Random Sparse Polynomial Systems

Gregorio Malajovich*† J. Maurice Rojas‡ §

December 11, 2000

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

Abstract

Let $f := (f^1, \ldots, f^n)$ be a sparse random polynomial system. This means that each f^i has fixed support (list of possibly non-zero coefficients) and each coefficient has a Gaussian probability distribution of arbitrary variance.

We express the expected number of roots of f inside a region U as the integral over U of a certain **mixed volume** form. When $U = (\mathbb{C}^*)^n$, the classical mixed volume is recovered.

The main result in this paper is a bound on the probability that the condition number of f on the region U is larger than $1/\varepsilon$. This bound depends on the integral of the mixed volume form over U, and on a certain intrinsic invariant of U as a subset of a toric manifold.

Polynomials with real coefficients are also considered, and bounds for the expected number of real roots and for the condition number are given.

The connection between zeros of sparse random polynomial systems, Kähler geometry, and mechanics (momentum maps) is discussed.

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

---

*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, 83 Tat Chee Ave, Kowloon, Hong Kong.

†G.M.'s visit to City University of Hong Kong was supported by CERG grants 9040393-730, 9040402-730, and 9040188

‡Department of Mathematics, City University of Hong Kong (83 Tat Chee Ave., Kowloon, HONG KONG) and Department of Mathematics, Texas A&M University (College Station, Texas 77843-3368, USA). `http://math.cityu.edu.hk/ mamrojas` . **e-mail:** `mamrojas@cityu.edu.hk` (before January 2001), `rojas@math.tamu.edu` (after January 2001).

§The authors work on this paper was supported by Hong Kong UGC grant #9040402-730, Hong Kong/France PROCORE Grant #9050140-730, and a US National Science Foundation Mathematical Sciences Postdoctoral Fellowship.

Contents

1 Introduction

A complexity theory of homotopy algorithms for solving dense systems of polynomial equations was developed in [?, ?, ?, ?, ?]. (see also [?, Ch. 8–14]). The number of steps for these homotopy algorithms was bounded in terms of a condition number and the Bézout number.

One of the main features of that theory was unitary invariance: the roots of a dense system of polynomial equations are points in projective space, and all the invariants of the theory are invariant under the action of the unitary group. However, unitary action does not preserve sparse coefficient structure.

In this paper, we give the one-distribution of the roots of random sparse polynomial systems. We also bound the probability that the condition number of a random sparse polynomial system 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.

1.1 Expected Number of Roots

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. There are deep reasons to write $\mathrm{Conv}(A)$ as a polytope in dual space, as the reader will see later on.

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 will 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 $\mathcal{T}^n \stackrel{\text{def}}{=} \mathbb{C}^n \pmod{2\pi\sqrt{-1}\,\mathbb{Z}^n}$ (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 of the roots of a polynomial system by exp. We leave out roots that have at least one coordinate equal to zero and roots at infinity. The differential 2-form ω_A corresponds to the pull-back of the canonical 2-form in a suitable Veronese variety (see Section 2).

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 [?], 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 [?] 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 [?]. ■

The general unmixed case ($A_1 = \cdots = A_n$, $C_1 = \cdots = C_n$) is a particular case of Theorem 8.1 in [?]. 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 [?] for the classical definition of Mixed Volume and main properties).

A version of Theorem 1 was known to Kazarnovskii [?, p. 351] and Khovanskii. In [?], 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 [?] 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 [?, ?, ?]).

1.2 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*

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

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 $G : \mathcal{F} \rightarrow \mathcal{T}^n$ 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 $g \rightarrow f$ in $\mathbb{P}(\mathcal{F})$. (Compare with [**?**, 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 [**?**, p. 234] becomes:

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

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

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,*

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

We can also obtain a version of Shub and Smale's condition number theorem (Theorem 3 in [**?**, 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*

$$C_i = \frac{1}{d_i} \operatorname{Diag}\left(\frac{d_i!}{(A_i)^\alpha_1!(A_i)^\alpha_2! \cdots (A_i)^\alpha_n!(d_i - \sum_{j=1}^n (A_i)^\alpha_j)!} \right) \quad .$$

Then,

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

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 [**?**, p. 233] (see Remark 5 p. 15 below).

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 Section 5 for a discussion.

Theorem 2 is very similar to Theorem D in [**?**], 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:

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

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 [?]) 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 [?]. Since we are in the linear case, this may be interpreted as the inverse of the condition number of f in the sense of [?, 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 [?]. Then, using Theorem 6 p. 254 *ibid*, we deduce that

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

1.3 Real Polynomials

Shub and Smale showed in [?] 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 [?]. Also, less explicit bounds for the multi-homogeneous case were given by [?].

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,*

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

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.

1.4 Acknowledgements

Steve Smale provided valuable inspiration for us to develop a theory similar to [?, ?, ?, ?, ?] 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 [?] was of foremost importance to this research. To the best of our knowledge, [?] 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 the generous support.

2 Symplectic Geometry and Polynomial Systems

2.1 About Symplectic Geometry

Definition 2 (Symplectic structure). Let M be a manifold. A 2-form on M is said to be non-degenerate if and only if for all $x \in M$, the only vector $u \in T_xM$ such that for all $v \in T_xM$, $\omega_x(u, v) = 0$ is the zero vector.

A *symplectic form* on M is a closed, non-degenerate 2-form ω on M. In that case, (M, ω) is said to be a symplectic manifold.

Definition 3 (Complex structure). Let M be a complex manifold. (We assume that M is given with a certain maximal holomorphic atlas). If $X : U \subset \mathbb{C}^n \to M$ is a chart of M, and $p = X(z) \in M$, then we define $J_p : T_pM \to T_pM$ such that the following diagram commutes:

$$\begin{array}{ccc} T_pM & \xrightarrow{J_p} & T_pM \\ {\scriptstyle DX_z}\uparrow & & {\scriptstyle DX_z}\uparrow \\ T_z\mathbb{C}^n & \xrightarrow{\text{Mult. by } \sqrt{-1}} & T_z\mathbb{C}^n \end{array} .$$

This is well-defined for each p in M. Indeed, if two charts X and Y overlap at p, then $Y \circ X^{-1} : \mathbb{C}^n \to \mathbb{C}^n$ is holomorphic so its derivative exists and commutes with multiplication by $\sqrt{-1}$.

The map

$$\begin{array}{rccc} J : & TM & \to & TM \\ & (p, \dot{p}) & \mapsto & (p, J_p\dot{p}) \end{array}$$

is called the *canonical complex structure* of M. (The complex structure may depend on the holomorphic atlas. We assume a canonical holomorphic atlas of M is given). Note that $-J^2$ is the identity on TM.

Definition 4 (Kähler manifolds). Let M be a complex manifold, with complex structure J. Let ω be a symplectic form on M (considered as a real manifold). Then ω is said to be a $(1,1)$-*form* if and only if $J^*\omega = \omega$. A $(1,1)$ form ω corresponds to a symmetric form $u, v \mapsto \omega(u, Jv)$. We say that ω is strictly positive if and only if the corresponding symmetric form is positive definite for all $p \in M$. Therefore, a strictly positive $(1,1)$-form defines a Riemann structure on M. Also, we can recover an Hermitian structure on M by setting $\langle u, v \rangle = \omega(u, Jv) + \sqrt{-1}\omega(u, v)$.

The triple (M, ω, J) is said to be a Kähler manifold when M, ω and J are as above.

Example 2 ($\mathbb{C}^M$). We identify $\mathbb{C}^M$ to $\mathbb{R}^{2M}$ and use coordinates $Z^i = X^i + \sqrt{-1}Y^i$. The *canonical* 2-form $\omega_Z = \sum_{i=1}^{M} dX_i \wedge dY_i$ makes $\mathbb{C}^M$ into a symplectic manifold.

The natural complex structure J is just the multiplication by $\sqrt{-1}$. The triple $(\mathbb{C}^M, \omega_Z, J)$ is a Kähler manifold. ■

Example 3 (Projective space). Projective space $\mathbb{P}^{M-1}$ admits a *canonical* 2-form defined as follows. Let $Z = (Z^1, \cdots, Z^M) \in (\mathbb{C}^M)^*$, and let $[Z] = (Z^1 : \cdots : Z^M) \in \mathbb{P}^{M-1}$ be the corresponding point in $\mathbb{P}^{M-1}$. The tangent space $T_{[Z]}\mathbb{P}^{M-1}$ may be modelled by $Z^\perp \subset T_Z\mathbb{C}^M$. Then we can define a two-form on $\mathbb{P}^{M-1}$ by setting:

$$\omega_{[Z]}(u, v) = \|Z\|^{-2}\omega_Z(u, v) \quad ,$$

where it is assumed that u and v are orthogonal to Z. The latter assumption tends to be quite inconvenient, and most people prefer to pull $\omega_{[Z]}$ back to $\mathbb{C}^M$ by the canonical projection $\pi : Z \mapsto [Z]$. It is standard to write the pull-back $\tau = \pi^*\omega_{[Z]}$ as:

$$\tau_Z = -\frac{1}{2} dJ^* d \, \frac{1}{2} \log \|Z\|^2 \quad ,$$

using the notation $d\eta = \sum_i \frac{\partial \eta}{p_i} \wedge dp_i + \frac{\partial \eta}{q_i} \wedge dq_i$, and where J^* denotes the pull-back by J.

Projective space also inherits the complex structure from $\mathbb{C}^M$. Then $\omega_{[Z]}$ is a strictly positive $(1, 1)$-form. The corresponding metric is called *Fubini-Study* metric in $\mathbb{C}^M$ or $\mathbb{C}^{M-1}$. ■

Remark 2. Some authors prefer to write $\sqrt{-1}\partial\bar{\partial}$ instead of $-\frac{1}{2}dJ^*d$. The following notation is assumed: $\partial\eta = \sum_i \frac{\partial\eta}{Z_i} \wedge dZ_i$ and $\bar{\partial}\eta = \sum_i \frac{\partial\eta}{\bar{Z}_i} \wedge d\bar{Z}_i$. Then they write τ_Z as:

$$\tau_Z = \frac{\sqrt{-1}}{2} \left(\frac{\sum_i dZ_i \wedge d\bar{Z}_i}{\|Z\|^2} - \frac{\sum_i Z_i d\bar{Z}_i \wedge \sum_j \bar{Z}_j dZ_j}{\|Z\|^4} \right) \quad . \; ■$$

Example 4. Let A be an $M \times n$ matrix with non-negative integer entries, and we associate every row A^α of A to the monomial $z^{A^\alpha} = z_1^{A_1^\alpha} \cdots z_n^{A_n^\alpha}$. We also assume (as in the introduction) that the corresponding polytope (the convex hull of all the rows) is n-dimensional. Also, as in the introduction, let C be a diagonal positive-definite matrix (that we called the variance matrix). The variance matrix was the matrix of the inner product in $\mathcal{F}_A$. Let $C^{1/2}$ be the diagonal positive-definite matrix such that $C = C^{1/2}C^{1/2}$. The right-multiplication of some $f \in \mathcal{F}_A$ by $C^{-1/2}$ makes the inner product canonical. The left-multiplication by $C^{1/2}$ is the pull-back of this operation in dual-space $\mathcal{F}_A^\vee$.

We define the map $\hat{V}_A$ from $\mathbb{C}^n$ into $\mathbb{C}^M$:

$$\hat{V}_A : z \mapsto C^{1/2} \begin{bmatrix} z^{A^1} \\ \vdots \\ z^{A^M} \end{bmatrix} \quad .$$

Because $C^{1/2}$ is diagonal, $\|\hat{V}_A(z)\|$ is invariant by the natural action $z_i \mapsto z_i e^{\theta_i \sqrt{-1}}$ of $(S^1)^n$ on the variable z. Moreover, we still have the pairing $f(z) = (f \cdot C^{-1/2}) \cdot \hat{V}_A(z)$. The variable $(f \cdot C^{-1/2})$ is corresponds to M_i i.i.d. Gaussian variables.

