BKK(\mathbf{n}-\mathbf{e}_j,\delta^{-i}).$$

(b) For all $j=1,\ldots,k$ such that $n_j>0$,

$$BKK(\mathbf{n},\delta)=\frac{1}{n_j}\sum_{i=1}^{n}\delta_{ij}\cdot BKK(\mathbf{n}-\mathbf{e}_j,\delta^{-i}).$$

(c) *Suppose there is some* $1 \leq k' < k$ *such that* $\delta_{ij} = 0$ *whenever* $q(i) \leq k' < j$*, and let* δ^{11}*,* δ^{21}*, and* δ^{22} *be as in Section 4. Then*

$$BKK(\mathbf{n}, \delta) = BKK((n_1, \ldots, n_{k'}), \delta^{11}) \cdot BKK((n_{k'+1}, \ldots, n_k), \delta^{22}).$$

(d) *If there are nonnegative integers* $d_1, \ldots, d_n$ *and* $e_1, \ldots, e_k$ *such that* $\delta'_{ij} = d_i \cdot e_j \cdot \delta_{ij}$*, then*

$$BKK(\mathbf{n}, \delta') = \Big(\prod_{i=1}^{n} d_i\Big) \cdot \Big(\prod_{j:n_j>0} e_j^{n_j}\Big) \cdot BKK(\mathbf{n}, \delta).$$

The recursive formulas (a) and (b) give obvious algorithms for computing $BKK(\mathbf{n}, \delta)$ that have computed values of BKK on the order of 10^{21}. (Cf. [MM97].)

In preparation for Theorem 3, we ask when $BKK(\mathbf{n}, \delta)$ can be computed by repeated applications of (a) in which the RHS has only one nonzero term. We say that the pair $(\mathbf{n}, \delta)$ is *simply reducible* if the following inductive definition is satisfied: there is some i for which there is at most one j with $n_j > 0$, $\delta_{ij} > 0$, and $BKK(\mathbf{n} - \mathbf{e}_j, \delta^{-i}) > 0$, and if $n > 1$ we require that for this j, $(\mathbf{n} - \mathbf{e}_j, \delta^{-i})$ is also simply reducible. This will clearly be the case when repeated applications of (c) reduces $BKK(\mathbf{n}, \delta)$ to a product of instances of the general homogeneous system. In fact this is the only way that $(\mathbf{n}, \delta)$ can be simply reducible, as we shall see in the next section.

We will need the following technical result. Let A be an $m \times n$ matrix of 0's and 1's. We say that an $m \times n$ matrix $D = (d_ij)$ is *A-sparse* if $d_{ij} = 0$ whenever $a_{ij} = 0$.

Lemma 6.3: *The following conditions are equivalent:*

(i) *there is an integer* $1 \leq k < m$ *such that, after relabelling of rows and columns,* A *has a* $k \times (n + 1 - k)$ *block of 0's.*

(ii) $\operatorname{per} A = 0$*;*

(iii) *all* A*-sparse matrices have row rank less than* m*.*

Proof: Clearly (i) implies (ii). The meaning of (ii) is that for each one-to-one $\sigma : \{1, \ldots, m\} \to \{1, \ldots, n\}$ there is some i such that $a_{i\sigma(i)} = 0$, which implies that all A-sparse matrices have no $m \times m$ submatrices of full rank, so (ii) implies (iii). Assuming that (iii) holds, we may assume without loss of the generality that the first k rows of A are minimally linearly independent: for a generic A-sparse matrix D their span agrees with the span of any $(k-1)$-element subset. Reordering columns, we may assume that, for generic D, the projection of the span of the first k rows onto the space of the first $k-1$ columns has full rank. Now the upper right hand $k \times (n-(k-1))$ block of A must vanish, since otherwise it is straightforward to construct an A-sparse matrix D whose first k rows are linearly independent. ∎

7. The Mean Exceeds the Square Root of the Maximum

Let $\Delta_{\frac{1}{2}}(\mathbf{n}, \delta)$ be the $n \times n$ matrix whose (i, jh)-entry is $\sqrt{\delta_{ij}}$. This section establishes the following generalization of the Shub–Smale formula.

Theorem 3:

$$\frac{\operatorname{per} \Delta_{\frac{1}{2}}(\mathbf{n}, \delta)}{n_1! \cdot \ldots \cdot n_k!} \geq E(\mathbf{n}, \delta) \geq \sqrt{BKK(\mathbf{n}, \delta)} = \sqrt{\frac{\operatorname{per} \Delta(\mathbf{n}, \delta)}{n_1! \cdot \ldots \cdot n_k!}}.$$

These inequalities hold with equality when $(\mathbf{n}, \delta)$ is simply reducible and not otherwise.

Theorem 3 will follow by induction from the following stronger result.

Theorem 4: *For all $i = 1, \ldots, n$,*

$$\sum_{j : n_j > 0} \sqrt{\delta_{ij}} \cdot E(\mathbf{n} - \mathbf{e}_j, \delta^{-i}) \geq E(\mathbf{n}, \delta) \geq \sqrt{\sum_{j : n_j > 0} \delta_{ij} \cdot E(\mathbf{n} - \mathbf{e}_j, \delta^{-i})^2},$$

These inequalities hold with equality if and only if there is at most one j with $\delta_{ij} > 0$ and $E(\mathbf{n} - \mathbf{e}_j, \delta^{-i}) > 0$.

