Polynomial Systems and the Momentum Map

Gregorio Malajovich*† J. Maurice Rojas‡ §

June 18, 2001

To Steve Smale on his $70^{\underline{\text{th}}}$ birthday.

Keywords: mixed volume, condition number, polynomial systems, sparse, random.
2000 Math Subject Classification: 65H10, 52A39.

1 Introduction

This paper outlines a Kähler-geometric approach to the study of random sparse polynomial systems, with a view toward understanding the distribution of solutions and their numerical conditioning. Section 2 develops the necessary mathematical framework while section 3 contains our main results for random sparse polynomial systems, including:

1. a new formula, in terms of a particular differential form, for the average number of complex roots in any measurable region
2. new bounds relating numerical conditioning of any given sparse polynomial system to the nearest ill-posed problem.

However, let us first give a simple preliminary illustration of some of the ideas we will develop.

*Departamento de Matemática Aplicada, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, CEP 21945–970, Rio de Janeiro, RJ, Brasil. `http://www.labma.ufrj.br/~gregorio` . **e-mail:** `gregorio@labma.ufrj.br` . On leave at the Department of Mathematics, City University of Hong Kong.

†G.M.'s visit to City University of Hong Kong was supported by CERG grants 9040393, 9040402, and 9040188

‡Department of Mathematics, City University of Hong Kong and Department of Mathematics, Texas A&M University, College Station, Texas 77843-3368, USA. `http://math.tamu.edu/~rojas` . **e-mail:** `rojas@math.tamu.edu` .

§Partially supported by Hong Kong UGC grant #9040402 and a US National Science Foundation Mathematical Sciences Postdoctoral Fellowship.

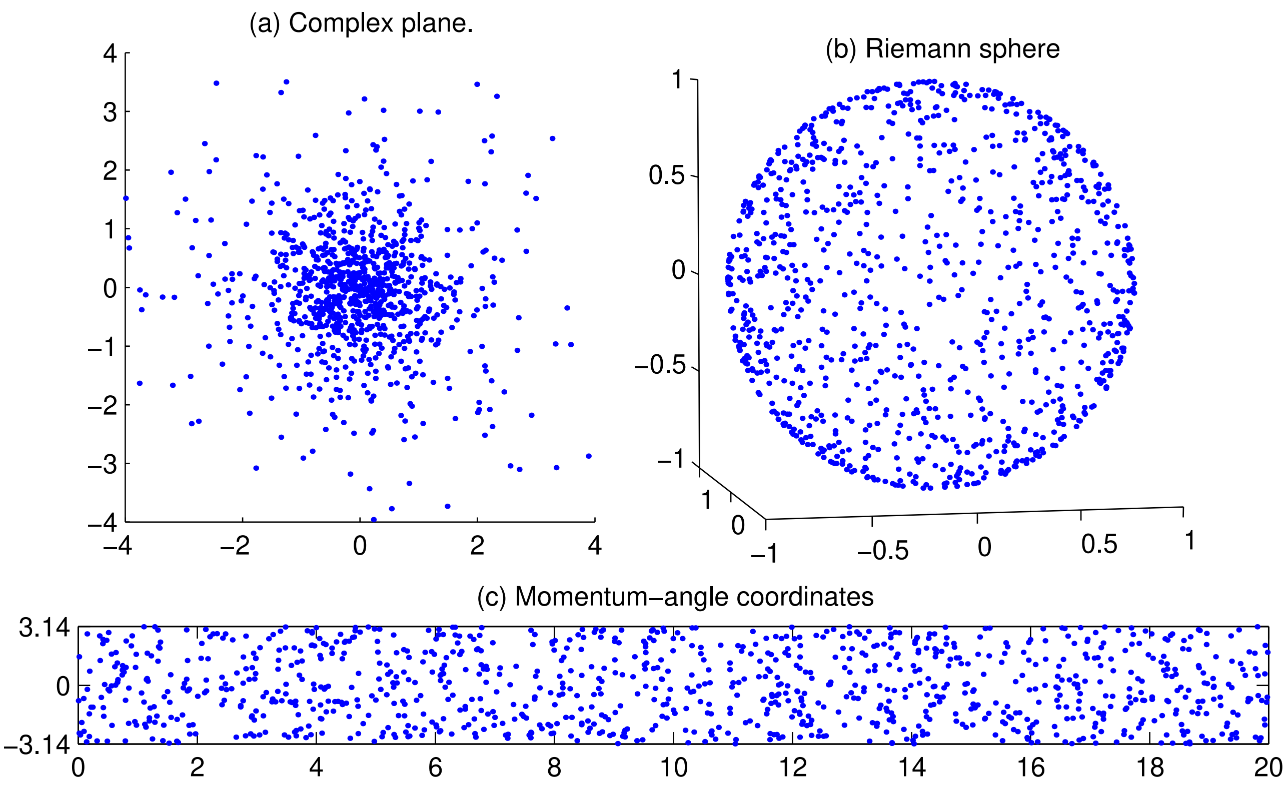

Figure 1: Root distribution when coefficient vector has a unitarily invariant Gaussian distribution