We can also compose with the projection into projective space, $V_A = \pi \circ \hat{V}_A : \mathbb{C}^n \rightarrow \mathbb{P}^{M-1}$,

$$V_A : z \mapsto [\hat{V}_A(z)] \quad .$$

When C is the identity, the Zariski closure of the image of V_A is called the *Veronese variety.* The map V_A is called the *Veronese embedding.* Notice that V_A is not defined for certain values of z, like $z = 0$. Those values are called the *exceptional set.* The exceptional set is contained in the union of the planes $z_i = 0$.

There is a natural symplectic structure on the closure of the image of V_A, given by the restriction of the Fubini-Study 2-form. We will see below (Lemma 1) that DV_A has rank n for $z \in (\mathbb{C}^*)^n$, because the polytope of A has dimension n. Thus, we can pull-back this structure to $(C^*)^n$ by:

$$\Omega_A = V_A^* \tau \quad .$$

Also, we can pull back the complex structure of $\mathbb{P}^{M-1}$, so that Ω_A becomes a strictly positive $(1,1)$-form.

Therefore, the matrix A defines a Kähler manifold $((C^*)^n, \Omega_A, J)$. ■

Example 5. Let $\mathcal{T}^n = \mathbb{C}^n \pmod{2\pi\sqrt{-1}\,\mathbb{Z}^n}$. We will use coordinates $p + q\sqrt{-1}$ for $\mathcal{T}^n$, where $p \in \mathbb{R}^n$ and $q \in \mathbb{T}^n = \mathbb{R}^n \pmod{2\pi\,\mathbb{Z}^n}$.

Given $(p,q) \in \mathcal{T}^n$, we define $\exp(p,q) = \left(\cdots, e^{p+q\sqrt{-1}}, \cdots \right) \in (\mathbb{C}^*)^n$.

For any matrix A as in the previous example, we can pull-back the Kähler structure of $((\mathbb{C}^*)^n, \Omega_A, J)$ to obtain another Kähler manifold $(\mathcal{T}^n, \omega_A, J)$. (Actually, it is the same object in logarithmic coordinates, minus points at "infinity".) An equivalent definition is to pull back the Kähler structure of the Veronese variety by

$$\hat{v}_A \stackrel{\text{def}}{=} \hat{V}_A \circ \exp \quad . \blacksquare$$

Remark 3. The Fubini-Study metric on $\mathbb{C}^M$ was constructed by applying the operator $-\frac{1}{2}dJ^*d$ to a certain convex function (in our case, $\frac{1}{2}\log\|Z\|^2$). This is a general standard way to construct Kähler structures. In [?], it is explained how to associate a (non-unique) convex function to any convex body, thus producing an associated Kähler metric. ■

Remark 4. Now a little bit of magic... $\omega_A = \hat{v}^*\tau = \hat{v}^*(-\frac{1}{2}dJ^*d)g$, where $g : Z \mapsto \frac{1}{2}\log\|Z\|^2$. Both d and J commute with pull-back, so

$$\begin{aligned} \omega_A &= -\hat{v}^*(\frac{1}{2}dJ^*d)g \\ &= -(\frac{1}{2}dJ^*d)\hat{v}_A^* g \\ &= -(\frac{1}{2}dJ^*d)(g \circ \hat{v}_A) \end{aligned}$$ ■

Remark 5. The same is true for $((\mathbb{C}^*)^n, \Omega_A, J)$. A particular case should be mentioned here. Unitary invariance played an important role in [?, ?, ?, ?, ?] and in [?]. Let us now recover that invariance for dense polynomials.

Suppose that the rows of our matrix A are the exponent vectors of all possible monomials of degree exactly D in $n+1$ variables. Let $A^\alpha = [I_1, \cdots, I_{n+1}]$. We set $C = \mathrm{Diag}\left(\frac{D!}{I_1!I_2!\cdots,I_{n+1}}\right)$. Then,

$$g \circ \hat{V}_A = \frac{1}{2}\log\|\hat{V}_A\|^2 = \frac{1}{2}\log\|z_1, \cdots, z_{n+1}\|^{2D} = D\frac{1}{2}\log\|z_1, \cdots, z_{n+1}\|^2 \quad .$$

So Ω_A is a multiple of the Fubini-Study metric, and we can actually extend Ω_A to $\mathbb{C}^{n+1}_*$.

Let $\tilde{f} = fC^{-1/2} \in (\mathbb{C}^M)^\vee$ represent the polynomial $z \mapsto f(z) = \tilde{f}\hat{V}_A(z)$. Then evaluation corresponds to the pairing $(\tilde{f}, z) \mapsto \tilde{f} \cdot V_A(z)$. The action of $U(n)$ on $\mathbb{C}^{n+1}$ induces an action on $(\mathbb{C}^M)^\vee$. All these actions are unitary, and the Hermitian structure of the space of polynomials (in the coordinates above) is invariant under such actions. ■

For the record, we state explicit formulæ for several of the invariants associated to the Kähler manifold $(\mathcal{T}^n, \omega_A, J)$. First of all, the function $g_A = g \circ \hat{v}_A$ is precisely:

Formula 2.1.1: The canonical Integral g_A (or *Kähler potential*) of the convex set associated to A

$$g_A(p) = \frac{1}{2} \log \Big((\exp(A \cdot p))^T \, C \, (\exp(A \cdot p)) \Big)$$

The terminology *integral* is borrowed from mechanics, and its refers to the invariance of g_A by $\mathbb{T}^n$-action (see appendix A for more analogies). Also, the gradient of g_A is called the *momentum map*. Recall that the Veronese embedding takes values in projective space. We will use the following notation: $v_A(p) = \hat{v}_A(p)/\|\hat{v}_A(p)\|$. This is independent of the representative of equivalence class $v_A(p)$. Now, let $v_A(p)^2$ mean coordinatewise squaring and $v_A(p)^{2T}$ be the transpose of $v_A(p)^2$. The gradient of g_A is then:

Formula 2.1.2: The Momentum Map associated to A

$$\nabla g_A = v_A(p)^{2T} A$$

Since $p \mapsto v_A(p)$ is a well-defined real function, we may write its derivative as

$$D v_A(p) = P_{v_A(p)} \mathrm{Diag}(v_A(p)) A$$

where P_v is the projection operator $I - \frac{vv^H}{\|v\|^2}$.

Then the second derivative of g_A is

Formula 2.1.3: Second derivative of g_A

$$D^2 g_A = 2 D v_A(p)^T D v_A(p)$$

Using the relation $-\frac{1}{2} dJ^* dg_A = \frac{1}{2} \sum (D^2 g_A)_{ij} dp_i \wedge dq_j$, one obtains an expression for ω_A:

Formula 2.1.4: The symplectic 2-form associated to A:

$$(\omega_A)_{(p,q)} = \frac{1}{2} \sum_{ij} (D^2 g_A)_{ij} dp_i \wedge dq_j$$

We still have to show that ω_A is a symplectic form. Clearly, $\frac{1}{2}d(dJ^* dg_A) = \frac{1}{2}d^2(J^* dg_A) = 0$. The remaining condition to check is non-degeneracy. In view of formulæ 2.1.3 and 2.1.4, this is a consequence of the following fact:

Lemma 1. *Let A be a matrix with non-negative integer entries, such that* $\mathrm{Conv}(A)$ *has dimension n. Then $(Dv_A)_p$ is injective, for all $p \in \mathbb{R}^n$.*

Proof. The conclusion of this Lemma can fail only if there are $p \in \mathbb{R}^n$ and $u \neq 0$ with $(Dv_A)_p u = 0$. This means that

$$P_{v_A(p)} \mathrm{diag}(v_A)_p A u = 0 \quad .$$

This can only happen if $\mathrm{diag}(v_A)_p A u$ is in the space spanned by $(v_A)_p$, or, equivalently, Au is in the space spanned by $(1, 1, \cdots, 1)^T$. This means that all the rows a of A satisfy $au = \lambda$ for some λ. Interpreting a row of A as a vertex of $\mathrm{Conv}A$, this means that $\mathrm{Conv}A$ is contained in the affine plane $\{a : au = \lambda\}$. ∎

We can also write down the Hermitian structure of $\mathcal{T}^n$ as:

Formula 2.1.5: Hermitian structure of $\mathcal{T}^n$ associated to A:

$$(\langle u, w \rangle_A)_{(p,q)} = u^H (\frac{1}{2} D^2 g_A)_p w$$

In general, the function v_A goes from $\mathcal{T}^n$ into projective space. Therefore, its derivative is a mapping

$$(Dv_A)_{(p,q)} : T_{(p,q)}\mathcal{T}^n \rightarrow T_{v_A(p+q\sqrt{-1})}\mathbb{P}^{M-1} \simeq \hat{v}_A(p + q\sqrt{-1})^\perp \subset \mathbb{C}^M \quad .$$

For convenience, we will write this derivative as a mapping into $\mathbb{C}^M$, with range $\hat{v}_A(p + q\sqrt{-1})^\perp$. Let P_v be the projection operator

$$P_v = I - \frac{1}{\|v\|^2} v v^H \quad .$$

Then,

Formula 2.1.6: Derivative of v_A

$$(Dv_A)_{(p,q)} = P_{\hat{v}_A(p+q\sqrt{-1})} \mathrm{Diag}\left(\frac{\hat{v}_A(p+q\sqrt{-1})}{\|\hat{v}_A(p+q\sqrt{-1}\|}\right) A$$

An immediate consequence of Formula 2.1.6 is:

Lemma 2. *Let $f \in \mathcal{F}_A$ and $(p, q) \in \mathcal{T}^n$ be such that $f \cdot \hat{v}_A(p + q\sqrt{-1}) = 0$. Then, $f \cdot (Dv_A)_{(p,q)} = \frac{1}{\|\hat{v}_A(p,q)\|}(D\hat{v}_A)_{(p,q)}$*

In other words, when $(f \circ \exp)(p + q\sqrt{-1})$ vanishes, Dv_A and $D\hat{v}_A$ are the same up to scaling.

Notice that the Hermitian metric is also

$$(\langle u, w\rangle_A)_{(p,q)} = u^h Dv_A(p, q)^H Dv_A(p, q)w \quad .$$

Finally, the volume element associated to A is

Formula 2.1.7: Volume element of $(\mathcal{T}^n, \omega_A, J)$

$$d\mathcal{T}^n_A = \det\left(\frac{1}{2}\ D^2 g_A(p)\right)\ dp_1 \wedge \cdots \wedge dp_n \wedge dq_1 \wedge \cdots \wedge dq_n$$

2.2 Toric Actions and the Momentum Map

The *momentum map*, also called *moment map*, was introduced in its modern formulation by Smale [?] and Souriau [?]. The reader may consult one of the many textbooks in the subject (such as Abraham and Marsden [?] or McDuff and Salamon [?]) for a general exposition.

In appendix A, we will explicitly show what Lie group action ∇g_A is the momentum of, and what the associated Hamiltonian dynamical system is.

In this Section we instead follow the point of view of Gromov [?]. The main results in this Section are that

Proposition 1. *The momentum map ∇g_A maps $\mathcal{T}^n$ onto the interior of* $\mathrm{Conv}(A)$*. When ∇g_A is restricted to the real n-plane $[q = 0] \subset \mathcal{T}^n$, this mapping is a bijection.* ■

This seems to be a particular case of the Atiyah-Guillemin-Sternberg theorem. However, technical difficulties prevent us from directly applying this result here (see appendix A).