Proof of Theorem 3: The asserted inequalities follow from an induction on n that begins with the convention that $E(\mathbf{n}, \delta) = BKK(\mathbf{n}, \delta) = 1$ when $n_1 = \ldots = n_k = 0$. The induction step is a matter of comparing (a) of Proposition 6.2 and the analogous formula for $\operatorname{per} \Delta_{\frac{1}{2}}(\mathbf{n}, \delta)$ with the inequalities in Theorem 4. Moreover, Theorem 4 implies that either of the inequalities in Theorem 3 holds with equality if and only if there is at most one j with $E(\mathbf{n} - \mathbf{e}_j, \delta^{-i}) > 0$, and $E(\mathbf{n} - \mathbf{e}_j, \delta^{-i})$ also satisfies the inequality with equality. In particular, it follows from induction that $E(\mathbf{n}, \delta) > 0$ if and only if $BKK(\mathbf{n}, \delta) > 0$, so either of the inequalities in Theorem 3 holds with equality if and only if $(\mathbf{n}, \delta)$ is simply reducible. ∎

Remark: We can now give a direct characterization of simple reducibility. Applying Theorem 3 to the situation laid out in Corollary 1 and (c) of Proposition 6.2 shows that $(\mathbf{n}, \delta)$ is simply reducible if and only if both $((n_1, \ldots, n_{k'}), \delta^{11})$ and $((n_{k'+1}, \ldots, n_k), \delta^{22})$ are simply reducible. Thus it suffices to characterize simple reducibility when the hypotheses of (c) of Proposition 6.2 are not satisfied: there is no $1 \leq k' < k$ such that (after any reordering of rows and columns) $\delta_{ij} = 0$ whenever $q(i) \leq k' < j$. The inequality of Theorem 3 cannot hold with equality unless all instances of the inequality in Theorem 4 hold with equality, so we see that if $(\mathbf{n}, \delta)$ is simply reducible, then for *any* $i = 1, \ldots, n$ there is at most one j with $n_j > 0$, $\delta_{ij} > 0$, and $BKK(\mathbf{n} - \mathbf{e}_j, \delta^{-i}) > 0$, with $(\mathbf{n} - \mathbf{e}_j, \delta^{-i})$ simply reducible if $n > 1$. If there is some i for which there exist distinct j, j' with $n_j > 0$, $n_{j'} > 0$, $\delta_{ij} > 0$, and $\delta_{ij'} > 0$, then either $BKK(\mathbf{n} - \mathbf{e}_j, \delta^{-i}) = 0$ or $BKK(\mathbf{n} - \mathbf{e}_{j'}, \delta^{-i}) = 0$, in which case Proposition 6.1 and Lemma 6.3 imply that $\Delta(\mathbf{n} - \mathbf{e}_j, \delta^{-i})$ or $\Delta(\mathbf{n} - \mathbf{e}_{j'}, \delta^{-i})$ has a block of zeros, as per (iii) of Lemma 6.3, and this implies that the hypotheses of (c) of Proposition 6.2 are satisfied by $(\mathbf{n}, \delta)$, contrary to assumption. For each i there is consequently at most one j with $n_j > 0$ and $\delta_{ij} > 0$. If there is more than one j with $n_j > 0$ it is again easy to show that the hypotheses of (c) of Proposition 6.2 are satisfied by $(\mathbf{n}, \delta)$, so $n_j = n$ for some j. That is, we have the general homogeneous case.

It remains to prove Theorem 4. For the random matrix $\tilde{Z}$ of Theorem 1, let $\tilde{Z}^i_{jh}$ be the determinant of the $(n-1) \times (n-1)$ minor obtained by eliminating row i and column jh. Observe that, by Theorem 1,

$$\mathbf{E}\big(|\tilde{Z}^i_{jh}|\big) = 2^{\frac{n-1}{2}} \frac{\Gamma(\frac{n_j}{2})}{\Gamma(\frac{n_j+1}{2})} \Big(\prod_{p=1}^{k} \frac{\Gamma(\frac{n_p+1}{2})}{\Gamma(\frac{1}{2})}\Big) E(\mathbf{n} - \mathbf{e}_j, \delta^{-i}),$$

so, applying Theorem 1 again to express $E(\mathbf{n}, \delta)$ in terms of $\mathbf{E}\big(|\det \tilde{Z}|\big)$, we quickly find that the assertion of Theorem 4 is equivalent to:

Proposition 7.1: *For all $i = 1, \ldots, n$,*

$$\sum_{j: n_j > 0} \sqrt{\delta_{ij}} \frac{\Gamma(\frac{n_j+1}{2})}{\Gamma(\frac{n_j}{2})} \mathbf{E}\big(|\tilde{Z}^i_{j1}|\big) \geq \frac{\mathbf{E}\big(|\det \tilde{Z}|\big)}{\sqrt{2}} \geq \sqrt{\sum_{j: n_j > 0} \delta_{ij} \Big(\frac{\Gamma(\frac{n_j+1}{2})}{\Gamma(\frac{n_j}{2})}\Big)^2 \mathbf{E}\big(|\tilde{Z}^i_{j1}|\big)^2}.$$

These inequalities hold with equality if and only if $\delta_{ij} \mathbf{E}\big(|\tilde{Z}^i_{j1}|\big) > 0$ for at most one j.

The proof of this will be our goal for the remainder of the section. The next result describes the source of the inaccuracy of the approximation.

Lemma 7.2: *If $\tilde{x}$ is a $\mathbb{R}^m_{\geq 0}$–valued random variable for which $\mathbf{E}(\tilde{x})$ is defined, then*

$$\sum_{h=1}^{m} \mathbf{E}\big(|\tilde{x}_h|\big) \geq \mathbf{E}\big(\|\tilde{x}\|\big) \geq \big\|\mathbf{E}(\tilde{x})\big\|.$$

The first inequality holds with equality if and only if the support of the distribution of $\tilde{x}$ is contained in the union of the coordinate axes. The second inequality holds with equality if and only if the support of the distribution of $\tilde{x}$ is contained in a single ray emanating from the origin.

