

# The Maximum Likelihood Degree

Fabrizio Catanese, Serkan Hoşten,  
Amit Khetan, and Bernd Sturmfels

## Abstract

Maximum likelihood estimation in statistics leads to the problem of maximizing a product of powers of polynomials. We study the algebraic degree of the critical equations of this optimization problem. This degree is related to the number of bounded regions in the corresponding arrangement of hypersurfaces, and to the Euler characteristic of the complexified complement. Under suitable hypotheses, the maximum likelihood degree equals the top Chern class of a sheaf of logarithmic differential forms. Exact formulae in terms of degrees and Newton polytopes are given for polynomials with generic coefficients.

## 1 Introduction

In algebraic statistics [13, 21, 22], a *model for discrete data* is a map  $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^n$  whose coordinates  $f_1, \dots, f_n$  are polynomial functions in the parameters  $(\theta_1, \dots, \theta_d) =: \theta$ . The parameter vector  $\theta$  ranges over an open subset  $\mathcal{U}$  of  $\mathbb{R}^d$  such that  $f(\theta)$  lies in the positive orthant  $\mathbb{R}_{>0}^n$ . The image  $f(\mathcal{U})$  represents a family of probability distributions on an  $n$ -element state space, provided we make the extra assumption that  $f_1 + \dots + f_n - 1$  is the zero polynomial.

A given data set is a vector  $u = (u_1, \dots, u_n)$  of positive integers. The problem of *maximum likelihood estimation* is to find parameters  $\theta$  which best explain the data  $u$ . This leads to the following optimization problem:

$$\text{Maximize } f_1(\theta)^{u_1} f_2(\theta)^{u_2} \cdots f_n(\theta)^{u_n} \quad \text{subject to } \theta \in \mathcal{U}. \quad (1)$$

Under suitable assumptions we have an optimal solution  $\hat{\theta}$  to the problem (1), which is an algebraic function of the data  $u$ . Our goal is to compute the degree of that algebraic function. We call this number the *maximum*

*likelihood degree* of the model  $\mathbf{f}$ . Equivalently, the *ML degree* is the number of complex solutions to the critical equations of (1), for a general data vector  $u$ . In this paper we prove results of the following form:

**Theorem 1.** *Let  $f_1, \dots, f_n$  be polynomials of degrees  $b_1, \dots, b_n$  in  $d$  unknowns. If the maximum likelihood degree of the model  $\mathbf{f} = (f_1, \dots, f_n)$  is finite then it is less than or equal to the coefficient of  $z^d$  in the generating function*

$$\frac{(1-z)^d}{(1-zb_1)(1-zb_2)\cdots(1-zb_n)}. \quad (2)$$

*Equality holds if the coefficients of the polynomials  $f_i$  are sufficiently generic.*

As an example, consider a model given by  $n = 4$  quadratic polynomials in  $d = 2$  parameters. The solution to (1) satisfies the two critical equations

$$\frac{u_1}{f_1} \frac{\partial f_1}{\partial \theta_1} + \frac{u_2}{f_2} \frac{\partial f_2}{\partial \theta_1} + \frac{u_3}{f_3} \frac{\partial f_3}{\partial \theta_1} + \frac{u_4}{f_4} \frac{\partial f_4}{\partial \theta_1} = \frac{u_1}{f_1} \frac{\partial f_1}{\partial \theta_2} + \frac{u_2}{f_2} \frac{\partial f_2}{\partial \theta_2} + \frac{u_3}{f_3} \frac{\partial f_3}{\partial \theta_2} + \frac{u_4}{f_4} \frac{\partial f_4}{\partial \theta_2} = 0.$$

If the  $f_i$ 's are general quadrics then these equations have 25 complex solutions. The formula for the maximum likelihood degree in Theorem 1 gives

$$\frac{(1-z)^2}{(1-2z)^4} = 1 + 6z + \underline{25}z^2 + 88z^3 + 280z^4 + \dots.$$

For special quadrics  $f_i$ , the ML degree can be much lower than 25. A familiar example is the *independence model for two binary random variables*:

$$f_1 = \theta_1\theta_2, \quad f_2 = (1-\theta_1)\theta_2, \quad f_3 = \theta_1(1-\theta_2), \quad f_4 = (1-\theta_1)(1-\theta_2). \quad (3)$$