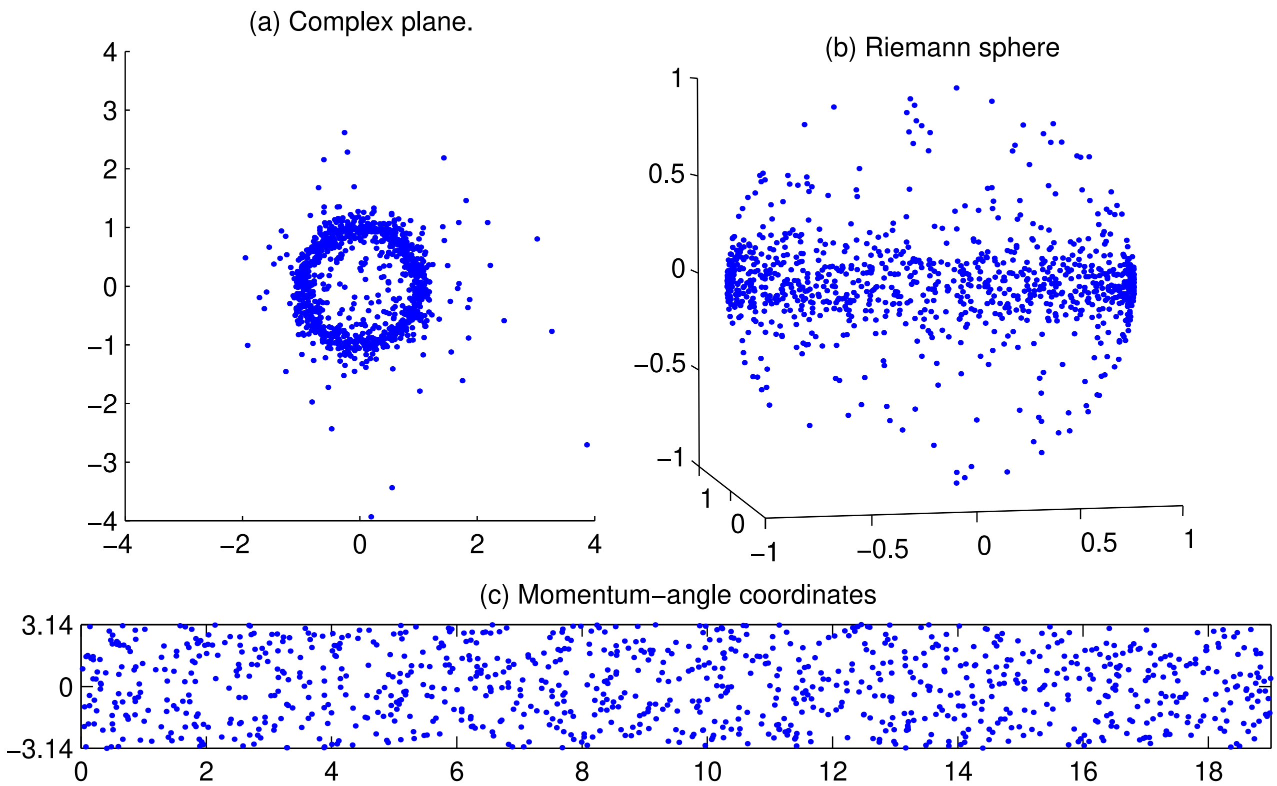

Figure 2: Root distribution when the coefficients are identically and independently distributed Gaussians

Figures 1(a) through 2(c) feature one random root of each of 1000 random degree 20 univariate polynomials. The j–th coefficient of each polynomial is a pseudo-random Gaussian variable of variance $\left(\begin{array}{c} d \\ j \end{array}\right)$ in Figure 1(a–c), and of variance 1 in Figure 2(a–c). The first distribution is invariant under a naturally defined action of the 2×2 unitary matrix group on the complex projective line $\mathbb{P}^1$: the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ sends the point with projective coordinates $[x_0 : x_1]$ to $[ax_0 + bx_1 : cx_0 + dx_1]$. The second distribution lacks this "unitary" invariance.

The roots are plotted in three different "phase spaces". In Figures 1(a) and 2(a), the roots are plotted in the complex plane. We notice that in figure 2(a), the roots seem to accumulate on the unit circle. This is a well-known phenomenon which can be made more rigorous.

In Figures 1(b) and 2(b), we show the same roots as points in the Riemann Sphere. We can see that the selected roots of the unitarily invariant distributed polynomials are uniformly distributed on the Riemann Sphere, while those of the variance-one polynomials are not (this shows graphically that the distribution of the roots is not unitarily invariant).

There is strong evidence in [Kos] and [BCSS98] suggesting that unitarily invariant distribution would be the natural distribution for Gaussian random polynomials. While the unitarily invariant paradigm was fundamental to the development in [SS93b, SS93a, SS93c, SS96, SS94], we will see here that a more general approach is viable and fruitful.

In figures 1(c) and 2(c), the same roots are plotted in a different "phase space", that we will construct below. The distribution seems uniform, it is indeed uniform. This suggests that there is another approach to random polynomials, that allows arbitrary variances through different coordinate systems.

This approach extends to systems of polynomial equations. While the unitary-invariant paradigma precluded the treatment of sparse polynomial systems, the approach we suggest extends naturally to systems of equations.

For instance, it is possible to bound the probability that the condition number of a random sparse polynomial system (arbitrary variances) is large.

The roots of a sparse polynomial system are known to belong to a certain toric variety. However, in order to obtain the theorems below, we needed to endow the toric variety with a certain geometrical structure, as explained below. The main insight comes from mechanics, and from symplectic and

Kähler geometry. The main tool is the Momentum map.

The proofs of the results mentioned herein may be found in [MR] and its references.

2 General setting

Let A be an $M \times n$ matrix, with non-negative integer entries. To the matrix A we associate the convex polytope $\mathrm{Conv}(A)$ given by the convex hull of all the rows, $\{A^\alpha\}_{\alpha \in \{1,\dots,M\}}$, of A:

$$\mathrm{Conv}(A) \stackrel{\mathrm{def}}{=} \left\{ \sum_{\alpha=1}^{M} t_\alpha A^\alpha \ : \ 0 \le t_\alpha \le 1, \ \sum_{\alpha=1}^{M} t_\alpha = 1 \right\} \subset (\mathbb{R}^n)^\vee \quad .$$

Here, we use the notation $X^\vee$ to denote the dual of a vector space X.