Proposition 2. *The momentum map ∇g_A is a volume-preserving map from the manifold $(\mathcal{T}^n, \omega_A, J)$ into* $\mathrm{Conv}(A)$*, up to a constant, in the following sense: if U is a measurable region of* $\mathrm{Conv}(A)$*, then*

$$\mathrm{Vol}\left((\nabla g_A)^{-1}(U)\right) = \pi^n \ \mathrm{Vol} U \quad .$$

We prove Proposition 2 by first assuming Proposition 1.

Proof of Proposition 2. Consider the mapping

$$\begin{array}{rccc} M: & \mathcal{T}^n & \to & \frac{1}{2}\mathrm{Conv}(A) \times \mathbb{T}^n \\ & (p, q) & \mapsto & (\frac{1}{2}\nabla g_A(p), q) \end{array} \quad .$$

Since we assume $\dim \mathrm{Conv}(A) = n$, we can apply Proposition 1 and conclude that M is a diffeomorphism.

The pull-back of the canonical symplectic structure in $\mathbb{R}^{2n}$ by M is precisely ω_A, because of Formulæ 2.1.3 and 2.1.4. Diffeomorphisms with that property are called *symplectomorphisms*. Since the volume form of a symplectic manifold depends only of the canonical 2-form, symplectomorphisms preserve volume. We compose with a scaling by $\frac{1}{2}$ in the first n variables, that divides $\mathrm{Vol} U$ by 2^n, and we are done. ■

Remark 6. Symplectomorphisms are also known to preserve a few other invariants such as the symplectic width (see [**?**]). However, symplectomorphisms are not required to preserve the complex structure and therefore need not be isometries.

However, it is explained in [**?**] how to define a new complex structure in $\mathrm{Conv}(A) \times \mathbb{T}^n$ that will make the map M a Kähler isomorphism, hence an isometry. ■

Before proving Proposition 1, we will need the following result about convexity. We follow here Convexity Theorem 1.2 in [**?**], attributed to Legendre:

Theorem (Legendre). *If f is convex and of class $\mathcal{C}^2$ on $\mathbb{R}^n$, then the closure of the image $\{\nabla f_r : r \in \mathbb{R}^n\}$ in $(R^n)^\vee$ is convex.*

Proof. Let L_f be the set of covectors $y \in (\mathbb{R}^n)^\vee$ with the property that

$$\exists c \in \mathbb{R} \ \forall x \in \mathbb{R}^n f(x) \geq y \cdot x - c \quad .$$

Notice that L_f is a convex subset of $(\mathbb{R}^n)^\vee$. Geometrically, the planes in L_f with c minimal correspond to the *envelope* of the graph of f.

The set L_f contains $\{\nabla f_r : r \in \mathbb{R}^n\}$: For any given r, we set $c_r = \nabla f_r \cdot r - f(r)$. Since f is convex,

$$f(x) \geq \nabla f_r \cdot x - c_r \quad .$$

For the converse, assume that there is $y \in L_f$ not in the closure of $\{\nabla f_r : r \in \mathbb{R}^n\}$. Then there is some $\varepsilon > 0$ such that

$$\forall r \in \mathbb{R}^n, \|y - \nabla f_r\| > \varepsilon \quad .$$

We define the following gradient vector field in $\mathbb{R}^n$:

$$\dot{x} = \frac{(y - \nabla f_x)^T}{\|y - \nabla f_x\|}$$

Because the denominator is bounded below by ε, this field is well-defined and Lispchitz in all of $\mathbb{R}^n$. Let us fix an arbitrary initial condition $x(0) \in \mathbb{R}^n$, and let $x(t)$ denote a maximal solution of the vector field. Since the vector field has norm 1, $x(t)$ cannot diverge in finite time and therefore $x(t)$ is well-defined for all $t \in \mathbb{R}$.

Now we look at the function $t \mapsto y \cdot x(t) - f(x(t))$. Its derivative w.r.t. t is $(y - \nabla f_{x(t)})\dot{x}(t) > \varepsilon$. Therefore, $\lim_{t\to\infty} y \cdot x(t) - f(x(t)) = \infty$. We deduce from there that $\sup_{r\in\mathbb{R}^n} y \cdot r - f(r) = \infty$ Hence, $y \notin L_f$, a contradiction. ■

By replacing f by g_A, we conclude that the image of the momentum map ∇g_A is convex.

Proof of Proposition 1. The momentum map ∇g_A maps $\mathcal{T}^n$ onto the interior of Conv$A$. Indeed, let $a = A^\alpha$ be a row of A, associated to a vertex of Conv$A$. Then there is a direction $v \in \mathbb{R}^n$ such that

$$a \cdot v = \max_{x \in \mathrm{Conv}A} x \cdot v$$

for some unique a.

We claim that $a \in \overline{\nabla g_A(\mathbb{R}^n)}$. Indeed, let $x(t) = v_A(tv)$, t a real parameter. If b is another row of A,

$$e^{a \cdot tv} = e^{ta \cdot v} \gg e^{tb \cdot v} = e^{b \cdot tv}$$

as $t \to \infty$. We can then write $\hat{v}_A(tv)^{2T}$ as:

$$\hat{v}_A(tv) = \begin{bmatrix} \vdots \\ e^{ta \cdot v} \\ \vdots \end{bmatrix}^T C \mathrm{Diag} \begin{bmatrix} \vdots \\ e^{ta \cdot v} \\ \vdots \end{bmatrix} \quad .$$

Since C is positive definite, $C_{\alpha\alpha} > 0$ and

$$\lim_{t \to \infty} v_A(tv)^{2T} = \lim_{t \to \infty} \frac{\hat{v}_A(tv)^{2T}}{\|\hat{v}_A(tv)\|^2} = \mathrm{e}_a^T \frac{C_{\alpha\alpha}}{C_{\alpha\alpha}} = \mathrm{e}_a^T \quad ,$$

where e_a is the unit vector in $\mathbb{R}^M$ corresponding to the row a. It follows that $\lim_{t\to\infty} \nabla g_A(tv) = a$

When we set $q = 0$, we have $\det D^2 g_A \neq 0$ on $\mathbb{R}^n$, so we have a local diffeomorphism at each point $p \in \mathbb{R}^n$. Assume that $(\nabla g_A)_p = (\nabla g_A)_{p'}$ for $p \neq p'$. Then, let $\gamma(t) = (1-t)p + tp'$. The function $t \mapsto (\nabla g_A)_{\gamma(t)} \gamma'(t)$ has the same value at 0 and at 1, hence by Rolle's Theorem its derivative must vanish at some $t^* \in (0, 1)$.

In that case,

$$(D^2 g_A)_{\gamma(t^*)}(\gamma'(t^*), \gamma'(t^*)) = 0$$

and since $\gamma'(t^*) = p' - p \neq 0$, $\det D^2 g_A$ must vanish in some $p \in \mathbb{R}^n$. This contradicts Lemma 1. ■

2.3 More Properties of the Momentum Map

We can also give an interpretation of the derivative Dv_A in terms of the momentum map (see figure 1).

Lemma 3.

$$(Dv_A^\alpha)_p u = |(v_A^\alpha)_p| \, (A^\alpha - \nabla g_A(p)) \cdot u \quad .$$

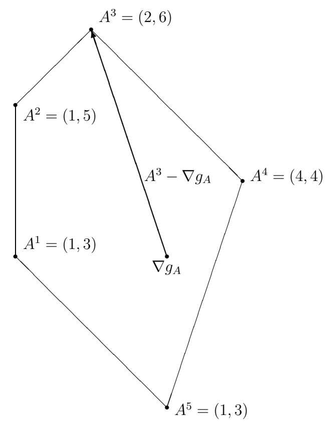

Figure 1: Geometric interpretation of Dv_A^α

where $|(v_A^\alpha)_p|$ stands for $\frac{|(\hat{v}_A^\alpha)_p|}{\|(\hat{v}_A)_p\|}$, and where A^α and ∇g_A are co-vectors.

Proof. By formula 2.1.6,

$$(Dv_A)_p u = \operatorname{diag} v_A(p) A u - v_A(p) v_A(p)^T \operatorname{diag}\left(v_A(p)\right) A u \quad .$$

Hence its α-th coordinate is:

$$(Dv_A^\alpha)_p u = (v_A^\alpha)_p \left(A^\alpha u - (v_A)_p^{2T} A u\right) = (v_A^\alpha)_p (A^\alpha - \nabla g_A(p)) u \quad .$$

■

Bearing in mind that $\sum v_A^\alpha(p)^2 = 1$, we obtain an immediate consequence:

Lemma 4. *For all* $(p,q) \in \mathcal{T}^n$,

$$\|Dv_A(p,q)\| \leq \operatorname{diam}(\operatorname{Conv} A) \quad \textit{and}$$

$$\|\frac{1}{2} D^2 g_A(p)\| \leq (\operatorname{diam}(\operatorname{Conv} A))^2$$

2.4 Evaluation Map and Condition Matrix

In the setting of Theorem 1, we can identify each space of polynomials $(\mathcal{F}_{A_i}, \langle\cdot,\cdot\rangle_{A_i})$ to the (co)vector space $(\mathbb{C}^{M_i})^\vee$, endowed with the canonical inner product. The value of f_i at $\exp(p + q\sqrt{-1})$ is then precisely $f_i \cdot \hat{v}_{A_i}(p + q\sqrt{-1})$.

More generally, we can define the *evaluation map* by

$$\begin{array}{rcl} ev: \ (\mathcal{F}_{A_1} \times \cdots \times \mathcal{F}_{A_n}) \times \mathcal{T}^n & \rightarrow & \mathbb{C}^n \\ (f_1, \cdots, f_n; p + q\sqrt{-1}) & \mapsto & \begin{bmatrix} f^1 \cdot \hat{v}_{A_1}(p + q\sqrt{-1}) \\ \vdots \\ f^n \cdot \hat{v}_{A_n}(p + q\sqrt{-1}) \end{bmatrix} \end{array} .$$

Following [?], we look at the linearization of the implicit function $p + q\sqrt{-1} = G(f)$ for the equation $ev(f, p + q\sqrt{-1}) = 0$.

Definition 5. The *condition matrix* of ev at $(f, p + q\sqrt{-1})$ is

$$DG = D_{\mathcal{T}^n}(ev)^{-1} D_{\mathcal{F}}(ev) \quad ,$$

where $\mathcal{F} = \mathcal{F}_{A_1} \times \cdots \times \mathcal{F}_{A_n}$.

Above, $D_{\mathcal{T}^n}(ev)$ is a linear operator from an n-dimensional complex space into $\mathbb{C}^n$, while $D_{\mathcal{F}}(ev)$ goes from an $M_1 + \cdots + M - n$-dimensional complex space into $\mathbb{C}^n$.

Lemma 5. *Assume that* $ev(f; p + q\sqrt{-1}) = 0$. *Then,*

$$\det \left(DG DG^H\right)^{-1} dp_1 \wedge dq_1 \wedge \cdots \wedge dp_n \wedge dq_n = (-1)^{n(n-1)/2} \bigwedge \sqrt{-1} f^i \cdot (Dv_{A_i})_{(p,q)} dp \wedge \\ \wedge \bar{f}^i \cdot (Dv_{A_i})_{(p,-q)} dq \quad .$$

Note that although $f^i \cdot (Dv_{A_i})_{(p,q)} dp$ is a complex-valued form, each wedge $f^i \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}^i \cdot (Dv_{A_i})_{(p,-q)} dq$ is a real-valued 2-form.