Proof: Since $\sum_h \mathbf{E}\big(|\tilde{x}_h|\big) = \mathbf{E}\big(\sum_h |\tilde{x}_h|\big)$, the first inequality follows from $\sum_{h=1}^m |\tilde{x}_h| \geq \|\tilde{x}\|$, and it holds with equality if and only if, with probability one, $\sum_{h=1}^m |\tilde{x}_h| = \|\tilde{x}\|$. The second inequality follows from Jensen's inequality, and it holds with equality if and only if $\|(1-\alpha)x_0 + \alpha x_1\| = (1-\alpha)\|x_0\| + \alpha\|x_1\|$ for any x_0, x_1 in the support of the distribution of $\tilde{x}$ and any $0 \leq \alpha \leq 1$. ∎

We will need the following technical fact.

Lemma 7.3: *Let $\tilde{\epsilon} = (\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_m)$ where $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_m$ are independent identically distributed normal random variables with mean zero and unit variance. Then*

$$\mathbf{E}\big(\|\tilde{\epsilon}\|\big) = \frac{\sqrt{2} \cdot \Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}. \tag{3}$$

Proof: We compute that

$$\begin{aligned}
\mathbf{E}\big(\|\tilde{\epsilon}\|\big) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \|x\| \cdot \Big(\frac{1}{\sqrt{2\pi}} e^{-x_1^2/2}\, dx_1\Big) \cdot \ldots \cdot \Big(\frac{1}{\sqrt{2\pi}} e^{-x_m^2/2}\, dx_m\Big) \\
&= (2\pi)^{-m/2} \int_{\mathbb{R}^m} \|x\| \cdot e^{-\|x\|^2/2}\, dx \\
&= (2\pi)^{-m/2} \int_0^{\infty} (re^{-r^2/2}) \cdot \mathrm{vol}(S^{m-1}) \cdot r^{m-1}\, dr.
\end{aligned}$$

The asserted formula is now obtained from the formula (e.g., [Fed69, p. 251])

$$\mathrm{vol}(S^{m-1}) = 2\frac{\Gamma(\frac{1}{2})^m}{\Gamma(\frac{m}{2})} \quad (m \geq 1) \tag{4}$$

the change of variables $t := r^2/2$, the fact that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and the definition of $\Gamma(\cdot)$. ∎

The next result expresses the central idea of the method, which exploits a special property of random normal variables, in its simplest form. Random matrices have been studied extensively [Gir90, Meh91, Mui82] but there seems to be little prior work on mean absolute values of random determinants.

Proposition 7.4: *Let $\tilde{E}$ be an $n \times n$ matrix whose entries $\tilde{\epsilon}_{ab}$ are independently distributed normal random variables with mean zero and variance σ_{ab}^2. For $1 \le a, b \le n$ let $\tilde{E}^{ab}$ be the determinant of the $(n-1) \times (n-1)$ minor of $\tilde{E}$ obtained by eliminating row a and column b. Then for any $a = 1, \ldots, n$:*

$$\sqrt{2/\pi} \cdot \sum_{b=1}^{n} \sigma_{ab} \cdot \mathbf{E}\big(|\tilde{E}^{ab}|\big) \;\ge\; \mathbf{E}\big(|\det \tilde{E}|\big) \;\ge\; \sqrt{2/\pi} \cdot \left(\sum_{b=1}^{n} \sigma_{ab}^2 \cdot \mathbf{E}\big(|\tilde{E}^{ab}|\big)^2\right)^{1/2}.$$

Proof: The expansion of the determinant by minors along row a is

$$\det \tilde{E} = \sum_{b=1}^{n} (-1)^{a+b} \tilde{\epsilon}_{ab} \tilde{E}^{ab}.$$

For any numbers $E^{a1}, \ldots, E^{an}$, elementary properties of Gaussian random variables imply that $\sum_{b=1}^{n} (-1)^{a+b} \tilde{\epsilon}_{ab} E^{ab}$ is a normally distributed random variable with mean 0 and variance $\sum_{b=1}^{n} \sigma_{ab}^2 (\tilde{E}^{ab})^2$. Since $(\tilde{\epsilon}_{a1}, \ldots, \tilde{\epsilon}_{an})$ and $(\tilde{E}^{a1}, \ldots, \tilde{E}^{an})$ are statistically independent, Fubini's theorem and (3) in the case $m = 1$ yield

$$\mathbf{E}\big(|\det \tilde{E}|\big) = \frac{\sqrt{2}}{\Gamma(\frac{1}{2})} \mathbf{E}\left(\sqrt{\sum_{b=1}^{n} \sigma_{ab}^2 (\tilde{E}^{ab})^2} \right).$$

Recalling that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, the claim follows from Lemma 7.2. ∎

Let Σ_1 be the $n \times n$ matrix with entries σ_{ab}, and let Σ_2 be the $n \times n$ matrix with entries σ_{ab}^2. By an induction on n we now have:

Corollary:

$$(2/\pi)^{n/2} \cdot \operatorname{per} \Sigma_1 \geq \mathbf{E}\big(|\det \tilde{E}|\big) \geq (2/\pi)^{n/2} \cdot \sqrt{\operatorname{per} \Sigma_2}.$$

The upper and lower bounds in Lemma 7.2 correspond to the extreme cases in which the distribution of $\tilde{x}$ is concentrated on the coordinate axes or on the ray through $\mathbf{E}\big(\tilde{x}\big)$. When the distribution of $\tilde{x}$ is known to be invariant under the action of a group, it can be possible to show that it is far from these extremes. In the specific case we have in mind the group

$$H = SO(\mathbb{R}^{n_1}) \times \ldots \times SO(\mathbb{R}^{n_k})$$

acts on the space of $n \times n$ matrices Z by simultaneously acting on each row of Z, where each row is viewed as an element of $\mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_k}$. Then, because the determinant of a linear transformation between inner product spaces is invariant under composition with orientation preserving orthogonal transformations of the domain or range, we have $\det(\eta Z) = \det Z$ for all $\eta \in H$ and all $n \times n$ matrices Z.