Assume that $\dim(\mathrm{Conv}(A)) = n$. Then we can associate to the matrix A the space $\mathcal{F}_A$ of polynomials with support contained in $\{A^\alpha : 1 \le \alpha \le M\}$. This is a linear space, and there are many reasonable choices of an inner product in $\mathcal{F}_A$.

Let C be a diagonal positive definite $M \times M$ matrix. Its inverse C^{-1} is also a diagonal positive definite $M \times M$ matrix. This inverse matrix defines the inner product:

$$\langle z^{A^\alpha}, z^{A^\beta} \rangle_{C^{-1}} = (C^{-1})_{\alpha,\beta} \quad .$$

The matrix C will be called the *variance matrix.* This terminology arises when we consider random normal polynomials in $\mathcal{F}_A$ with variance $C_{\alpha\alpha}$ for the α–th coefficient. We will refer to these randomly generated functions as *random normal polynomials*, for short.

We may also produce several objects associated to the matrix A (and to the variance matrix C). The most important one for this paper will be a Kähler manifold $(\mathcal{T}^n, \omega_A, J)$. This manifold is a natural "phase space" for the roots of polynomial systems with support in A. It is *the* natural phase space for the roots of systems of random normal polynomials in $(\mathcal{F}_A, \langle \cdot, \cdot \rangle_{C^{-1}})$.

More explicitly, let $\mathcal{T}^n \stackrel{\mathrm{def}}{=} \mathbb{C}^n \pmod{2\pi\sqrt{-1}\,\mathbb{Z}^n}$ (which, as a real manifold, happens to be an n-fold product of cylinders). Let $\exp : \mathcal{T}^n \to (\mathbb{C}^*)^n$ denote coordinatewise exponentiation. Then we will look at the preimages

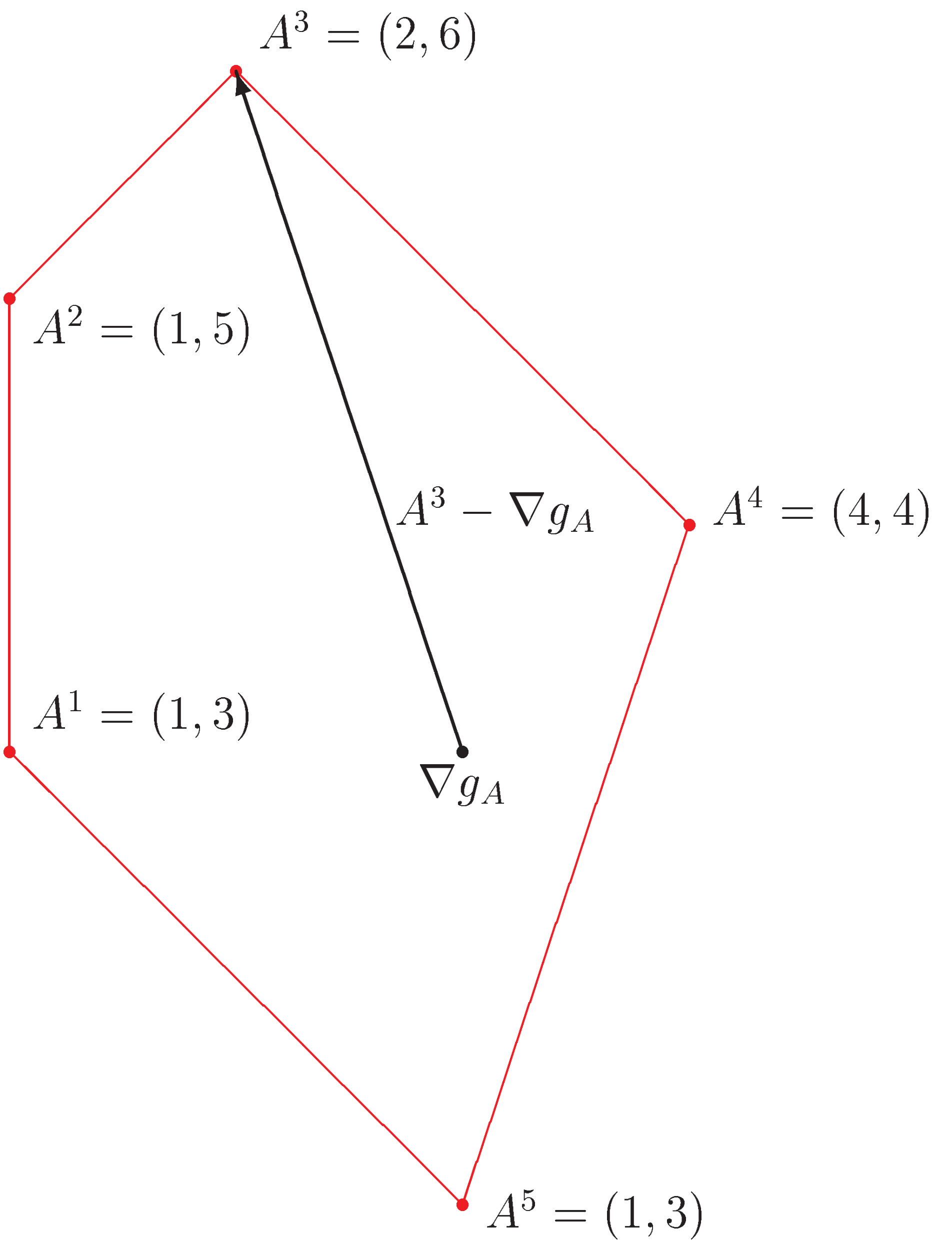

Figure 3: Geometric interpretation of the Momentum map ∇g_A. The momentum is related to the derivative of the Veronese embedding v_A by the formula: $Dv_A^\lambda = v_A^\lambda(A^\lambda - \nabla g_A)$

of the roots of a polynomial system by exp. We leave out roots that have one coordinate equal to zero and roots at infinity.