Here the ML degree is only one because the maximum likelihood estimate  $\hat{\theta}$  is a rational function (= algebraic function of degree one) of the data  $u$ :

$$\hat{\theta}_1 = \frac{u_1 + u_3}{u_1 + u_2 + u_3 + u_4} \quad \text{and} \quad \hat{\theta}_2 = \frac{u_1 + u_2}{u_1 + u_2 + u_3 + u_4}.$$

This paper is organized as follows. In Section 2 we present the algebraic geometry for studying critical points of a rational function  $f = f_1^{u_1} \cdots f_n^{u_n}$  on an irreducible projective variety  $X$ . The critical equations  $d \log(f) = 0$  are interpreted as sections of the sheaf  $\Omega^1(\log D)$  of 1-forms with logarithmic singularities along the divisor  $D$  defined by  $f$ . In Theorem 4, we show that if  $D$  is a global normal crossing divisor then the ML degree equals the degree

of the top Chern class of  $\Omega^1(\log D)$ . If  $X$  is projective  $d$ -space then this leads to Theorem 1. In Section 3 we study the case when  $X$  is a smooth toric variety, and we derive a formula for the ML degree when the  $f_i$ 's are Laurent polynomials which are generic relative to their Newton polytopes. For instance, Example 8 shows that the ML degree is 13 if we replace (3) by

$$f_i = \alpha_i + \beta_i\theta_1 + \gamma_i\theta_2 + \delta_i\theta_1\theta_2 \quad (i = 1, 2, 3, 4).$$

Section 4 is concerned with the relationship of the ML degree to the bounded regions of the complement of  $\{f_i = 0\}$  in  $\mathbb{R}^d$ . The number of these regions is a lower bound to the number of real solutions of the critical equations, and therefore a lower bound to the ML degree. We show that for plane quadrics all three numbers can be equal. However, for other combinations of plane curves the ML degree and the number of bounded regions diverge, and we prove a tight upper bound on the latter in Theorem 12. Also, following work of Terao [24] and Varchenko [25], we show in Theorem 13 that the ML degree coincides with the number of bounded regions of the arrangement of hyperplanes  $\{f_i = 0\}$  when the  $f_i$ 's are (not necessarily generic) linear forms.

Section 5 revisits the ML degree for toric varieties, replacing the smoothness assumption by a much milder condition. Theorem 15 gives a purely combinatorial formula for the ML degree in terms of the Newton polytopes of the polynomials  $f_i$ . This section also discusses how resolution of singularities can be used to compute the ML degree for nongeneric polynomials.

Section 6 deals with topological methods for determining the ML degree. Theorem 19 shows that, under certain restrictive hypotheses, it coincides with the Euler characteristic of the complex manifold  $X \setminus D$ , and Theorem 22 offers a general version of the semi-continuity principle which underlies the inequality in Theorem 1. In Section 7 we relate the ML degree to the sheaf of logarithmic vector fields along  $D$ , which is the sheaf dual to  $\Omega^1(\log D)$ .

This paper was motivated by recent appearances of the concept of ML degree in statistics and computational biology. Chor, Khetan and Snir [7] showed that the ML degree of a phylogenetic model equals 9, and Geiger, Meek and Sturmfels [14] proved that an undirected graphical model has ML degree one if and only if it is decomposable. The notion of ML degree also makes sense for certain parametrized models for continuous data: Drton and Richardson [10] showed that the ML degree of a Gaussian graphical model equals 5, and Bout and Richards [5] studied the ML degree of certain mixture models. The ML degree always provides an upper bound on the number of

local maxima of the likelihood function. Our ultimate hope is that a better understanding of the ML degree will lead to the development of custom-tailored algorithms for solving the critical equations  $d\log(f) = 0$ . There is a need for such new algorithms, given that methods currently used in statistics (notably the *EM-algorithm*) often produce only local maxima in (1).

## 2 Critical Points of Rational Functions

In this section we work in the following general set-up of algebraic geometry. Let  $X$  be a complete factorial algebraic variety over the complex numbers  $\mathbb{C}$ . We also assume that  $X$  is irreducible of dimension  $d \geq 1$ . In applications to statistics, the variety  $X$  will often be a smooth projective toric variety.