Let Z^i_{jh} denote the determinant of the $(n-1)\times(n-1)$ minor obtained from Z by eliminating row i and column jh. Define the function c_i from the space of $n \times n$ matrices to $\mathbb{R}^n$ by letting $c_i(Z)$ be the vector with components $c_i^{jh}(Z) = (-1)^{i+n_1+\ldots+n_{j-1}+h} Z^i_{jh}$. (Of course $c_i(Z)$ is independent of the i^{th} row of Z, and is called the *cross product* (cf. [Spi65], pp. 84-5) of the remaining $n-1$ rows.)

Lemma 7.5: *c_i is equivariant: $c_i(\eta Z) = \eta c_i(Z)$ for all $n \times n$ matrices Z and all $\eta \in H$.*

Proof: Let Z_i denote the i^{th} row of Z. Then for any $\eta \in H$ we have

$$Z_i \cdot c_i(Z) = \det Z = \det(\eta Z) = \eta Z_i \cdot c_i(\eta Z).$$

Since $c_i(Z)$ and $c_i(\eta Z)$ are independent of Z_i, and this holds for all Z_i, it must be the case that $c_i(\eta Z) = \eta c_i(Z)$ for all $n \times n$ matrices Z and all $\eta \in H$. ∎

Proof of 7.1: As in the last proof, we write $\det(\tilde{Z}) = \tilde{Z}_i \cdot c_i(\tilde{Z})$. As in the proof of Proposition 7.4, elementary properties of normal random variables and Fubini's theorem imply that

$$\mathbf{E}\big(|\det \tilde{Z}|\big) = \frac{\sqrt{2} \cdot \Gamma(1)}{\Gamma(\frac{1}{2})} \mathbf{E}\left(\sqrt{\sum_{j=1}^{k} \sum_{h=1}^{n_j} \delta_{ij} \cdot (\tilde{Z}^i_{jh})^2}\right)$$

$$= \frac{\sqrt{2}\cdot\Gamma(1)}{\Gamma(\frac{1}{2})}\mathbf{E}\Bigg(\sqrt{\sum_{j=1}^{k}\delta_{ij}\cdot\|\tilde{\xi}_{ij}\|^2}\Bigg).$$

Combining this with Lemma 7.2 yields

$$\frac{\sqrt{2}\cdot\Gamma(1)}{\Gamma(\frac{1}{2})}\sum_{j=1}^{k}\sqrt{\delta_{ij}}\mathbf{E}\big(\|\tilde{\xi}_{ij}\|\big) \;\geq\; \mathbf{E}\big(|\det\tilde{Z}|\big) \;\geq\; \frac{\sqrt{2}\cdot\Gamma(1)}{\Gamma(\frac{1}{2})}\sqrt{\sum_{j=1}^{k}\delta_{ij}\mathbf{E}\big(\|\tilde{\xi}_{ij}\|\big)^2}. \quad (5)$$

For $j=1,\dots,k$ let $\Pi_j : I\!R^n \to I\!R^{n_j}$ be the projection

$$\Pi_j(z^{11},\dots,z^{1n_1},\ \dots\ ,z^{k1},\dots,z^{kn_k}) = (z^{j1},\dots,z^{jn_j}).$$

Clearly Π_j is equivariant: $\Pi_j(\eta z) = \eta_j(\Pi_j(z))$ for all $z \in I\!R^n$ and $\eta = (\eta_1,\dots,\eta_k) \in H$. Therefore $\Pi_j \circ c_i$ is equivariant. By virtue of elementary properties of the multivariate normal, the distribution of the random matrix $\tilde{Z}$ on the space of $n\times n$ matrices is invariant under the action of H, so the distribution of $\tilde{\xi}_{ij} = \Pi_j(c_i(\tilde{Z}))$ is invariant under the action of $SO(I\!R^{n_j})$.

If $\tilde{x}$ is any $I\!R^{n_j}$–valued random variable whose distribution is invariant under the action of $O(I\!R^{n_j})$, the ratio $\mathbf{E}\big(|\tilde{x}_h|\big)/\mathbf{E}\big(\|\tilde{x}\|\big)$ must agree with the mean absolute value of the first component of a random vector that is uniformly distributed on the unit sphere in $I\!R^{n_j}$. In particular, by Lemma 7.3 we have

$$\frac{\mathbf{E}\big(\|\tilde{\xi}_{ij}\|\big)}{\mathbf{E}\big(\|\tilde{Z}_i^{j1}\|\big)} = \frac{\mathbf{E}\big(\|\tilde{\epsilon}\|\big)}{\mathbf{E}\big(|\tilde{\epsilon}_1|\big)} = \frac{\Gamma(\frac{n_j+1}{2})/\Gamma(\frac{n_j}{2})}{\Gamma(1)/\Gamma(\frac{1}{2})}$$

when $\tilde{\epsilon} = (\tilde{\epsilon}_1,\dots,\tilde{\epsilon}_{n_j})$ and $\tilde{\epsilon}_1,\dots,\tilde{\epsilon}_{n_j}$ are i.i.d. normal random variables with mean zero. The asserted inequality follows from substituing this into (5).