The Kähler manifold may be constructed as follows. The Complex Structure J is defined locally by

$$J_{(p,q)}(\dot{p}, \dot{q}) \stackrel{\text{def}}{=} (-\dot{q}, \dot{p})$$

The form ω_A is the pull-back of the Fubini-Study 2–form in $\mathbb{C}^M$ by the embedding

$$\begin{array}{rccl} \hat{v}_A : & \mathcal{T}^n & \rightarrow & \mathbb{C}^M \\ & (p,q) & \mapsto & C^{1/2} \cdot \exp(A(p + q\sqrt{-1})) \end{array}$$

If $v_A : \mathcal{T}^n \rightarrow \mathbb{P}^{M-1}$ is the composition of $\hat{v}_A$ with the canonical map $\mathbb{C}^M_* \rightarrow \mathbb{P}^{M-1}$, then ω_A is also the pull-back by v_A of the canonical symplectic form of $\mathbb{P}^{M-1}$.

Another important invariant is the real function:

$$\begin{array}{rccl} g_A : & \mathcal{T}^n & \rightarrow & \mathbb{R} \\ & p,q & \mapsto & \log \|\hat{v}_A(p)\| \end{array}$$

This function is independent of p and we may treat g_A as a function of p alone. In that setting, the gradient $\nabla g_A : \mathbb{R}^n \rightarrow (\mathbb{R}^n)^\vee$ maps $\mathbb{R}^n$ diffeomorphically onto the interior of the convex hull of the rows of A (See Figure 3 and [MR, Section 2.2]). The mapping

$$\nabla g_A : \mathcal{T}^n \rightarrow (\mathbb{R}^n)^\vee$$

is called the *Momentum map.* It was introduced in its modern formulation by Smale [Sma70] and Souriau [Sou70]. The reader may consult one of the many textbooks in the subject (such as Abraham and Marsden [AM78] or McDuff and Salamon [MS98]) for a general exposition.

We can say now what the "phase space" of Figures 1(c) and 2(c) was. It was the image of the toric manifold $(\mathcal{T}^n, \omega_A, J)$ by the "volume-preserving" (up to a constant) map:

$$(p,q) \mapsto ((\nabla g_A)_{|p}, q)$$

The potential g_A and the Kähler form ω_A are related as follows:

$$\omega_A = -\frac{1}{2} dJ^* dg_A$$

where d stands for exterior differentiation. The Hermitian metric associated to A has also the expressions:

$$\langle a, b\rangle \stackrel{\text{def}}{=} \omega_A(a, Jb) = a^H Dv_A^H Dv_A b = \frac{1}{2} D^2 g_A(a, b)$$

3 Mixed and Unmixed systems

Systems where all the polynomials have the same support are called *unmixed.* The general situation (*mixed* polynomial systems), where the polynomials may have different supports, is of greater practical interest. It is also a much more challenging situation. We shall consider systems of n polynomials in n variables, each polynomial in some inner product space of the form $(\mathcal{F}_{A_i}, \langle\cdot,\cdot\rangle_{C_i^{-1}})$ (where $i = 1, \cdots, n$ and each A_i and each C_i are as above).

In this realm, a mathematical object (that we may call a *mixed manifold*) seems to arise naturally. A mixed manifold is an $(n+2)$–tuple $(\mathcal{T}^n, \omega_{A_1}, \cdots, \omega_{A_n}, J)$ where for each i, $(\mathcal{T}^n, \omega_{A_i}, J)$ is a Kähler manifold. Mixed manifolds do **not** have a natural canonical Hermitian structure. They have n equally important Hermitian structures. However, they have one natural volume element, the *mixed volume form*, given by

$$d\mathcal{T}^n = \frac{(-1)^{n(n-1)/2}}{n!} \, \omega_{A_1} \wedge \cdots \wedge \omega_{A_n} \quad .$$

As explained in [Gro90], the volume of $\mathcal{T}^n$ relative to the mixed volume form is (up to a constant) the mixed volume of the n–tuple of polytopes $(\mathrm{Conv}(A_1), \cdots, \mathrm{Conv}(A_n))$.

We extend the famous result by Bernshtein [Ber76] on the number of roots of mixed systems of polynomials as follows:

Theorem 1. *Let $A_1, \cdots, A_n$ and $C_1, \cdots, C_n$ be as above. For each $i = 1, \cdots, n$, let f_i be an (independently distributed) normal random polynomial in $(\mathcal{F}_{A_i}, \langle\cdot,\cdot\rangle_{C_i^{-1}})$. Let U be a measurable region of $\mathcal{T}^n$. Then, the expected number of roots of the polynomial system $f(z) = 0$ in $\exp U \subseteq (\mathbb{C}^*)^n$ is*

$$\frac{n!}{\pi^n} \int_U d\mathcal{T}^n \quad .$$

Example 1. When each f_i is dense with a variance matrix C_i of the form:

$$C_i = \mathrm{Diag}\left(\frac{\deg f_i!}{I_1! I_2! \cdots , I_n! (\deg f_i - \sum_{j=1}^n I_j)!} \right) \quad ,$$

the volume element $d\mathcal{T}^n$ becomes the Bézout number $\prod \deg f_i$ times the pull-back to $\mathcal{T}^n$ of the Fubini-Study metric. We thus recover Shub and Smale's stochastic real version of Bézout's Theorem [SS93a]. ■

The general unmixed case ($A_1 = \cdots = A_n$, $C_1 = \cdots = C_n$) is a particular case of Theorem 8.1 in [EK95]. This is the only overlap, since neither theorem generalizes the other.

On the other hand, when one sets $U = \mathcal{T}^n$, one recovers Bernshtein's first theorem. The quantity $\pi^{-n} \int_{\mathcal{T}^n} d\mathcal{T}^n$ is precisely the *mixed volume* of polytopes $A_1, \cdots , A_n$ (see [SY93] for the classical definition of Mixed Volume and main properties).