Proof. We compute:

$$D_{\mathcal{F}}(ev)|_{(p,q)} = \begin{bmatrix} \sum_{\alpha=1}^{M_1} \hat{v}_{A_1}^{\alpha}(p + q\sqrt{-1}) df_{\alpha}^1 \\ \vdots \\ \sum_{\alpha=1}^{M_n} \hat{v}_{A_n}^{\alpha}(p + q\sqrt{-1}) df_{\alpha}^n \end{bmatrix} \quad ,$$

and hence

$$D_{\mathcal{F}}(ev) D_{\mathcal{F}}(ev)^H = \operatorname{diag} \|\hat{v}_{A_i}\|^2 \quad .$$

Also,

$$D_{\mathcal{T}^n}(ev) = \begin{bmatrix} f^1 \cdot D\hat{v}_{A_1} \\ \vdots \\ f^n \cdot D\hat{v}_{A_n} \end{bmatrix} \quad .$$

Therefore,

$$\det \left(DG_{(p,q)} DG_{(p,q)}^H \right)^{-1} = \left| \det \begin{bmatrix} f^1 \cdot \frac{1}{\|\hat{v}_{A_1}\|} D\hat{v}_{A_1} \\ \vdots \\ f^n \cdot \frac{1}{\|\hat{v}_{A_n}\|} D\hat{v}_{A_n} \end{bmatrix} \right|^2 \quad .$$

We can now use Lemma 2 to conclude the following:

Formula 2.4.1: Determinant of the Condition Matrix

$$\det \left(DG_{(p,q)} DG_{(p,q)}^H \right)^{-1} = \left| \det \begin{bmatrix} f^1 \cdot Dv_{A_1} \\ \vdots \\ f^n \cdot Dv_{A_n} \end{bmatrix} \right|^2$$

We can now write the same formula as a determinant of a block matrix:

$$\det \left(DG_{(p,q)} DG_{(p,q)}^H \right)^{-1} = \det \begin{bmatrix} f^1 \cdot Dv_{A_1} & \\ \vdots & \\ f^n \cdot Dv_{A_n} & \\ & \bar{f}^1 \cdot D\bar{v}_{A_1} \\ & \vdots \\ & \bar{f}^n \cdot D\bar{v}_{A_n} \end{bmatrix}$$

and replace the determinant by a wedge. The factor $(-1)^{n(n-1)/2}$ comes from replacing $dp_1 \wedge \cdots \wedge dp_n \wedge dq_1 \wedge \cdots \wedge dq_n$ by $dp_1 \wedge dq_1 \wedge \cdots \wedge dp_n \wedge dq_n$. ■

Proof of Theorem 1. Given $(p, q) \in \mathcal{T}^n$, we define $\mathcal{F}_{(p,q)}$ as the space of $f \in \mathcal{F}_{A_1} \times \cdots \times \mathcal{F}_{A_n}$ such that $ev(f; p + q\sqrt{-1})$ vanishes.

Using [**?**, Theorem 5 p. 243] (or Proposition 5 p. 42 below), we deduce that the average number of complex roots is:

$$\mathrm{Avg} = \int_{(p,q)\in U} \int_{f\in\mathcal{F}_{(p,q)}} \left(\prod \frac{e^{-\|f_i\|^2/2}}{(2\pi)^{M_i}} \right) \det \left(DG_{(p,q)} DG_{(p,q)}^H \right)^{-1} \quad .$$

By Lemma 5, we can replace the inner integral by a $2n$-form valued integral:

$$\mathrm{Avg} = (-1)^{n(n-1)/2} \int_{(p,q)\in U} \int_{f\in\mathcal{F}_{(p,q)}} \bigwedge_i \frac{e^{-\|f_i\|^2/2}}{(2\pi)^{M_i}} f^i (Dv_{A_i})_{(p,q)} dp \wedge \\ \wedge \bar{f}^i (Dv_{A_i})_{(p,-q)} dq \quad .$$

Since the image of Dv_{A_i} is precisely $\mathcal{F}_{A_i}|_{(p,q)} \subset \mathcal{F}_{A_i}$, one can add n extra variables corresponding to the directions $v_{A_i}(p+q\sqrt{-1})$ without changing the integral: we write $\mathcal{F}_{A_i} = \mathcal{F}_{A_i,(p,q)} \times \mathbb{C} v_{A_i}(p+q\sqrt{-1})$. Since $\left(f^i + t v_{A_i}(p+q\sqrt{-1})\right) Dv_{A_i}$ is equal to $f^i Dv_{A_i}$, the average number of roots is indeed:

$$\mathrm{Avg} = (-1)^{n(n-1)/2} \int_{(p,q)\in U} \int_{f\in\mathcal{F}} \bigwedge_i \frac{e^{-\|f_i\|^2/2}}{(2\pi)^{M_i+1}} f^i \cdot (Dv_{A_i})_{(p,q)} dp \wedge \\ \wedge \bar{f}^i \cdot (Dv_{A_i})_{(p,-q)} dq \quad .$$

In the integral above, all the terms that are multiple of $f^i_\alpha \bar{f}^i_\beta$ for some $\alpha \neq \beta$ will cancel out. Therefore,

$$\mathrm{Avg} = (-1)^{n(n-1)/2} \int_{(p,q)\in U} \int_{f\in\mathcal{F}} \bigwedge_i \frac{e^{-\|f_i\|^2/2}}{(2\pi)^{M_i+1}} \sum_\alpha |f^i_\alpha|^2 (Dv_{A_i})^\alpha_{(p,q)} dp \wedge \\ \wedge (Dv_{A_i})^\alpha_{(p,-q)} dq \quad .$$

Now, we apply the integral formula:

$$\int_{x\in\mathbb{C}^M} |x_1|^2 \frac{e^{-\|x\|^2/2}}{(2\pi)^M} = \int_{x_1\in\mathbb{C}} |x_1|^2 \frac{e^{-|x_1|^2/2}}{2\pi} = 2$$

to obtain:

$$\mathrm{Avg} = \frac{(-1)^{n(n-1)/2}}{\pi^n} \int_{(p,q)\in U} \bigwedge \sum_\alpha (Dv_{A_i})^\alpha_{(p,q)} dp \wedge (Dv_{A_i})^\alpha_{(p,-q)} dq \quad .$$

According to formulæ 2.1.3 and 2.1.4, the integrand is just $2^{-n} \bigwedge \omega_{A_i}$, and thus

$$\mathrm{Avg} = \frac{(-1)^{n(n-1)/2}}{\pi^n} \int_U \bigwedge_i \omega_{A_i} = \frac{n!}{\pi^n} \int_U d\mathcal{T}^n \quad .$$

■

3 The Condition Number

3.1 Proof of Theorem 2

Let $(p,q) \in \mathcal{T}^n$ and let $f \in \mathcal{F}_{(p,q)}$. Without loss of generality, we can assume that f is scaled so that for all i, $\|f^i\| = 1$.

Let $\delta f \in \mathcal{F}_{(p,q)}$ be such that $f + \delta f$ is singular at (p,q), and assume that $\sum \|\delta f^i\|^2$ is minimal. Then, due to the scaling we chose,

$$d_{\mathbb{P}}(f, \Sigma_{(p,q)}) = \sqrt{\sum \|\delta f^i\|^2} \quad .$$

Since $f + \delta f$ is singular, there is a vector $u \neq 0$ such that

$$\begin{bmatrix} (f^1 + \delta f^1) \cdot (D\hat{v}_{A_1})_{(p,q)} \\ \vdots \\ (f^n + \delta f^n) \cdot (D\hat{v}_{A_n})_{(p,q)} \end{bmatrix} u = 0$$

and hence

$$\begin{bmatrix} (f^1 + \delta f^1) \cdot (Dv_{A_1})_{(p,q)} \\ \vdots \\ (f^n + \delta f^n) \cdot (Dv_{A_n})_{(p,q)} \end{bmatrix} u = 0 \quad .$$

This means that

$$\left\{ \begin{array}{rcl} f^1 \cdot Dv_{A_1} u & = & -\delta f^1 \cdot Dv_{A_1} u \\ & \vdots & \\ f^n \cdot Dv_{A_n} u & = & -\delta f^n \cdot Dv_{A_n} u \end{array} \right. \quad .$$

Let $D(f)$ denote the matrix

$$D(f) \stackrel{\text{def}}{=} \begin{bmatrix} f^1 \cdot (Dv_{A_1})_{(p,q)} \\ \vdots \\ f^n \cdot (Dv_{A_n})_{(p,q)} \end{bmatrix} \quad .$$

Given $v = D(f)\ u$, we obtain:

$$\begin{cases} v_1 &= -\delta f^1 \cdot Dv_{A_1} D(f)^{-1} v \\ &\vdots \\ v_n &= -\delta f^n \cdot Dv_{A_n} D(f)^{-1} v \end{cases} \tag{3.1.1}$$

We can then scale u and v, such that $\|v\| = 1$.

Claim. *Under the assumptions above, δf^i is colinear to $\left(Dv_{A_i} D(f)^{-1} v\right)^H$.*

Proof. Assume that $\delta f^i = g + h$, with g colinear and h orthogonal to $\left(Dv_{A_i} D(f)^{-1} v\right)^H$. As the image of Dv_{A_i} is orthogonal to v_{A_i}, g is orthogonal to $v_{A_i}^H$, so $ev(g^i, (p,q)) = 0$ and hence $ev(h^i, (p,q)) = 0$. We can therefore replace δf^i by g without compromising equality (3.1.1). Since $\|\delta f\|$ was minimal, this implies $h = 0$. ■

We obtain now an explicit expression for δf^i in terms of v:

$$\delta f^i = -v_i \frac{\left(Dv_{A_i} D(f)^{-1} v\right)^H}{\|Dv_{A_i} D(f)^{-1} v\|^2} \quad .$$

Therefore,

$$\|\delta f^i\| = \frac{|v_i|}{\|Dv_{A_i} D(f)^{-1} v\|} = \frac{|v_i|}{\|\left(D(f)^{-1} v\right)\|_{A_i}} \quad .$$

So we have proved the following result:

Lemma 6. *Fix v so that $\|v\| = 1$ and let $\delta f \in \mathcal{F}_{(p,q)}$ be such that equation (3.1.1) holds and $\|\delta f\|$ is minimal. Then,*

$$\|\delta f^i\| = \frac{|v_i|}{\|D(f)^{-1} v\|_{A_i}} \quad .$$

Lemma 6 provides an immediate lower bound for $\|\delta f\| = \sqrt{\sum \|\delta f^i\|^2}$: Since

$$\|\delta f^i\| \geq \frac{|v_i|}{\max_j \|D(f)^{-1}v\|_{A_j}} \quad ,$$

we can use $\|v\| = 1$ to deduce that

$$\sqrt{\sum_i \|\delta f^i\|^2} \geq \frac{1}{\max_j \|D(f)^{-1}v\|_{A_j}} \geq \frac{1}{\max_j \|D(f)^{-1}\|_{A_j}} \quad .$$

Also, for any v with $\|v\| = 1$, we can choose δf minimal so that equation (3.1.1) applies. Using Lemma 6, we obtain:

$$\|\delta f^i\| \leq \frac{|v_i|}{\min_j \|D(f)^{-1}v\|_{A_j}} \quad .$$

Hence

$$\sqrt{\sum_i \|\delta f^i\|^2} \leq \frac{1}{\min_j \|D(f)^{-1}v\|_{A_j}} \quad .$$

Since this is true for any v, and $\|\delta f\|$ is minimal for all v, we have

$$\sqrt{\sum_i \|\delta f^i\|^2} \leq \frac{1}{\max_{\|v\|=1} \min_j \|D(f)^{-1}\|_{A_j}}$$

and this proves Theorem 2.