With respect to conditions under which the inequalities hold strictly, if the vector $(\|\tilde{\xi}_{i1}\|,\dots,\|\tilde{\xi}_{ik}\|)$ has two components that are nonzero with positive probability, then both inequalities in (5) hold strictly, by Lemma 7.2, and otherwise they do not. ∎

8. A Reformulation

This and the following three sections constitute the proof of Theorem 1. We now reformulate the problem in a framework that is more amenable to calculation, and which allows the application of the methods of [SS93] and [BCS98]. Let $M_i \subset \mathcal{H}_i$ be the unit sphere defined by $\langle \cdot, \cdot \rangle_i$, and let

$$M := M_1 \times \ldots \times M_n.$$

As a submanifold of $\mathcal{H}$, M inherits a measure corresponding to the intuitive notion of volume which we denote by $\mathrm{vol}(\cdot)$ or (when no confusion is possible) M. The *uniform distribution* on M is $\mathbf{U}_M(\cdot) := \mathrm{vol}(\cdot)/\mathrm{vol}(M)$. The analogous notation will occur in connection with other manifolds as well. The roots of $f \in \mathcal{H}$ depend only on $(f_1/\|f_1\|, \ldots, f_n/\|f_n\|)$, and the random system $(\tilde{f}_1/\|\tilde{f}_1\|, \ldots, \tilde{f}_n/\|\tilde{f}_n\|)$ is uniformly distributed in M, by virtue of standard facts concerning the multivariate normal distribution.

We regard P_j as the space of unordered pairs $[\zeta_j] = \{\zeta_j, -\zeta_j\}$ of antipodal points in N_j, where N_j is the unit sphere in $\mathbb{R}^{n_j+1}$. Let

$$N := N_1 \times \ldots \times N_k.$$

For each root $[\zeta] \in P$ of $f \in \mathcal{H}$ there are 2^k corresponding roots in N.

For each i let $\theta_i : N \to \mathcal{H}_i$ be the function with components $\theta_{ia}(\zeta) := \eta(a)^{-1}\zeta^a$. Let $F : M \times N \to \mathbb{R}^n$ be the evaluation map with components

$$F_i(f, \zeta) := f_i(\zeta) = \langle f_i, \theta_i(\zeta) \rangle_i.$$

The *incidence variety* is $V = F^{-1}(0)$. Let π_1 and π_2 be the projections from V to M and N respectively. We now have

$$E(\mathbf{n}, \delta) = 2^{-k} \int_M \#(\pi_1^{-1}(f))\, d\mathbf{U}_M. \tag{6}$$

In preparation for the result of the next section we discuss some technical matters.

Lemma 8.1: *Each θ_i is equivariant with respect to the actions of G on N and $\mathcal{H}_i$: $\theta_i(O\zeta) = O\theta_i(\zeta)$ for all $\zeta \in N$ and $O \in G$. The image of θ_i is contained in the unit sphere of $\mathcal{H}_i$.*

Proof: We have

$$\langle Of_i, O\theta_i(\zeta) \rangle_i = \langle f_i, \theta_i(\zeta) \rangle_i = f_i(\zeta) = f_i(O^{-1}(O\zeta)) = \langle Of_i, \theta_i(O\zeta) \rangle_i.$$

Here the first equality is the invariance established in Lemma 3.1, and the other three equalities are essentially matters of definition. For given ζ this holds for all f_i, so $\theta_i(O\zeta) = O\theta_i(\zeta)$. Consequently $\|\theta_i(O\zeta)\| = \|\theta_i(\zeta)\|$ for all ζ and O. Clearly $\theta_i(\zeta)$ is a standard basis vector of $\mathcal{H}_i$ if $\zeta_1, \dots, \zeta_k$ are all standard basis vectors in $I\!R^{n_1+1}, \dots, I\!R^{n_k+1}$ respectively, so the second claim follows from the fact that the action of G on N is transitive. ∎

The equation $f_i(\zeta) = 0$ means precisely that f_i and $\theta_i(\zeta)$ are orthogonal, so for $(f, \zeta) \in V$ we may construe $\theta_i(\zeta)$ as a tangent vector in $T_{f_i}M_i$, and clearly

$$\frac{\partial F_i}{\partial f}(f, \zeta)(0, \dots, \theta_{i'}(\zeta), \dots, 0)$$

is nonzero according to whether $i' = i$. Thus (f, ζ) is a regular point of F, and 0 is a regular value of F, so the regular value theorem (e.g., [GP65]) implies:

Lemma 8.2: *V is a C^∞ submanifold of $M \times N$ with $\dim V = \dim M$.*

Abusing notation, we let V_ζ denote both of the "fibers"

$$\pi_2^{-1}(\zeta) \subset V \subset M \times N \quad \text{and} \quad \{\, f \in M : (f, \zeta) \in \pi_2^{-1}(\zeta) \,\}$$

over a point $\zeta \in N$, with the appropriate interpretation to be inferred from context. For each i let $V_{\zeta,i}$ be the set of $f_i \in M_i$ with $f_i(\zeta) = 0$. As the intersection of M_i with a hyperplane, this set is a subsphere of M_i of codimension one. Thus $V_\zeta = V_{\zeta,1} \times \ldots \times V_{\zeta,n}$ has a simple topology that is independent of ζ, and, as one might expect:

Lemma 8.3: *$\pi_2 : V \to N$ is a C^∞ fibration.*

As usual, to argue this point in detail would be a longwinded and mundane affair, and we shall not do so. It is, perhaps, worth mentioning that the "group" of the fibration may be taken to be the group G introduced in Section 3, and that a suitable atlas of coordinate functions[(4)] is given by the following maps: given $\zeta_0 \in N$, a neighborhood $W \subset N$ of ζ_0, and a C^∞ map $h : W \to G$ satisfying $h(\zeta)\zeta_0 = \zeta$ for all $\zeta \in W$, let $\phi : V_{\zeta_0} \times W \to \pi_2^{-1}(W)$ be given by $\phi\big(f, \zeta\big) := \big(h(\zeta)f, \zeta\big)$.

[(4)] This terminology, and the definition of "fibration" we are appealing to, are from [Ste51, §2].

9. An Integral Formula

Sard's theorem implies that almost all points of M are regular values of π_1, so we need only consider such points in computing the average number of roots. Consider a regular point (f,ζ) of π_1. Since $T_{(f,\zeta)}V$ is mapped surjectively onto $T_f M$ by $D\pi_1(f,\zeta)$, the restriction of $DF(f,\zeta)$ to $T_\zeta N \subset T_{(f,\zeta)}(M \times N)$ must be nonsingular, else (f,ζ) would not be a regular point of F. The implicit function theorem implies that there is a neighborhood $U \subset M$ of f for which there is a smooth $G : U \to N$ with $G(f) = \zeta$ whose graph is contained in V. The *condition matrix* at (f,ζ) is the matrix of $DG(f)$ which, by the implicit function theorem, is

$$C(f,\zeta) := -\Big(\frac{\partial F}{\partial \zeta}(f,\zeta)\Big)^{-1}\frac{\partial F}{\partial f}(f,\zeta) : T_f M \to T_\zeta N.$$

This linear transformation gives a description of the way polynomial systems f are associated with roots near (f,ζ). Let $C^*(f,\zeta) : T_\zeta N \to T_f M$ be the adjoint of $C(f,\zeta)$.

Proposition 9.1: ([BCS98, p. 240]) *For any open $U \subset V$,*

$$\int_M \#(\pi_1^{-1}(f) \cap U)\, dM = \int_N \int_{V_y \cap U} \det\big(C(f,\zeta)C^*(f,\zeta)\big)^{-1/2} dV_\zeta dN.$$

Lemma 9.2: *If $(f,\zeta) \in V$ is a regular point of π_1, then*

$$\det\big(C(f,\zeta)C^*(f,\zeta)\big)^{-1/2} = |\det Df(\zeta)|.$$

Proof: For $v \in T_0\mathbb{R}^n$ and $\phi \in T_f M = T_{f_1}M_1 \times \ldots \times T_{f_n}M_n$ we compute that

$$\begin{aligned}\Big\langle \frac{\partial F}{\partial f}(f,\zeta)\phi, v\Big\rangle &= \Big\langle \big(\langle \phi_1, \theta_1(\zeta)\rangle_1, \ldots, \langle \phi_n, \theta_n(\zeta)\rangle_n\big), v\Big\rangle \\ &= \sum_{i=1}^n \langle \phi_i, v_i\theta_i(\zeta)\rangle_i = \Big\langle \phi, \big(v_1\theta_1(\zeta), \ldots, v_n\theta_n(\zeta)\big)\Big\rangle.\end{aligned}$$

This means precisely that the map $v \mapsto (v_1\theta_1(\zeta), \ldots, v_n\theta_n(\zeta))$ is the adjoint $\frac{\partial F}{\partial f}(f,\zeta)^*$ of $\frac{\partial F}{\partial f}(f,\zeta)$, and in particular $\frac{\partial F}{\partial f}(f,\zeta)\frac{\partial F}{\partial f}(f,\zeta)^*$ is the identity on $T_0\mathbb{R}^n$. Since the matrix of the adjoint of a linear transformation is the transpose of the transformation's matrix, substituting the definition of the condition matrix leads to

$$\det\big(C(f,\zeta)C^*(f,\zeta)\big)^{-1/2} = \left(\det\Big(\frac{\partial F}{\partial \zeta}(f,\zeta)^{-1}\big(\frac{\partial F}{\partial \zeta}(f,\zeta)^{-1}\big)^*\Big)\right)^{-1/2}$$

$$= \Big|\det \frac{\partial F}{\partial \zeta}(f,\zeta)\Big| = |\det Df(\zeta)|. \blacksquare$$

Combining the last two results, for any open $U \subset V$ we have

$$\int_{f\in\pi_1(U)} \#(\pi_1^{-1}(f))\, dM = \int_N \int_{V_\zeta \cap U} |\det Df(\zeta)|\; dV_\zeta dN. \tag{7}$$

10. Invariance

Combining the actions of G on the various $\mathcal{H}_i$ (recall Section 3) we obtain an action of G on $\mathcal{H}$ given by

$$Of := (f_1 \circ O^{-1}, \ldots, f_n \circ O^{-1}).$$

We will exploit this symmetry to further simplify the RHS of the formula above.