A version of Theorem 1 was known to Kazarnovskii [Kaz81, p. 351] and Khovanskii. In [Kaz81], the supports A_i are allowed to have complex exponents. However, uniform variance ($C_i = I$) is assumed. His method may imply this special case of Theorem 1, but the indications given in [Kaz81] were insufficient for us to reconstruct a proof.

The idea of working with roots of polynomial systems in logarithmic coordinates seems to be extremely classical, yet it gives rise to interesting and surprising connections (see the discussions in [MZa, MZb, Vir00]).

4 The Condition Number

Let $\mathcal{F} = \mathcal{F}_{A_1} \times \cdots \times \mathcal{F}_{A_n}$, and let $f \in \mathcal{F}$. A *root* of f will be represented by some $p + q\sqrt{-1} \in \mathcal{T}^n$. (Properly speaking, the root of f is $\exp(p + q\sqrt{-1})$).

In this discussion, we assume that the "root" $p+q\sqrt{-1}$ is non-degenerate. This means that the derivative of the *evaluation map*

$$\begin{array}{rccc} ev : & \mathcal{F} \times \mathcal{T}^n & \to & \mathbb{C}^n \\ & (f, p + q\sqrt{-1}) & \mapsto & (f \circ \exp)(p + q\sqrt{-1}) \end{array}$$

with respect to the variable in $\mathcal{T}^n$ at the point $p+q\sqrt{-1}$ has rank $2n$. We are then in the situation of the implicit function theorem, and there is (locally) a smooth function $G : \mathcal{F} \to \mathcal{T}^n$ such that for $\hat{f}$ in a neighborhood of f, we have $ev(\hat{f}, G(\hat{f})) \equiv 0$ and $G(f) = p + q\sqrt{-1}$.

The condition number of f at $(p + q\sqrt{-1})$ is usually defined as

$$\boldsymbol{\mu}(f; p + q\sqrt{-1}) = \|DG_f\| \quad .$$

This definition is sensitive to the norm used in the space of linear maps between tangent spaces $L(T_f\mathcal{F}, T_{(p,q)}\mathcal{T}^n)$. In general, one would like to use an operator norm, related to some natural Hermitian or Riemannian structure on $\mathcal{F}$ and $\mathcal{T}^n$.

In the previous section, we already defined an inner product in each coordinate subspace $\mathcal{F}_{A_i}$, given by the variance matrix C_i. Since the evaluation function is homogeneous in each coordinate, it makes sense to projectivize each of the coordinate spaces $\mathcal{F}_{A_i}$ (with respect to the inner product $\langle \cdot, \cdot \rangle_{C_i^{-1}}$). Alternatively, we can use the Fubini-Study metric in each of the $\mathcal{F}_{A_i}$'s. By doing so, we are endowing $\mathcal{F}$ with a Fubini-like metric that is scaling-invariant. We will treat $\mathcal{F}$ as a multiprojective space, and write $\mathbb{P}(\mathcal{F})$ for $\mathbb{P}(\mathcal{F}_{A_1}) \times \cdots \times \mathbb{P}(\mathcal{F}_{A_n})$.

Another useful metric in $\mathbb{P}(\mathcal{F})$ is given by

$$d_{\mathbb{P}}(f, g)^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \left(\min_{\lambda \in \mathbb{C}^*} \frac{\|f^i - \lambda g^i\|}{\|f^i\|} \right)^2 \quad .$$

Each of the terms in the sum above corresponds to the square of the sine of the Fubini (or angular) distance between f^i and g^i. Therefore, $d_{\mathbb{P}}$ is never larger than the Hermitian distance between points in $\mathcal{F}$, but is a correct first-order aproximation of the distance when $g \to f$ in $\mathbb{P}(\mathcal{F})$. (Compare with [BCSS98, Ch. 12]).

While $\mathcal{F}$ admits a natural Hermitian structure, the solution-space $\mathcal{T}^n$ admits n possibly different Hermitian structures, corresponding to each of the Kähler forms ω_{A_i}.

In order to elucidate what the natural definition of a condition number for mixed systems of polynomials is, we will interpret the condition number as the inverse of the distance to the *discriminant locus*. Given $p + q\sqrt{-1} \in \mathcal{T}^n$, we set:

$$\mathcal{F}_{(p,q)} = \{f \in \mathcal{F} : ev(f; (p, q)) = 0\}$$

and we set $\Sigma_{(p,q)}$ as the space of degenerate polynomial systems in $\mathcal{F}_{(p,q)}$. Since the fiber $\mathcal{F}_{(p,q)}$ inherits the metric structure of $\mathcal{F}$, we can speak of the distance to the discriminant locus along a fiber. In this setting, Theorem 3 in [BCSS98, p. 234] becomes:

Theorem 2 (Condition number theorem). *Under the notations above, if (p,q) is a non-degenerate root of f,*

$$\max_{\|\dot f\|\leq 1} \min_i \|DG_f \dot f\|_{A_i} \leq \frac{1}{d_{\mathbb{P}}(f, \Sigma_{(p,q)})} \leq \max_{\|\dot f\|\leq 1} \max_i \|DG_f \dot f\|_{A_i} \quad .$$

There are two interesting particular cases. First of all, if $A_1 = \cdots = A_n$ and $C_1 = \cdots = C_n$, we obtain an equality:

Corollary 2.1 (Condition number theorem for unmixed systems). *Let $A_1 = \cdots = A_n$ and $C_1 = \cdots = C_n$, then under the hypotheses of Theorem 2,*

$$\boldsymbol{\mu}(f;(p,q)) \stackrel{\text{def}}{=} \max_{\|\dot f\|\leq 1} \min_i \|DG_f \dot f\|_{A_i} = \max_i \max_{\|\dot f\|\leq 1} \|DG_f \dot f\|_{A_i} = \frac{1}{d_{\mathbb{P}}(f, \Sigma_{(p,q)})} \quad .$$