3.2 Idea of the Proof of Theorem 3

The proof of Theorem 3 is long. We first sketch the idea of the proof. Recall that $\mathcal{F}_{(p,q)}$ is the set of all $f \in \mathcal{F}$ such that $ev(f; p + q\sqrt{-1}) = 0$, and that $\Sigma_{(p,q)}$ is the restriction of the discriminant to the fiber $\mathcal{F}_{(p,q)}$:

$$\Sigma_{(p,q)} \stackrel{\text{def}}{=} \{f \in \mathcal{F}_{(p,q)} : D(f)_{(p,q)} \text{ does not have full rank}\} \quad .$$

The space $\mathcal{F}$ is endowed with a Gaussian probability measure, with volume element

$$\frac{e^{-\|f\|^2/2}}{(2\pi)^{\sum M_i}} d\mathcal{F} \quad ,$$

where $d\mathcal{F}$ is the usual volume form in $\mathcal{F} = (\mathcal{F}_{A_1}, \langle\cdot,\cdot\rangle_{A_1}) \times \cdots \times (\mathcal{F}_{A_n}, \langle\cdot,\cdot\rangle_{A_n})$ and $\|f\|^2 = \sum \|f^i\|^2_{A_i}$. For U a set in $\mathcal{T}^n$, we defined earlier (in the statement of Theorem 3) the quantity:

$$\nu^A(U,\varepsilon) \stackrel{\text{def}}{=} \text{Prob}[\boldsymbol{\mu}(f,U) > \varepsilon^{-1}] = \text{Prob}[\exists (p,q) \in U : d_{\mathbb{P}}(f, \Sigma_{(p,q)}) < \varepsilon] \quad .$$

The naïve idea for bounding $\nu^A(U,\varepsilon)$ is as follows: Let $V(\varepsilon) \stackrel{\text{def}}{=} \{(f,(p,q)) \in \mathcal{F} \times U : ev(f;(p,q)) = 0 \text{ and } d_{\mathbb{P}}(f,\Sigma_{(p,q)}) < \varepsilon\}$. We also define $\pi : V(\varepsilon) \rightarrow \mathcal{F}$ as the canonical projection mapping $\mathcal{F} \times U$ to $\mathcal{F}$, and set $\#_{V(\varepsilon)}(f) \stackrel{\text{def}}{=} \#\{(p,q) \in U : (f,(p,q)) \in V(\varepsilon)\}$. Then,

$$\begin{aligned} \nu^A(U,\varepsilon) &= \int_{f \in \mathcal{F}} \chi_{\pi(V(\varepsilon))}(f) \; \frac{e^{-\|f\|^2/2}}{(2\pi)^{\sum M_i}} d\mathcal{F} \\ &\leq \int_{f \in \mathcal{F}} \#_{V(\varepsilon)} \; \frac{e^{-\|f\|^2/2}}{(2\pi)^{\sum M_i}} d\mathcal{F} \end{aligned}$$

with equality in the linear case.

Now we apply the coarea formula [**?**, Theorem 5 p. 243] to obtain:

$$\nu^A(U,\varepsilon) \leq \int_{(p,q) \in U \subset \mathcal{T}^n} \int_{\substack{f \in \mathcal{F}_{(p,q)} \\ d_{\mathbb{P}}(f,\Sigma_{(p,q)}) < \varepsilon}} \frac{1}{NJ(f;(p,q))} \; \frac{e^{-\|f\|^2/2}}{(2\pi)^{\sum M_i}} d\mathcal{F} \; dV_{\mathcal{T}^n} \quad ,$$

where $dV_{\mathcal{T}^n}$ stands for Lebesgue measure in $\mathcal{T}^n$. Again, in the linear case, we have equality.

We already know from Lemma 5 that

$$1/NJ(;(p,q)) = \bigwedge_{i=1}^{n} f^i \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}^i \cdot (D\bar{v}_{A_i})_{(p,q)} dq \quad .$$

We should focus now on the inner integral. In each coordinate space $\mathcal{F}_{A_i}$, we can introduce a new orthonormal system of coordinates (depending on (p,q)) by decomposing:

$$f^i = f^i_{\mathrm{I}} + f^i_{\mathrm{II}} + f^i_{\mathrm{III}} \quad ,$$

where f^i_{I} is the component colinear to $v_{A_i}^H$, f^i_{II} is the projection of f^i to $(\text{range } Dv_{A_i})^H$, and f^i_{III} is orthogonal to f^i_{I} and f^i_{II}.

Of course, $f^i \in (\mathcal{F}_{A_i})_{(p,q)}$ if and only if $f^i_{\mathrm{I}} = 0$.

Also,

$$\bigwedge_{i=1}^{n} f^i \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}^i \cdot (D\bar{v}_{A_i})_{(p,q)} dq =$$
$$= \bigwedge_{i=1}^{n} f^i_{\mathrm{II}} \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}^i_{\mathrm{II}} \cdot (D\bar{v}_{A_i})_{(p,q)} dq \quad .$$

It is an elementary fact that

$$d_{\mathbb{P}}(f^i_{\mathrm{II}} + f^i_{\mathrm{III}}, \Sigma_{(p,q)}) \leq d_{\mathbb{P}}(f^i_{\mathrm{II}}, \Sigma_{(p,q)}) \quad .$$

It follows that for $f \in \mathcal{F}_{(p,q)}$:

$$d_{\mathbb{P}}(f, \Sigma_{(p,q)}) \leq d_{\mathbb{P}}(f_{\mathrm{II}}, \Sigma_{(p,q)}) \quad ,$$

with equality in the linear case. Hence, we obtain:

$$\nu^A(U, \varepsilon) \leq \int_{(p,q)\in U \subset \mathcal{T}^n} \int_{\substack{f \in \mathcal{F}_{(p,q)} \\ d_{\mathbb{P}}(f_{\mathrm{II}}, \Sigma_{(p,q)}) < \varepsilon}} \left(\bigwedge_{i=1}^{n} f^i_{\mathrm{II}} \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}^i_{\mathrm{II}} \cdot (D\bar{v}_{A_i})_{(p,q)} dq \right) \cdot$$
$$\cdot \frac{e^{-\|f^i_{\mathrm{II}} + f^i_{\mathrm{III}}\|^2/2}}{(2\pi)^{\sum M_i}} d\mathcal{F} \; dV_{\mathcal{T}^n} \quad ,$$

with equality in the linear case. We can integrate the $\sum(M_i - n - 1)$ variables f_{III} to obtain:

Proposition 3.

$$\nu^A(U, \varepsilon) \leq \int_{(p,q)\in U \subset \mathcal{T}^n} \int_{\substack{f_{\mathrm{II}} \in \mathbb{C}^{n^2} \\ d_{\mathbb{P}}(f_{\mathrm{II}}, \Sigma_{(p,q)}) < \varepsilon}} \left(\bigwedge_{i=1}^{n} f^i_{\mathrm{II}} \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}^i_{\mathrm{II}} \cdot (D\bar{v}_{A_i})_{(p,q)} dq \right) \cdot$$
$$\cdot \frac{e^{-\|f^i_{\mathrm{II}}\|^2/2}}{(2\pi)^{n(n+1)}} \; dV_{\mathcal{T}^n} \quad .$$

with equality in the linear case. ■

3.3 From Gaussians to Multiprojective Spaces

The domain of integration in Proposition 3 makes integration extremely difficult. In order to estimate the inner integral, we will need to perform a change of coordinates.

Unfortunately, the Gaussian in Proposition 3 makes that change of coordinates extremely hard, and we will have to restate Proposition 3 in terms of integrals over a product of projective spaces.

The domain of integration will be $\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}$. Translating an integral in terms of Gaussians to an integral in terms of projective spaces is not immediate, and we will use the following elementary fact about Gaussians:

Lemma 7. *Let $\varphi : \mathbb{C}^n \to \mathbb{R}$ be $\mathbb{C}^*$-invariant (in the sense of the usual scaling action). Then we can also interpret φ as a function from $\mathbb{P}^{n-1}$ into $\mathbb{R}$, and:*

$$\frac{1}{\mathrm{Vol}(\mathbb{P}^{n+1})} \int_{[x]\in\mathbb{P}^{n-1}} \varphi(x) d[x] = \int_{x\in\mathbb{C}^n} \varphi(x) \frac{e^{-\|x\|^2/2}}{(2\pi)^n} dx \quad ,$$

where, respectively, the natural volume forms on $\mathbb{P}^{n-1}$ and $\mathbb{C}^n$ are understood for each integral.

Now the integrand in Proposition 3 is not $\mathbb{C}^*$-invariant. This is why we will need the following formula:

Lemma 8. *Under the hypotheses of Lemma 7,*

$$\frac{1}{\mathrm{Vol}(\mathbb{P}^{n+1})} \int_{[x]\in\mathbb{P}^{n-1}} \varphi(x) d[x] = \frac{1}{2n} \int_{x\in\mathbb{C}^n} \|x\|^2 \varphi(x) \frac{e^{-\|x\|^2/2}}{(2\pi)^n} dx \quad .$$

where, respectively, the natural volume forms on $\mathbb{P}^{n-1}$ and $\mathbb{C}^n$ are understood for each integral.

Proof.

$$\int_{x\in\mathbb{C}^n} \|x\|^2 \varphi(x) \frac{e^{-\|x\|^2/2}}{(2\pi)^n} dx = \int_{\Theta\in S^{2n-1}} \int_{r=0}^{\infty} |r|^{2n+1} \varphi(\Theta) \frac{e^{-|r|^2/2}}{(2\pi)^n} dr d\Theta$$

$$= \int_{\Theta\in S^{2n-1}} \left(- \left[|r|^{2n} \frac{e^{-|r|^2/2}}{(2\pi)^n} \right]_0^\infty + 2n \int_{r=0}^{\infty} |r|^{2n-1} \frac{e^{-|r|^2/2}}{(2\pi)^n} dr \right) \varphi(\Theta) d\Theta$$

$$= 2n \int_{x \in \mathbb{C}^n} \varphi(x) \frac{e^{-\|x\|^2/2}}{(2\pi)^n} dx$$

■

We can now introduce the notation:

$$\mathrm{WEDGE}^A(f_{\mathrm{I\!I}}) \stackrel{\mathrm{def}}{=} \bigwedge_{i=1}^{n} \frac{1}{\|f_{\mathrm{I\!I}}^i\|^2} f_{\mathrm{I\!I}}^i \cdot (Dv_{A_i})_{(p,q)} dp \wedge \bar{f}_{\mathrm{I\!I}}^i \cdot (D\bar{v}_{A_i})_{(p,q)} dq \quad .$$

This function is invariant under the $(\mathbb{C}^*)^n$-action $\lambda \star f_{\mathrm{I\!I}} : f_{\mathrm{I\!I}} \mapsto (\lambda_1 f_{\mathrm{I\!I}}^1, \cdots, \lambda_n f_{\mathrm{I\!I}}^n)$.

We adopt the following conventions: $\mathcal{F}_{\mathrm{I\!I}} \subset \mathcal{F}$ is the space spanned by coordinates $f_{\mathrm{I\!I}}$ and $\mathbb{P}(\mathcal{F}_{\mathrm{I\!I}})$ is its quotient by $(\mathbb{C}^*)^n$.