Each M_i is invariant under the action of G on $\mathcal{H}_i$, of course, so M is an invariant of the action of G on $\mathcal{H}$, and the restriction of this action to M is an action of G on M. Of course N is invariant under the usual action of G on $\prod_{j=1}^k I\!R^{n_j+1}$. Combining these actions, we derive an action of G on $M \times N$ given by $O(f,\zeta) := (Of, O\zeta)$. For any $O \in G$, $f \in M$, and $\zeta \in N$ we have $Of(O\zeta) = f \circ O^{-1}(O\zeta) = f(\zeta)$, so:

Lemma 10.1: *V is an invariant of the action of G on $M \times N$: $OV = V$ for all $O \in G$. Consequently (for either interpretation of the symbol V_ζ) $O(V_\zeta) = V_{O\zeta}$ for all ζ and O.*

Proposition 10.2: *The quantity $\int_{V_\zeta} |\det Df(\zeta)|\, dV_\zeta$ is independent of ζ.*

Proof: Observe that

$$D(Of)(O\zeta) = D(f \circ O^{-1})(O\zeta) = Df(\zeta) \circ O^{-1}$$

so that $|\det D(Of)(O\zeta)| = |\det Df(\zeta)|$. We now have the calculation that

$$\int_{V_{O\zeta}} |\det Df(O\zeta)|\, dV_{O\zeta} = \int_{V_\zeta} |\det D(Of)(O\zeta)|\, dV_\zeta = \int_{V_\zeta} |\det Df(\zeta)|\, dV_\zeta.$$

Here the first equality is an application of the change of variables formula with the change of variables function an isometry, so that the Jacobean is identically one. The claim now follows from the fact that the action of G on N is transitive. ∎

Applying this to (7), for any open $W \subset N$ and any $\zeta \in N$ we have

$$\int_M \#(\pi_1^{-1}(f) \cap \pi_2^{-1}(W))\, dM = \mathrm{vol}(W) \cdot \int_{V_\zeta} |\det Df(\zeta)|\, dV_\zeta. \tag{8}$$

Clearly (b) of Theorem 1 follows directly from this. The remaining task is to prove (a) of that result.

11. The Final Calculations

Fixing $\zeta \in N$, let $\tilde{f}_\zeta = (\tilde{f}_{\zeta,1}, \dots, \tilde{f}_{\zeta,n})$ be the orthogonal projection of $\tilde{f}$ onto the subspace of polynomial systems for which ζ is a root. For each i, $\|\tilde{f}_{\zeta,i}\|$ and $\tilde{f}_{\zeta,i}/\|\tilde{f}_{\zeta,i}\|$ are statistically independent, and the normalized vector is uniformly distributed in $V_{\zeta,i}$, so

$$\begin{aligned}\int_{\mathcal{H}} |\det D\tilde{f}_\zeta(\zeta)|\, d\mu &= \int_{\mathcal{H}} \Big(\prod_{i=1}^n \|\tilde{f}_{\zeta,i}\|\Big) \cdot \Big|\det D\Big(\frac{\tilde{f}_{\zeta,1}}{\|\tilde{f}_{\zeta,1}\|}, \dots, \frac{\tilde{f}_{\zeta,n}}{\|\tilde{f}_{\zeta,n}\|}\Big)(\zeta)\Big|\, d\mu \\ &= \Big(\prod_{i=1}^n \mathbf{E}\big(\|\tilde{f}_{\zeta,i}\|\big)\Big) \int_{V_\zeta} |\det Df(\zeta)|\, d\mathbf{U}_{V_\zeta}.\end{aligned}$$

Combining this with (6) and (8), we now obtain

$$E(\mathbf{n}, \delta) = 2^{-k} \frac{\mathrm{vol}(N) \cdot \mathrm{vol}(V_\zeta)}{\mathrm{vol}(M) \cdot \prod_{i=1}^n \mathbf{E}\big(\|\tilde{f}_{\zeta,i}\|\big)} \int_{\mathcal{H}} |\det D\tilde{f}_\zeta(\zeta)|\, d\mu.$$

The formula (4) for sphere volume gives

$$\mathrm{vol}(N_j) = 2\frac{\Gamma(\frac{1}{2})^{n_j+1}}{\Gamma(\frac{n_j+1}{2})}, \quad \mathrm{vol}(M_i) = 2\frac{\Gamma(\frac{1}{2})^{\dim \mathcal{H}_i}}{\Gamma(\frac{\dim \mathcal{H}_i}{2})}, \quad \mathrm{vol}(V_{\zeta,i}) = 2\frac{\Gamma(\frac{1}{2})^{\dim \mathcal{H}_i - 1}}{\Gamma(\frac{\dim \mathcal{H}_i - 1}{2})},$$

and Lemma 7.3 yields

$$\mathbf{E}\big(\|\tilde{f}_{\zeta,i}\|\big) = \frac{\sqrt{2}\cdot\Gamma(\frac{\dim\mathcal{H}_i}{2})}{\Gamma(\frac{\dim\mathcal{H}_i - 1}{2})}.$$

Since $\mathrm{vol}(M) = \mathrm{vol}(M_1) \times \ldots \times \mathrm{vol}(M_n)$, and similarly for N and V_ζ, we now have

$$E(\mathbf{n},\delta) = 2^{-n/2}\cdot\Big(\prod_{j=1}^{k}\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{n_j+1}{2})}\Big)\cdot\int_{\mathcal{H}}|\det D\tilde{f}_\zeta(\zeta)|\,d\mu. \tag{9}$$