We can also obtain a version of Shub and Smale's condition number theorem (Theorem 3 in [BCSS98, p. 243]) for dense systems as a particular case, once we choose the correct variance matrices:

Corollary 2.2 (Condition number theorem for dense systems). *Let $d_1, \cdots, d_n$ be positive integers, and let A_i be the n-columns matrix having all possible rows with non-negative entries adding up to at most d_i. Let*

$$C_i = \frac{1}{d_i} \operatorname{Diag}\left(\frac{d_i!}{(A_i)_1^\alpha! (A_i)_2^\alpha! \cdots (A_i)_n^\alpha! (d_i - \sum_{j=1}^n (A_i)_j^\alpha)!} \right) \quad .$$

Then,

$$\boldsymbol{\mu}(f;(p,q)) \stackrel{\text{def}}{=} \max_{\|\dot f\|\leq 1} \min_i \|DG_f \dot f\|_{A_i} = \max_i \max_{\|\dot f\|\leq 1} \|DG_f \dot f\|_{A_i} = \frac{1}{d_{\mathbb{P}}(f, \Sigma_{(p,q)})} \quad .$$

The factor $\frac{1}{d_i}$ in the definition of the variance matrix C_i corresponds to the factor $\sqrt{d_i}$ in the definition of the normalized condition number in [BCSS98, p. 233].

In the general mixed case, we would like to interpret the two "minmax" bounds as condition numbers related to some natural Hermitian or Finslerian structures on $\mathcal{T}^n$. See [MR, Section 5] for a discussion

Theorem 2 is very similar to Theorem D in [Ded96], but the philosophy here is radically different. Instead of changing the metric in the fiber $\mathcal{F}_{(p,q)}$, we consider the inner product in $\mathcal{F}$ as the starting point of our investigation. Theorem 2 gives us some insight about reasonable metric structures in $\mathcal{T}^n$.

As in Theorem 1, let U be a measurable set of $\mathcal{T}^n$. In view of Theorem 2, we define a restricted condition number (with respect to U) by:

$$\boldsymbol{\mu}(f;U) \stackrel{\text{def}}{=} \frac{1}{\min_{(p,q)\in U} d_{\mathbb{P}}(f, \Sigma_{(p,q)})}$$

where the distance $d_{\mathbb{P}}$ is taken along the fiber $\mathcal{F}_{(p,q)} = \{f : (f \circ \exp)(p + q\sqrt{-1}) = 0\}$.

Although we do not know in general how to bound the expected value of $\boldsymbol{\mu}(f;\mathcal{T}^n)$, we can give a convenient bound for $\boldsymbol{\mu}(f;U)$ whenever U is compact and in some cases where U is not compact.

The group $GL(n)$ acts on $T_{(p,q)}\mathcal{T}^n$ by sending $(\dot{p}, \dot{q})$ into $(L\dot{p}, L\dot{q})$, for any $L \in GL(n)$. In more intrinsic terms, J and the $GL(n)$-action commute. With this convention, we can define an intrinsic invariant of the mixed structure $(\mathcal{T}^n, \omega_{A_1}, \cdots, \omega_{A_n}, J)$:

Definition 1. The *mixed dilation* of the tuple $(\omega_{A_1}, \cdots, \omega_{A_n})$ is:

$$\kappa(\omega_{A_1}, \cdots, \omega_{A_n}; (p,q)) \stackrel{\text{def}}{=} \min_{L\in GL(n)} \max_i \frac{\max_{\|u\|=1} (\omega_{A_i})_{(p,q)}(Lu, JLu)}{\min_{\|u\|=1} (\omega_{A_i})_{(p,q)}(Lu, JLu)} \quad .$$

Given a set U, we define:

$$\kappa_U \stackrel{\text{def}}{=} \sup_{(p,q)\in U} \kappa(\omega_{A_1}, \cdots, \omega_{A_n}; (p,q)) \quad ,$$

provided the supremum exists, and $\kappa_U = \infty$ otherwise.

We will bound the expected number of roots with condition number $\boldsymbol{\mu} > \varepsilon^{-1}$ on U in terms of the mixed volume form, the mixed dilation κ_U and the expected number of ill-conditioned roots in the *linear case*. The linear case corresponds to polytopes and variances below:

$$A_i^{\text{Lin}} = \begin{bmatrix} 0 & \cdots & 0 \\ 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \qquad\qquad C_i^{\text{Lin}} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

Theorem 3 (Expected value of the condition number). *Let $\nu^{\mathrm{Lin}}(n,\varepsilon)$ be the probability that a random n–variate linear complex polynomial has condition number larger than ε^{-1}. Let $\nu^{A}(U,\varepsilon)$ be the probability that $\boldsymbol{\mu}(f,U) > \varepsilon^{-1}$ for a normal random polynomial system f with supports $A_1,\cdots,A_n$ and variance $C_1,\cdots,C_n$.*

Then,

$$\nu^{A}(U,\varepsilon) \le \frac{\int_U \bigwedge \omega_{A_i}}{\int_U \bigwedge \omega_{A_i^{\mathrm{Lin}}}}\ \nu^{\mathrm{Lin}}(n,\sqrt{\kappa_U}\varepsilon) \quad .$$

There are a few situations where we can assert that $\kappa_U = 1$. For instance,

Corollary 3.1. *Under the hypotheses of Theorem 3, if $A = A_1 = \cdots = A_n$ and $C = C_1 = \cdots = C_n$, then*

$$\nu^{A}(U,\varepsilon) \le \mathrm{Vol}(U)\ \nu^{\mathrm{Lin}}(n,\varepsilon) \quad .$$

The dense case (Theorem 1 p. 237 in [BCSS98]) is also a consequence of Theorem 3.