We apply n times Lemma 8 and obtain:

Proposition 4. *Let* $\mathrm{VOL} \stackrel{\mathrm{def}}{=} \mathrm{Vol}(\mathbb{P}^{n-1})^n$. *Then,*

$$\nu^A(U, \varepsilon) \leq \frac{(2n)^n}{\mathrm{VOL}} \int_{(p,q) \in U \subset \mathcal{T}^n} \int_{\substack{f_{\mathrm{I\!I}} \in \mathbb{P}(\mathcal{F}_{\mathrm{I\!I}}) \\ d_{\mathbb{P}}(f_{\mathrm{I\!I}}, \Sigma_{(p,q)}) < \varepsilon}} \mathrm{WEDGE}^A(f_{\mathrm{I\!I}}) \; d\mathbb{P}(\mathcal{F}_{\mathrm{I\!I}}) \; dV_{\mathcal{T}^n}$$

and in the linear case,

$$\nu^{\mathrm{Lin}}(U, \varepsilon) = \frac{(2n)^n}{\mathrm{VOL}} \int_{(p,q) \in U \subset \mathcal{T}^n} \int_{\substack{g_{\mathrm{I\!I}} \in \mathbb{P}(\mathcal{F}_{\mathrm{I\!I}}^{\mathrm{Lin}}) \\ d_{\mathbb{P}}(g_{\mathrm{I\!I}}, \Sigma^{\mathrm{Lin}}_{(p,q)}) < \varepsilon}} \mathrm{WEDGE}^{\mathrm{Lin}}(g_{\mathrm{I\!I}}) \; d(P\mathcal{F}_{\mathrm{I\!I}}^{\mathrm{Lin}}) dV_{\mathcal{T}^n} \; ■$$

Now we introduce the following change of coordinates. Let $L \in GL(n)$ be such that the minimum in Definition 1 p. 8 is attained:

$$\begin{array}{rccl} \varphi: & \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1} & \to & \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1} \\ & f_{\mathrm{I\!I}} & \mapsto & g_{\mathrm{I\!I}} \stackrel{\mathrm{def}}{=} \varphi(f_{\mathrm{I\!I}}) \quad \text{, such} \\ & & & \text{that } g_{\mathrm{I\!I}}^i = f_{\mathrm{I\!I}}^i \cdot Dv_{A_i} L \quad . \end{array}$$

Without loss of generality, we scale L such that $\det L = 1$. The following property follows from the definition of WEDGE:

$$\mathrm{WEDGE}^A(f_{\mathrm{I\!I}}) = \mathrm{WEDGE}^{\mathrm{Lin}}(g_{\mathrm{I\!I}}) \prod_{i=1}^{n} \frac{\|g_{\mathrm{I\!I}}^i\|^2}{\|f_{\mathrm{I\!I}}^i\|^2} \tag{3.3.1}$$

Assume now that $d_{\mathbb{P}}(f_{\mathrm{I\!I}}, \Sigma_{(p,q)}) < \varepsilon$. Then there is $\delta f \in \mathcal{F}_{\mathrm{I\!I}}$, such that $f + \delta f \in \Sigma^{\mathrm{Lin}}_{(p,q)}$ and $\|\delta f\| \le \varepsilon$ (assuming the scaling $\|f^i_{\mathrm{I\!I}}\| = 1$ for all i).

Setting $g_{\mathrm{I\!I}} = \varphi(f_{\mathrm{I\!I}})$ and $\delta g = \varphi(g)$, we obtain that $g + \delta g \in \Sigma^{\mathrm{Lin}}_{(p,q)}$.

$$d_{\mathbb{P}}(g, \Sigma^{\mathrm{Lin}}_{(p,q)}) \le \sqrt{\sum_{i=1}^{n} \frac{\|\delta g^i\|^2}{\|g^i_{\mathrm{I\!I}}\|^2}}$$

At each value of i,

$$\frac{\|\delta g^i\|}{\|g^i_{\mathrm{I\!I}}\|} \le \frac{\|\delta f^i\|}{\|f^i_{\mathrm{I\!I}}\|} \kappa(D_{f^i_{\mathrm{I\!I}}} \varphi^i)$$

where κ denotes Wilkinson's condition number of the linear operator $D_{f^i_{\mathrm{I\!I}}} \varphi^i$. This is precisely $\kappa(Dv_{A_i} L)$. Thus,

$$d_{\mathbb{P}}(g, \Sigma^{\mathrm{Lin}}_{(p,q)}) \le \varepsilon \max_i \kappa(Dv_{A_i} L) = \max_i \sqrt{\kappa(\omega_{A_i})}$$

Thus, an ε-neighborhood of $\Sigma^{A}_{(p,q)}$ is mapped into a $\sqrt{\kappa_U}\varepsilon$ neighborhood of $\Sigma^{\mathrm{Lin}}_{(p,q)}$.

We use this property and equation (3.3.1) to bound:

$$\nu^A(U, \varepsilon) \le \frac{(2n)^n}{\mathrm{VOL}} \int_{(p,q) \in U \subset \mathcal{T}^n} \int_{\substack{g_{\mathrm{I\!I}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1} \\ d_{\mathbb{P}}(g_{\mathrm{I\!I}}, \Sigma^{\mathrm{Lin}}_{(p,q)}) < \sqrt{\kappa_U}\varepsilon}} \mathrm{WEDGE}^{\mathrm{Lin}}(g_{\mathrm{I\!I}}) \cdot$$
$$\cdot \prod_{i=1}^{n} \frac{\|g^i_{\mathrm{I\!I}}\|^2}{\|f^i_{\mathrm{I\!I}}\|^2} |J_{g_{\mathrm{I\!I}}} \varphi^{-1}|^2 \; d(\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}) \; dV_{\mathcal{T}^n} \quad (3.3.2)$$

where $J_{g_{\mathrm{I\!I}}} \varphi^{-1}$ is the Jacobian of φ^{-1} at $g_{\mathrm{I\!I}}$.

Remark 7. Considering each Dv_{A_i} as a map from $\mathbb{C}^n$ into $\mathbb{C}^n$, the Jacobian is:

$$J_{g_{\mathrm{I\!I}}} \varphi^{-1} = \prod_{i=1}^{n} \frac{\|\varphi^{-1}(g_{\mathrm{I\!I}})^i\|^n}{\|g^i_{\mathrm{I\!I}}\|^n} \left(\det Dv^H_{A_i} Dv_{A_i} \right)^{-1/2} \quad .$$

We will not use this value in the sequel. ■

In order to simplify the expressions for the bound on $\nu^A(U, \varepsilon)$, it is convenient to introduce the following notations:

$$\begin{aligned}
dP &\stackrel{\text{def}}{=} \frac{(2n)^n}{\text{VOL}} \text{WEDGE}^{\text{Lin}}(g_{\text{II}}) \frac{d(\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1})}{n!\ (\omega_{\text{Lin}})^{\wedge n}} \\
H &\stackrel{\text{def}}{=} \prod_{i=1}^{n} \frac{\|g_{\text{II}}^i\|^2}{\|f_{\text{II}}^i\|^2} |J_g \varphi^{-1}|^2 \\
\chi_\delta &\stackrel{\text{def}}{=} \chi_{\left\{g : d_{\mathbb{P}}(g, \Sigma^{\text{Lin}}_{(p,q)}) < \delta\right\}}
\end{aligned}$$

Now equation (3.3.2) becomes:

$$\nu^A(U, \varepsilon) \le n! \int_{(p,q) \in U \subset \mathcal{T}^n} (\omega_{\text{Lin}})^{\wedge n} \int_{g_{\text{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP\ H(g_{\text{II}})\ \chi_{\sqrt{\kappa_U}\varepsilon}(g_{\text{II}}) \tag{3.3.3}$$

Lemma 9. *Let (p, q) be fixed. Then $\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}$ together with density function dP, is a probability space.*

Proof. The expected number of roots in U for a linear system is

$$n! \int_{(p,q) \in U} \omega_{\text{Lin}}^{\wedge n} \int_{g_{\text{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP \quad .$$

It is also $n! \int_U \omega_{\text{Lin}}^{\wedge n}$. This holds for all U, hence the volume forms are the same and

$$\int_{g_{\text{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP = 1 \quad .$$

■

This allows us to interpret the inner integral of equation (3.3.3) as the expected value of a product. This is less than the product of the expected values, and:

$$\begin{aligned}
\nu^A(U, \varepsilon) \le n! \int_{(p,q) \in U \subset \mathcal{T}^n} (\omega_{\text{Lin}})^{\wedge n} &\left(\int_{g_{\text{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP\ H(g_{\text{II}}) \right) \cdot \\
&\cdot \left(\int_{g_{\text{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP\ \chi_{\sqrt{\kappa_U}\varepsilon}(g_{\text{II}}) \right)
\end{aligned}$$

Because generic systems of linear equations have one root, we can also consider U as a probability space, with probability measure $\frac{1}{\mathrm{Vol}^{\mathrm{Lin}} U} n! \omega_{\mathrm{Lin}}^{\bigwedge n}$. Therefore, we can bound:

$$\nu^A(U, \varepsilon) \leq \frac{1}{\mathrm{Vol}^{\mathrm{Lin}} U} \left(\int_{(p,q) \in U} n! (\omega_{\mathrm{Lin}})^{\bigwedge n} \int_{g_{\mathrm{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP \ H(g_{\mathrm{II}}) \right) \cdot \cdot \left(\int_{(p,q) \in U} n! (\omega_{\mathrm{Lin}})^{\bigwedge n} \int_{g_{\mathrm{II}} \in \mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}} dP \ \chi_{\sqrt{\kappa_U} \varepsilon}(g_{\mathrm{II}}) \right)$$

The first parenthesis is $\mathrm{Vol}^A(U)$. The second parenthesis is $\nu^{\mathrm{Lin}}(\sqrt{\kappa_U}\varepsilon, U)$. This concludes the proof of Theorem 3.

Proof of Corollary 3.1. We set $L = Dv_{A_i}^{\dagger} | \mathrm{range} Dv_{A_i}$, then $\kappa(\omega_{A_1}, \cdots, \omega_{A_n}; (p,q)) = 1$. ∎

4 Real Polynomials

4.1 Proof of Theorem 4

Proof of Theorem 4. As in the complex case (Theorem 1), the expected number of roots can be computed by applying the co-area formula:

$$AVG = \int_{p \in U} \int_{f \in \mathcal{F}_p^{\mathbb{R}}} \prod_{i=1}^{n} \frac{e^{-\|f^i\|^2/2}}{\sqrt{2\pi}^{M_i}} \det(DG \ DG^H)^{-1/2} \quad .$$

Now there are three big diferences. The set U is in $\mathbb{R}^n$ instead of $\mathcal{T}^n$, the space $\mathcal{F}_p^{\mathbb{R}}$ contains only real polynomials (and therefore has half the dimension), and we are integrating the square root of $1/\det(DG \ DG^H)$.