In the further evaluation of this quantity we are free to let ζ be any convenient point in N. We will compute at $\zeta_0 = (\mathbf{e}_{10},\ldots,\mathbf{e}_{k0}) \in N$ where, for $1 \le j \le k$, $\mathbf{e}_{j0},\mathbf{e}_{j1},\ldots,\mathbf{e}_{jn_j}$ are the standard unit basis vectors of $\mathbb{R}^{n_j+1}$. For each i and j let $a^0_{ij} = (\delta_{ij},0,\ldots,0) \in \mathcal{A}_{ij}$, and for each i let $a^0_i = (a^0_{i1},\ldots,a^0_{ik}) \in \mathcal{A}_i$. Since $\zeta_0^a = 0$ for all $a \in \mathcal{A}_i$ other than a^0_i, and $\zeta_0^{a^0_i} = 1$, for each i $V_{\zeta_0,i} := \{\, f_i \in M_i : f_{ia^0_i} = 0\,\}$. For i and j such that $\delta_{ij} > 0$ and each $h = 1,\ldots,n_j$, let a^{jh}_i be a^0_i with a^0_{ij} replaced by $(\delta_{ij}-1,0,\ldots,0,1,0,\ldots,0)$ (the '1' is component h). Then $T_{\zeta_0}N$ is spanned by the n vectors

$$\mathbf{b}_{jh} := (0,\ldots,\mathbf{e}_{jh},\ldots,0) \qquad (1 \le j \le k, 1 \le h \le n_j),$$

and elementary calculus yields

$$D\tilde{f}_{\zeta_0,i}(\zeta_0)\mathbf{b}_{jh} = \begin{cases} \tilde{f}_{ia^{jh}_i} & \text{if } \delta_{ij} > 0, \\ 0 & \text{if } \delta_{ij} = 0. \end{cases}$$

In this way we obtain a description of $D\tilde{f}_{\zeta_0}(\zeta_0)$ as an $n \times n$ matrix with rows indexed by $f_1,\ldots,f_n$, columns indexed by the pairs (j,h), and this (i,jh)–entry. Recalling from Section 3 that the variance of $\tilde{f}_{ia^{jh}_i}$ is δ_{ij}, we see that the matrix of $D\tilde{f}_{\zeta_0}(\zeta_0)$ has the same distribution as $\tilde{Z}$. In view of (9) this observation completes the proof of Theorem 1.

References

[Ber75] D. N. Bernshtein, The number of roots of a system of equations, *Functional Analysis and its Applications* **9** (1975), 183–185.

[BCS93] L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and Real Computation*, Springer-Verlag, New York, (1998).

[BP32] A. Bloch and G. Pólya, On the roots of a certain algebraic equation, *Proc. London Math. Soc.* **33** (1932), 102–114.

[EK95] A. Edelman and E. Kostlan, How many zeros of a random polynomial are real?, *Bulletin of the American Mathematical Society* **32** (1995), 1–37.

[Ego96] G. P. Egorychev, Van der Waerden conjecture and applications, *Handbook of Algebra*, Vol. I, Elsevier, Amsterdam (1996), 3–26.

[Ewa96] G. Ewald, *Combinatorial Convexity and Algebraic Geometry*, Springer, New York, (1996).

[Fed69] H. Federer, *Geometric Measure Theory*, Springer, New York, (1969).

[FT91] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, Cambridge, (1991).

[Gir90] V.L. Girko, *The Theory of Random Determinants*, Kluwer, Boston, (1990).

[GP65] V. Guillemin and A. Pollack, *Differential Topology*, Prentice-Hall, Englewood Cliffs, (1965).

[Kac43] M. Kac, On the average number of real roots of a random algebraic equation, *Bulletin of the American Mathematical Society* **49** (1943), 314–320 and 938.

[Kho78] A. G. Khovanskii, Newton polyhedra and the genus of complete intersections, *Functional Analysis and its Applications* **12** (1978), 51–61.

[Kos93] E. Kostlan, On the distribution of roots of random polynomials, *From Topology to Computation: Proceedings of the Smalefest*, Hirsch, M., Marsden, J., and Shub, M. (eds) (1993).

[Kus75] A. G. Kushnirenko, The Newton polyhedron and the number of solution of a system of k equations in k unknowns, *Upsekhi Mat. Nauk.* **30** (1975), 266–267.

[MM97] R. D. McKelvey and A. McLennan, The maximal number of regular totally mixed Nash equilibria, *Journal of Economic Theory* **72** (1997), 411–425.

[McL97] A. McLennan, On the Expected Number of Nash Equilibria of a Normal Form Game, mimeo, University of Minnesota, (1997).

[McL98] A. McLennan, The maximal number of real roots of a multihomogeneous system of polynomial equations, forthcoming in *Beiträge zur Algebra und Geometrie*, (1998).

[Meh91] M.L. Mehta, *Random Matrices*, Academic Press, New York (1991).

[Mui82] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York (1982).

[Roj96] J. M. Rojas, On the average number of real roots of certain random sparse polynomial systems, *Lectures on Applied Mathematics Series*, ed. by J. Renegar, M. Shub, and S. Smale, American Mathematical Society, (1996).

[SS93] M. Shub and S. Smale, Complexity of Bezout's theorem II: volumes and probabilities, *Computational Algebraic Geometry* (F. Eyssette and A. Galligo, eds.), Progr. Math., vol. 109 (1993), Birkhauser, Boston, 267–285.

[Spi65] M. Spivak, *Calculus on Manifolds : a Modern Approach to Classical Theorems of Advanced Calculus*, Benjamin, New York, (1965).

[Ste51] N. Steenrod, *The Topology of Fibre Bundles*, Princeton University Press, Princeton (1951).

Andrew McLennan
Department of Economics
University of Minnesota
271 19th Ave. S.
Minneapolis, MN 55455, USA

mclennan@atlas.socsci.umn.edu
http://www.econ.umn.edu/~mclennan