Remark 1. We interpret $\nu^{\mathrm{Lin}}(n,\varepsilon)$ as the probability that a random linear polynomial f is at multiprojective distance less than ε from the discriminant variety $\Sigma_{(p,q)}$. Let $g \in \Sigma_{(p,q)}$ be such that the following minimum is attained:

$$d_{\mathbb{P}}(f,\Sigma_{(p,q)})^2 = \inf_{\substack{g\in\Sigma_{(p,q)}\\ \lambda\in(\mathbb{C}^*)^n}} \sum_{i=1}^{n} \frac{\|f^i - \lambda_i g^i\|^2}{\|f^i\|^2} \quad .$$

Without loss of generality, we may scale g such that $\lambda_1 = \cdots = \lambda_n = 0$. In that case,

$$d_{\mathbb{P}}(f,\Sigma_{(p,q)})^2 = \sum_{i=1}^{n} \frac{\|f^i - g^i\|^2}{\|f^i\|^2} \ge \frac{\sum_{i=1}^{n}\|f^i - g^i\|^2}{\sum_{i=1}^{n}\|f^i\|^2} \quad .$$

The right hand term is the projective distance to the discriminant variety along the fiber, in the sense of [BCSS98]. Since we are in the linear case, this may be interpreted as the inverse of the condition number of f in the sense of [BCSS98, Prop. 4 and Remark 2 p. 250].

Recall that each f^i is an independent random normal linear polynomial of degree 1, and that C_i is the identity. Therefore, each f^i_α is an i.i.d. Gaussian variable. If we look at the system f as a random variable in $\mathbb{P}^{n(n+1)-1}$, then we obtain the same probability distribution as in [BCSS98]. Then, using Theorem 6 p. 254 *ibid*, we deduce that

$$\nu^{\mathrm{Lin}}(n, \varepsilon) \leq \frac{n^3(n+1)\Gamma(n^2+n)}{\Gamma(n^2+n-2)}\varepsilon^4 \quad . \blacksquare$$

5 Real Polynomials

Shub and Smale showed in [SS93a] that the expected number of real roots, in the dense case (with unitarily invariant probability measure) is exactly the square root of the expected number of roots.

Unfortunately, this result seems to be very hard to generalize to the unmixed case. Under certain conditions, explicit formulæ for the unmixed case are available [Roj96]. Also, less explicit bounds for the multi-homogeneous case were given by [McL00].

Here, we will give a very coarse estimate in terms of the square root of the mixed volume:

Theorem 4. *Let U be a measurable set in $\mathbb{R}^n$, with total Lebesgue volume $\lambda(U)$. Let $A_1, \cdots, A_n$ and $C_1, \cdots, C_n$ be as above. Let f be a normal random real polynomial system. Then the average number of real roots of f in $\exp U \subset (R_*^+)^n$ is bounded above by*

$$(4\pi^2)^{-n/2}\sqrt{\lambda(U)}\sqrt{\int_{\substack{(p,q)\in\mathcal{T}^n \\ p\in U}} n! d\mathcal{T}^n} \quad .$$

This is of interest when n and U are fixed. In that case, the expected number of positive real roots (hence of real roots) grows as the square root of the mixed volume.

It is somewhat easier to investigate real random polynomials in the unmixed case.

Let $\nu_{\mathbb{R}}(n, \varepsilon)$ be the probability that a linear random real polynomial has condition number larger than ε^{-1}.

Theorem 5. *Let $A = A_1 = \cdots = A_n$ and $C = C_1 = \cdots = C_n$. Let $U \subseteq \mathbb{R}^n$ be measurable. Let f be a normal random real polynomial system. Then,*

$$\mathrm{Prob}\left[\boldsymbol{\mu}(f, U) > \varepsilon^{-1}\right] \leq E(U)\ \nu_{\mathbb{R}}(n, \varepsilon)$$

where $E(U)$ is the expected number of real roots on U.

Notice that $E(U)$ depends on C. Even if we make $U = \mathbb{R}^n$, we may still obtain a bound depending on C.

6 Mechanical Interpretation

The momentum map defined above is also the the momentum map (in the sense of [Sma70]) associated to a certain Lie group action, namely the natural action of the n-torus on the *toric* manifold $\mathcal{T}^n$:

The n-torus $\mathbb{T}^n = \mathbb{R}^n \pmod{2\pi\,\mathbb{Z}^n}$ acts on $\mathcal{T}^n$ by

$$\rho : (p, q) \mapsto (p, q + \rho) \quad ,$$

where $\rho \in \mathbb{T}^n$.

This action preserves the symplectic structure, since it fixes the p-variables and translates the q-variables. Also, the Lie algebra of $\mathbb{T}^n$ is $\mathbb{R}^n$. An element ξ of $\mathbb{R}^n$ induces an *infinitesimal* action (i.e. a vector field) X_ξ in $\mathcal{T}^n$.

This vector field is the derivation that to any smooth function f associates:

$$(X_\xi)_{(p,q)}(f) = \iota_\xi(\frac{1}{2}\omega_A)_{(p,q)}(df) \stackrel{\text{def}}{=} (\frac{1}{2}\omega_A)_{(p,q)}(\xi, df) \quad .$$

If we write $df = d_p f\, dp + d_q f\, dq$, then this formula translates to:

$$(X_\xi)_{(p,q)}(f) = -\xi^T (D^2 g_A)_p d_p f$$

This vector field is Hamiltonian: if $(p(t), q(t))$ is a solution of the equation

$$(\dot{p}(t), \dot{q}(t)) = (X_\xi)_{p(t),q(t)}$$

then we can write

$$\left\{ \begin{array}{rcl} \dot{p} & = & \frac{\partial H_\xi}{\partial q} \\ \dot{q} & = & -\frac{\partial H_\xi}{\partial p} \end{array} \right. \quad ,$$

where $H_\xi = \frac{1}{2}\nabla g_A(p) \cdot \xi$.