Since we do not know in general how to integrate such a square root, we bound the inner integral as follows. We consider the real Hilbert space of functions integrable in $\mathcal{F}_p^{\mathbb{R}}$ endowed with Gaussian probability measure. The inner product in this space is:

$$\langle \varphi, \psi \rangle \stackrel{\mathrm{def}}{=} \int_{\mathcal{F}_p^{\mathbb{R}}} \varphi(f) \psi(f) \prod_{i=1}^{n} \frac{e^{-\|f^i\|^2/2}}{\sqrt{2\pi}^{M_i - 1}} dV \quad ,$$

where dV is Lebesgue volume. If $\mathbf{1}$ denotes the constant function equal to 1, we interpret

$$AVG = \int_{p \in U} (2\pi)^{-n/2} \left\langle \det(DG\ DG^H)^{-1/2}, \mathbf{1} \right\rangle \quad .$$

Hence Cauchy-Schwartz inequality implies:

$$AVG \leq \int_{p \in U} (2\pi)^{-n/2} \| \det(DG\ DG^H)^{-1/2} \| \|\mathbf{1}\| \quad .$$

By construction, $\|\mathbf{1}\| = 1$, and we are left with:

$$AVG \leq \int_{p \in U} (2\pi)^{-n/2} \sqrt{\int_{\mathcal{F}_p^{\mathbb{R}}} \prod_{i=1}^{n} \frac{e^{-\|f^i\|^2/2}}{\sqrt{2\pi}^{M_i - 1}} \det(DG\ DG^H)^{-1}} \quad .$$

As in the complex case, we add extra n variables:

$$AVG \leq (2\pi)^{-n/2} \int_{p \in U} \sqrt{\int_{\mathcal{F}^{\mathbb{R}}} \prod_{i=1}^{n} \frac{e^{-\|f^i\|^2/2}}{\sqrt{2\pi}^{M_i}} \det(DG\ DG^H)^{-1}} \quad ,$$

and we interpret $\det(DG\ DG^H)^{-1}$ in terms of a wedge. Since

$$\int_{x \in \mathbb{R}^M} |x_1|^2 \frac{e^{-\|x\|^2/2}}{\sqrt{2\pi}^M} = \int_{y \in \mathbb{R}} y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} = \int_{y \in \mathbb{R}} \frac{e^{-y^2/2}}{\sqrt{2\pi}} = 1 \quad ,$$

we obtain:

$$AVG \leq (2\pi)^{-n/2} \int_{p \in U} \sqrt{n! d\mathcal{T}^n} = (2\pi)^{-n/2} \int_{p \in U} \sqrt{n! d\mathcal{T}^n} \quad .$$

Now we would like to use Cauchy-Schwartz again. This time, the inner product is defined as:

$$\langle \varphi, \psi \rangle \stackrel{\text{def}}{=} \int_{p \in U} \varphi(p) \psi(p) dV \quad .$$

Hence,

$$AVG \leq (2\pi)^{-n/2} \langle n! d\mathcal{T}^n, \mathbf{1} \rangle \leq (2\pi)^{-n/2} \|n! d\mathcal{T}^n\| \|\mathbf{1}\| \quad .$$

This time, $\|\mathbf{1}\|^2 = \lambda(U)$, so we bound:

$$\begin{aligned} AVG &\leq (2\pi)^{-n/2}\sqrt{\lambda(U)}\sqrt{\int_U n! d\mathcal{T}^n} \\ &\leq (4\pi^2)^{-n/2}\sqrt{\lambda(U)}\sqrt{\int_{(p,q)\in\mathcal{T}^n, p\in U} n! d\mathcal{T}^n} \quad . \end{aligned}$$

■

4.2 Proof of Theorem 5

Proof of Theorem 5. Let $\varepsilon > 0$. As in the mixed case, we define:

$$\begin{aligned} \nu_{\mathbb{R}}(U,\varepsilon) &\stackrel{\text{def}}{=} \text{Prob}_{f\in\mathcal{F}}\left[\boldsymbol{\mu}(f;U) > \varepsilon^{-1}\right] \\ &= \text{Prob}_{f\in\mathcal{F}}\left[\exists p \in U : ev(f;p) = 0 \text{ and } \ d_{\mathbb{P}}(f,\Sigma_p) < \varepsilon\right] \end{aligned}$$

where now $U \in \mathbb{R}^n$.

Let $V(\varepsilon) \stackrel{\text{def}}{=} \{(f,p) \in F_{\mathbb{R}} \times U : ev(f;p) = 0 \text{ and } \ d_{\mathbb{P}}(f,\Sigma_p) < \varepsilon\}$. We also define $\pi : V(\varepsilon) \to \mathbb{P}(\mathcal{F})$ to be the canonical projection mapping $F_{\mathbb{R}} \times U$ to $F_{\mathbb{R}}$ and set $\#_{V(\varepsilon)}(f) \stackrel{\text{def}}{=} \#\{p \in U : (f,p) \in V(\varepsilon)\}$. Then,

$$\begin{aligned} \nu_{\mathbb{R}}(U,\varepsilon) &= \int_{f\in\mathcal{F}^{\mathbb{R}}} \frac{e^{-\sum_i \|f^i\|^2/2}}{\sqrt{2\pi}^{\sum M_i}} \chi_{\pi(V(\varepsilon))}(f) \ d\mathcal{F}^{\mathbb{R}} \\ &\leq \int_{f\in\mathcal{F}^{\mathbb{R}}} \frac{e^{-\sum_i \|f^i\|^2/2}}{\sqrt{2\pi}^{\sum M_i}} \#_{V(\varepsilon)} d\mathcal{F}^{\mathbb{R}} \\ &\leq \int_{p\in U\subset\mathbb{R}^n} \int_{\substack{f\in\mathcal{F}_p^{\mathbb{R}} \\ d_{\mathbb{P}}(f,\Sigma_p)<\varepsilon}} \frac{e^{-\sum_i \|f^i\|^2/2}}{\sqrt{2\pi}^{\sum M_i}} \frac{1}{NJ(f;p)} d\mathcal{F}_p^{\mathbb{R}} \ dV_{\mathcal{T}^n} \end{aligned}$$

As before, we change coordinates in each fiber of $\mathcal{F}_A^{\mathbb{R}}$ by

$$f = f_{\text{I}} + f_{\text{II}} + f_{\text{III}}$$

with f_{I}^i colinear to v_A^T, $(f_{\text{II}}^i)^T$ in the range of Dv_A, and f_{III}^i othogonal to f_{I}^i and f_{II}^i. This coordinate system is dependent on $p + q\sqrt{-1}$.

In the new coordinate system, formula 2.4.1 splits as follows:

$$\det\left(DG_{(p)}DG_{(p)}^H\right)^{-1/2} dV_{\mathcal{T}^n} =$$

$$= \left|\det \begin{bmatrix} (f_{\mathrm{I}}^1)_1 & \dots & (f_{\mathrm{I}}^1)_n \\ \vdots & & \vdots \\ (f_{\mathrm{I}}^n)_1 & \dots & (f_{\mathrm{I}}^n)_n \end{bmatrix}\right| \left|\det \begin{bmatrix} (Dv_A{}^{\mathrm{I}})_1^1 & \dots & (Dv_A{}^{\mathrm{I}})_n^1 \\ \vdots & & \vdots \\ (Dv_A{}^{\mathrm{I}})_1^n & \dots & (Dv_A{}^{\mathrm{I}})_n^n \end{bmatrix}\right| dV$$

$$= \left|\det \begin{bmatrix} (f_{\mathrm{I}}^1)_1 & \dots & (f_{\mathrm{I}}^1)_n \\ \vdots & & \vdots \\ (f_{\mathrm{I}}^n)_1 & \dots & (f_{\mathrm{I}}^n)_n \end{bmatrix}\right| \sqrt{\det Dv_A^H Dv_A}$$

The integral $E(U)$ of $\sqrt{\det Dv_A Dv_A^H}$ is the expected number of real roots on U, therefore

$$\nu_{\mathbb{R}}(U,\varepsilon) \leq E(U) \int_{\substack{f_{\mathrm{I}}+f_{\mathrm{II}} \in \mathcal{F}_p^{\mathbb{R}} \\ d_{\mathbb{P}}(f_{\mathrm{I}}+f_{\mathrm{II}}, \Sigma_p) < \varepsilon}} \frac{e^{-\sum_i \|f_{\mathrm{I}}^i + f_{\mathrm{II}}^i\|^2/2}}{\sqrt{2\pi}^{\sum M_i}} \cdot$$

$$\cdot \left|\det \begin{bmatrix} (f_{\mathrm{I}}^1)_1 & \dots & (f_{\mathrm{I}}^1)_n \\ \vdots & & \vdots \\ (f_{\mathrm{I}}^n)_1 & \dots & (f_{\mathrm{I}}^n)_n \end{bmatrix}\right| d\mathcal{F}_p^{\mathbb{R}} \quad .$$

In the new system of coordinates, Σ_p is defined by the equation:

$$\det \begin{bmatrix} (f_{\mathrm{I}}^1)_1 & \dots & (f_{\mathrm{I}}^1)_n \\ \vdots & & \vdots \\ (f_{\mathrm{I}}^n)_1 & \dots & (f_{\mathrm{I}}^n)_n \end{bmatrix} = 0 \quad .$$

Since $\|f_{\mathrm{I}} + f_{\mathrm{II}}\| \geq \|f_{\mathrm{I}}\|$,

$$d_{\mathbb{P}}(f_{\mathrm{I}} + f_{\mathrm{II}}, \Sigma_p) < \varepsilon \Longrightarrow d_{\mathbb{P}}(f_{\mathrm{I}}, \Sigma_p) < \varepsilon \quad .$$

This implies:

$$\nu_{\mathbb{R}}(U,\varepsilon) \leq E(U) \int_{\substack{f_{\mathrm{I}}+f_{\mathrm{II}} \in F_p^{\mathbb{R}} \\ d_{\mathbb{P}}(f_{\mathrm{I}}, [\det=0]) < \varepsilon}} \frac{e^{-\sum_i \|f_{\mathrm{I}}^i + f_{\mathrm{II}}^i\|^2/2}}{\sqrt{2\pi}^{\sum M_i}} \cdot$$

$$\cdot \left|\det \begin{bmatrix} (f_{\mathrm{I}}^1)_1 & \dots & (f_{\mathrm{I}}^1)_n \\ \vdots & & \vdots \\ (f_{\mathrm{I}}^n)_1 & \dots & (f_{\mathrm{I}}^n)_n \end{bmatrix}\right| d\mathcal{F}_p^{\mathbb{R}} \quad .$$

We can integrate the $(\sum M_i - n - 1)$ variables f_{III} to obtain:

$$\nu_{\mathbb{R}}(U,\varepsilon) = E(U) \int_{\substack{f_{\mathrm{II}} \in \mathbb{R}^{n^2} \\ d_{\mathbb{P}}(f_{\mathrm{II}}, [\det = 0]) < \varepsilon}} \frac{e^{-\sum_i \|f_{\mathrm{II}}^i\|^2/2}}{\sqrt{2\pi}^{n^2}} |\det f_{\mathrm{II}}|^2 \; d\mathbb{R}^{n^2} \quad .$$

This is $E(U)$ times the probability $\nu(n,\varepsilon)$ for the linear case. ■

5 Mixed Thoughts about Mixed Manifolds

Let $X : E \rightarrow F$ be a linear operator. Here, we assume that E has a canonical Riemannian structure, and that F has n possibly different Riemannian structures $\langle \cdot, \cdot \rangle_{A_i}$.

We would like to interpret the quantities

$$\max_{\|v\| \leq 1} \min_j \|Xv\|_{A_i}$$

and

$$\max_{\|v\| \leq 1} \max_j \|Xv\|_{A_i}$$

in terms of an operator norm on X. Let B denote the unit ball in E, and let B_i denote the unit ball in $(F, \langle \cdot, \cdot \rangle_{A_i})$. Note that

$$\max_{\|v\| \leq 1} \max_j \|Xv\|_{A_i} \leq 1 \iff X(B) \subseteq \bigcap B_i \quad ,$$

while

$$\max_{\|v\| \leq 1} \min_j \|Xv\|_{A_i} \leq 1 \iff X(B) \subseteq \bigcup B_i \quad .$$

This should be compared to:

$$\min_j \max_{\|v\| \leq 1} \|Xv\|_{A_i} \leq 1 \iff \exists i : X(B) \subseteq B_i \quad .$$

One standard way to define norms is by choosing an arbitrary symmetric convex set, and equating that set to the unit ball. (Such norms are called **Minkowski norms**.)

There are two immediate obvious choices:

1. We can use $\bigcap B_i$ as the unit ball.

2. We can use $\mathrm{Conv} \bigcup B_i$ as the unit ball.

In the first case, we can endow $\mathcal{T}^n$ with a $\mathcal{C}^0$ Finsler structure, while in the second case we can obtain a $\mathcal{C}^1$ Finsler structure.

Using $\mathrm{Conv} \bigcup B_i$ would have the advantage of a known probabilistic bound for $\boldsymbol{\mu} > \varepsilon^{-1}$. However, $\bigcap B_i$ seems to be more convenient for the study of polyhedral homotopy [**?**].