This construction associates to every $\xi \in \mathbb{R}^n$, the Hamiltonian function $H_\xi = \nabla g_A(p) \cdot \xi$. The term $\nabla g_A(p)$ is a function of p, with values in $(\mathbb{R}^n)^\vee$ (the dual of $\mathbb{R}^n$). In more general Lie group actions, the momentum map takes values in the dual of the Lie algebra, so that the pairing $\nabla g_A(p) \cdot \xi$ always makes sense. A Lie group action with such an expression for the Hamiltonian is called *Hamiltonian* or *Strongly Hamiltonian.*

7 Acknowledgements

Steve Smale provided valuable inspiration for us to develop a theory similar to [SS93b, SS93a, SS93c, SS96, SS94] for sparse polynomial systems. He also provided examples on how to eliminate the dependency upon unitary invariance in the dense case.

The paper by Gromov [Gro90] was of foremost importance to this research. To the best of our knowledge, [Gro90] is the only clear exposition available of mixed volume in terms of a wedge of differential forms. We thank Mike Shub for pointing out that reference, and for many suggestions.

We would like to thank Jean-Pierre Dedieu for sharing his thoughts with us on Newton iteration in Riemannian and quotient manifolds.

Also, we would like to thank Felipe Cucker, Alicia Dickenstein, Ioannis Emiris, Askold Khovanskii, Eric Kostlan, T.Y. Li, Martin Sombra, and Jorge P. Zubelli for their suggestions and support.

This paper was written while G.M. was visiting the Liu Bie Ju Center for Mathematics at the City University of Hong Kong. He wishes to thank CityU for their generous support.

References

[AM78] Ralph Abraham and Jerrold E. Marsden, *Foundations of mechanics*, Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass., 1978, Second edition, revised and enlarged, With the assistance of Tudor Ratiu and Richard Cushman.

[BCSS98] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998, With a foreword by Richard M. Karp.

[Ber76] D. N. Bernstein, *The number of roots of a system of equations*, Functional Anal. Appl. 9 **9 (1975)** (1976), no. 3, 183–185.

[Ded96] Jean-Pierre Dedieu, *Approximate solutions of numerical problems, condition number analysis and condition number theorem*, The mathematics of numerical analysis (Park City, UT, 1995), Amer. Math. Soc., Providence, RI, 1996, pp. 263–283.

[EK95] Alan Edelman and Eric Kostlan, *How many zeros of a random polynomial are real?*, Bull. Amer. Math. Soc. (N.S.) **32** (1995), no. 1, 1–37.

[Gro90] M. Gromov, *Convex sets and Kähler manifolds*, Advances in differential geometry and topology, World Sci. Publishing, Teaneck, NJ, 1990, pp. 1–38.

[Kaz81] B. Ja. Kazarnovskiĭ, *On zeros of exponential sums*, Soviet Math. Doklady **23** (1981), no. 2, 347–351.

[Kos] Eric Kostlan, *Random polynomials and the statistical fundamental theorem of algebra*, Preprint, MSRI, 1987.

[McL00] Andrew McLennan, *The expected number of real roots of a multihomogeneous system of polynomial equations*, Preprint, Dept of Economics, Univ. of Minnesota, February 2000.

[MR] Gregorio Malajovich and J. Maurice Rojas, *Random sparse polynomial systems*, Preprint, City University of Hong Kong and Math Archive, 2000.

[MS98] Dusa McDuff and Dietmar Salamon, *Introduction to symplectic topology*, second ed., The Clarendon Press Oxford University Press, New York, 1998.

[MZa] Gregorio Malajovich and Jorge Zubelli, *On the geometry of graeffe iteration*, Journal of Complexity (To appear).

[MZb] Gregorio Malajovich and Jorge Zubelli, *Tangent graeffe iteration*, Numerische Mathematik (To appear).

[Roj96] J. Maurice Rojas, *On the average number of real roots of certain random sparse polynomial systems*, Lectures in Applied Mathematics **32** (1996), 689–699.

[Sma70] S. Smale, *Topology and mechanics. I*, Invent. Math. **10** (1970), 305–331.

[Sou70] J.-M. Souriau, *Structure des systèmes dynamiques*, Dunod, Paris, 1970, Maîtrises de mathématiques.

[SS93a] M. Shub and S. Smale, *Complexity of Bezout's theorem. II. Volumes and probabilities*, Computational algebraic geometry (Nice, 1992), Birkhäuser Boston, Boston, MA, 1993, pp. 267–285.

[SS93b] Michael Shub and Steve Smale, *Complexity of Bézout's theorem. I. Geometric aspects*, J. Amer. Math. Soc. **6** (1993), no. 2, 459–501.

[SS93c] Michael Shub and Steve Smale, *Complexity of Bezout's theorem. III. Condition number and packing*, J. Complexity **9** (1993), no. 1, 4–14, Festschrift for Joseph F. Traub, Part I.

[SS94] M. Shub and S. Smale, *Complexity of Bezout's theorem. V. Polynomial time*, Theoret. Comput. Sci. **133** (1994), no. 1, 141–164, Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993).

[SS96] Michael Shub and Steve Smale, *Complexity of Bezout's theorem. IV. Probability of success; extensions*, SIAM J. Numer. Anal. **33** (1996), no. 1, 128–148.

[SY93] J. R. Sangwine-Yager, *Mixed volumes*, Handbook of convex geometry, Vol. A, B, North-Holland, Amsterdam, 1993, pp. 43–71.

[Vir00] Oleg Viro, *Dequantization of real algebraic geometry on logarithmic paper*, June 2000, Preprint, Mathematics ArXiv AG/0005163, http://front.math.ucdavis.edu.