Finsler structures are legitimate ways to endow a non-Riemannian manifold with a few familiar concepts. For instance, once we define a Finsler structure $\|\!|\cdot|\!\|_x$, the length of a curve $x(t)$, $t \in [0, 1]$ is defined to be $\int_0^1 \|\!|\dot{x}(t)|\!\|_{x(t)} \, dt$.

A general discussion on Finsler geometry can be found in [**?**, Ch. 8].

Remark 8. The proof of Theorem 2 strongly suggests that the geometry of mixed manifolds should be determined by a much more fundamental invariant, a norm in the space $L(\mathbb{C}^n, T_{(p,q)}\mathcal{T}^n)$, which we can take to be either side of the following equality:

$$\max_v \left(\sqrt{\sum_{i=1}^n \frac{|v^i|^2}{\|Xv\|_{A_i}^2}} \right)^{-1} = \max_u \left(\sqrt{\sum_{i=1}^n \frac{|(X^{-1}u)^i|^2}{\|u\|_{A_i}^2}} \right)^{-1} . \blacksquare$$

Remark 9. There is a class of polynomial systems that are not unmixed, but nevertheless can be treated as if they were unmixed. For instance, in the dense case, the potentials g_{A_i}, $i = 1, \cdot, n$ are all multiples of one another, therefore $\kappa \equiv 1$. The toric variety associated to those systems admits therefore a (possibly singular) Hermitian structure. That structure is non-singular provided that the A_i's satisfy Delzant's condition [**?**] (see also Appendix A below). Roughly speaking, Delzant's condition is an assertion about the angle cones of the Minkowski sum of the $\mathrm{Conv}(A_i)$. ■

A Mechanical Interpretation of the Momentum Map

The objective of this Section is to clarify the analogy between the geometry of polynomial roots and Hamiltonian mechanics. The key for that analogy was the existance of a *momentum map* associated to convex sets.

In the case the convex set is the support of a polynomial, that *momentum map* is also the momentum map 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 (see Formula 2.1.1). 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:

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

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

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

by Formula 2.1.4.

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

$$\begin{cases} \dot{p} &= \frac{\partial H_\xi}{\partial q} \\ \dot{q} &= -\frac{\partial H_\xi}{\partial p} \end{cases} \quad ,$$

where $H_\xi = \nabla g_A(p) \cdot \xi$.

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*.

The theorem of Atiyah, Guillemin and Sternberg asserts that under certain conditions, the image of the momentum map is the convex hull of the points A^α. One of those conditions requires that the action should take place in a compact symplectic manifold (as in [**?**, Th. 1] or [**?**, Th. 4]) or (sometimes) a compact Kähler manifold [**?**, Th 2].

We may consider a compactification of $\mathcal{T}^n$, such as the closure of $\exp(\mathcal{T}^n)$. Unfortunately, there are situations where this compactification is not a symplectic manifold, because the form ω_A vanishes in the preimage of one or more A^α's (thus differing from the assumptions of [**?**, p. 127]).

In [**?**], a necessary and sufficient condition for the compactification of $\mathcal{T}^n$ to be a symplectic manifold is given. Namely, the polytope $\mathrm{Conv}(A)$ should be simple (i.e., every vertex should be incident to exactly n edges) and unimodular (the integer points along the rays generated by each such n-tuple of edges should span $\mathbb{Z}^n$ as a $\mathbb{Z}$-module). If all the polytopes $\mathrm{Conv}(A_1), \dots, \mathrm{Conv}(A_n)$ satisfy Delzant's condition then we can construct a corresponding compactifaction $\mathcal{T}^n$ and apply Atiyah-Guillemin-Sternberg's theorem.

Another possibility is to blow-up the singularities, as explained in [**?**]. If we do so, polytopes $A_1, \cdots, A_n$ will be "shaved:" locally, the cone emanating from a vertex will be truncated by the intersection with a half-space with boundary infinitesimally close to a supporting hyperplane. From another point of view, the underlying normal fan [**?**] will be refined. However, the relation between the original polynomial system and the new momentum map is not yet clear.

B The Coarea Formula

This is an attempt to give a short proof of the coarea formula, in a version suitable to the setting of this paper. This means we take all manifolds and functions smooth and avoid measure theory as much as possible.

Proposition 5.

1. *Let X be a smooth Riemann manifold, of dimension M and volume form $|dX|$.*

2. *Let Y be a smooth Riemann manifold, of dimension n and volume form $|dY|$.*

3. *Let U be an open set of X, and $F : U \to Y$ be a smooth map, such that DF_x is surjective for all x in U.*

4. Let $\varphi : X \to \mathbb{R}^+$ be a smooth function with compact support contained in U.

Then for almost all $z \in F(U)$, $V_z \stackrel{\text{def}}{=} F^{-1}(z)$ is a smooth Riemann manifold, and

$$\int_X \varphi(x) NJ(F;x)|dX| = \int_{z\in Y} \int_{x \in V_z} \varphi(x)|dV_z||dY|$$

where $|dV_z|$ is the volume element of V_z and $NJ(F,x) = \sqrt{\det DF_x^H DF_x}$ is the product of the singular values of DF_x. ■

By the implicit function theorem, whenever V_z is non-empty, it is a smooth $(N-n)$-dimensional Riemann submanifold of X. By the same reason, $V \stackrel{\text{def}}{=} \{(z,x) : x \in V_z\}$ is also a smooth manifold.

Let η be the following N-form restricted to V:

$$\eta = dY \wedge dV_z \quad .$$

This is **not** the volume form of V. The proof of Proposition 5 is divided into two steps:

Lemma 10.

$$\int_V \varphi(x)|\eta| = \int_X \varphi(x) NJ(F;x)|dX|$$

.

Lemma 11.

$$\int_V \varphi(x)|\eta| = \int_{z\in Y} \int_{x\in V_z} \varphi(x)|dV_z||dY| \quad .$$

Proof of Lemma 10. We parametrize:

$$\begin{array}{rccl} \psi : & X & \to & V \\ & x & \mapsto & (F(x), x) \end{array} \quad .$$

Then,

$$\int_V \varphi(x)|\eta| = \int_X (\varphi \circ \psi)(x)|\psi^* \eta| \quad .$$

We can choose an orthonormal basis $u_1, \cdots, u_M$ of T_xX such that $u_{n+1}, \cdots, u_M \in \ker DF_x$. Then,

$$D\psi(u_i) = \begin{cases} (DF_x u_i, u_i) & i = 1, \cdots, n \\ (0, u_i) & i = n+1, \cdots, M \end{cases} .$$

Thus,

$$\begin{aligned} |\psi^* \eta(u_1, \cdots, u_M)| &= |\eta(D\psi u_1, \cdots, D\psi u_M)| \\ &= |dY(DF_x u_1, \cdots, DF_x u_n)|\ |dV_z(u_{n+1}, \cdots, u_M)| \\ &= |\det DF_x|_{\ker DF_x^\perp}| \\ &= NJ(F, x) \end{aligned}$$

and hence

$$\int_V \varphi(x)|\eta| = \int_X \varphi(x) NJ(F; x)|dX| \quad .$$

■

Proof of Lemma 11. We will prove this Lemma locally, and this implies the full Lemma through a standard argument (partitions of unity in a compact neighborhood of the support of φ).

Let x_0, z_0 be fixed. A small enough neighborhood of $(x_0, z_0) \subset V_{z_0}$ admits a fibration over V_{z_0} by planes orthogonal to $\ker DF_{x_0}$.

We parametrize:

$$\begin{array}{llll} \theta : & Y \times V_{z_0} & \to & V \\ & (z, x) & \mapsto & (z, \rho(x, z)) \end{array} ,$$

where $\rho(x, z)$ is the solution of $F(\rho) = z$ in the fiber passing through (z_0, x). Remark that $\theta^* dY = dY$, and $\theta^* dV_z = \rho^* DV_z$. Therefore,

$$\theta^*(dY \wedge dV_z) = dY \wedge (\rho^* dV_z) \quad .$$

Also, if one fixes z, then ρ is a parametrization $V_{z_0} \to V_z$. We have:

$$\begin{aligned} \int_V \varphi(x)|\eta| &= \int_{Y \times V_{z_0}} \varphi(\rho(x, z))|\theta^* \eta| \\ &= \int_{z \in Y} \left(\int_{x \in V_{z_0}} \varphi(\rho(x, z)|\rho^* dV_z| \right) |dY| \\ &= \int_{z \in Y} \left(\int_{x \in V_z} \varphi(x)|dV_z| \right) |dY| \end{aligned}$$

■

The proposition below is essentially Theorem 3 p. 240 of [?]. However, we do not require our manifolds to be compact. We assume all maps and manifolds are smooth, so that we can apply proposition 5.

Proposition 6.

1. *Let X be a smooth M-dimensional manifold with volume element $|dX|$.*
2. *Let Y be a smooth n-dimensional manifold with volume element $|dY|$.*
3. *Let V be a smooth M-dimensional submanifold of $X \times Y$, and let $\pi_1 : V \to X$ and $\pi_2 : V \to Y$ be the canonical projections from $X \times Y$ to its factors.*
4. *Let Σ' be the set of critical points of π_1, we assume that Σ' has measure zero and that Σ' is a manifold.*
5. *We assume that π_2 is regular (all points in $\pi_2(V)$ are regular values).*
6. *For any open set $U \subset V$, for any $x \in X$, we write: $\#_U(x) \stackrel{\text{def}}{=} \#\{\pi_1^{-1}(x) \cap U\}$. We assume that $\int_{x \in X} \#_V(x)|dX|$ is finite.*

Then, for any open set $U \subset V$,

$$\int_{x \in \pi_1(U)} \#_U(x)|dX| = \int_{z \in Y} \int_{\substack{x \in V_z \\ (x,z) \in U}} \frac{1}{\sqrt{\det DG_x DG_x^H}} |dV_z||dY|$$

where G is the implicit function for $(\hat{x}, G(\hat{x})) \in V$ in a neighborhood of $(x, z) \in V \setminus \Sigma'$. ■

Proof. Every $(x, z) \in U \setminus \Sigma'$ admits an open neighborhood such that π_1 restricted to that neighborhood is a diffeomorphism. This defines an open covering of $U \setminus \Sigma'$. Since $U \setminus \Sigma'$ is locally compact, we can take a countable subcovering and define a partition of unity $(\varphi_\lambda)_{\lambda \in \Lambda}$ subordinated to that subcovering.

Also, if we fix a value of z, then $(\varphi_\lambda)_{\lambda\in\Lambda}$ becomes a partition of unity for $\pi_1(\pi_1^{-1}(V_z) \cap U)$. Therefore,

$$\begin{aligned}
\int_{x\in\pi_1(U)} \#_U(x)|dX| &= \sum_{\lambda\in\Lambda} \int_{x,z\in\mathrm{Supp}\varphi_\lambda} \varphi_\lambda(x,z)|dX| \\
&= \sum_{\lambda\in\Lambda} \int_{z\in Y} \int_{x,z\in\mathrm{Supp}\varphi_\lambda} \frac{\varphi_\lambda(x,z)}{NJ(G,x)}|dX| \\
&= \int_{z\in Y} \sum_{\lambda\in\Lambda} \int_{x,z\in\mathrm{Supp}\varphi_\lambda} \frac{\varphi_\lambda(x,z)}{NJ(G,x)}|dX| \\
&= \int_{z\in Y} \int_{x\in V_z} \frac{1}{NJ(G,x)}|dX|
\end{aligned}$$

where the second equality uses Proposition 5 with $\varphi = \varphi_\lambda/NJ$. Since $NJ = \sqrt{\det DG_x DG_x^H}$, we are done. ■