Suppose that  $f \in \mathbb{C}(X)$  is a rational function on  $X$ . Since  $X$  is factorial, the local rings  $\mathcal{O}_{X,x}$  are unique factorization domains. This means that the function  $f$  has a global factorization which is unique up to constants:

$$f = F_1^{u_1} F_2^{u_2} \cdots F_r^{u_r}. \quad (4)$$

Here  $F_i$  is a prime section of an invertible sheaf  $\mathcal{O}_X(D_i)$  where  $D_i$  is the divisor on  $X$  defined by  $F_i$ . In our applications we usually assume that  $r \geq n$  where  $n$  is the number considered in the Introduction. For instance, if  $f_1, \dots, f_n$  are polynomials and  $X = \mathbb{P}^d$  then  $r = n + 1$ ; namely,  $F_1, \dots, F_n$  are the homogenizations of  $f_1, \dots, f_n$  using  $\theta_0$ , and  $F_{n+1} = \theta_0$  (see the proof of Theorem 1 for details).

By (4), we can write the divisor of the rational function  $f$  uniquely as

$$\operatorname{div}(f) = \sum_{i=1}^r u_i D_i,$$

where the  $u_i$ 's are (possibly negative) integers. Let  $D$  be the reduced union of the codimension one subvarieties  $D_i \subset X$ , or, as a divisor,  $D := \sum_{i=1}^r D_i$ .

We are interested in computing the critical points of the rational function  $f$  on the open set  $V := X \setminus D$  complementary to the divisor  $D$ . Especially, we wish to know the number of critical points, counted with multiplicities.

A critical point is by definition a point  $x \in X$  where the differential 1-form  $df$  vanishes. If  $x$  is a smooth point on  $X$ , and  $x_1, \dots, x_d$  are local coordinates, then  $df = \sum_{j=1}^d (\partial f / \partial x_j) dx_j$ . Hence  $x$  is a critical point of  $f$  if

and only if

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_d} = 0. \quad (5)$$

We next rewrite the critical equations (5) using the factorization (4). Around each point  $x \in X$ , we may choose a local trivialization for the sheaf  $\mathcal{O}_X(D_i)$  and express  $F_i$  locally by a regular function. By slight abuse of notation, we denote that regular function also by  $F_i$ . For instance, if  $X = \mathbb{P}^d$  then this means replacing the homogeneous polynomial  $F_i$  by a dehomogenization.

Since  $f$  has neither zeros nor poles on the open set  $V$ , the vanishing of  $df$  is equivalent to the vanishing of the logarithmic derivative

$$d\log(f) = \frac{df}{f} = \sum_{i=1}^r u_i d\log(F_i) = \sum_{i=1}^r u_i \frac{dF_i}{F_i}. \quad (6)$$

We now recall some classical definitions and results concerning the sheaf of differential 1-forms with logarithmic singularities along  $D$ . The standard references on this subject are Deligne's book [9] and Saito's paper [23]. We define  $\Omega_X^1(\log D)$  as a subsheaf of the sheaf  $\Omega_X^1(D)$  of 1-forms with poles at most on  $D$  and of order one. This sheaf is the image of the natural map

$$\Omega_X^1 \oplus \mathcal{O}_X^r \longrightarrow \Omega_X^1(D)$$

which is given by the inclusion  $\Omega_X^1 \rightarrow \Omega_X^1(D)$  and the homomorphisms sending  $1 \in \mathcal{O}_X \rightarrow d\log(F_i)$ . For experts we note that our definition differs from the one in [23] when  $D$  is not normal crossing. Saito's sheaf is the double dual of our  $\Omega_X^1(\log D)$ , which explains why his is always locally free when  $X$  is a surface [23, Corollary 1.7]. Ours need not be locally free even for surfaces. However, our definition gives a natural exact sequence.

**Lemma 2.** *If  $X$  is factorial and complete then we have an exact sequence*

$$0 \rightarrow \Omega_X^1 \rightarrow \Omega_X^1(\log D) \rightarrow \bigoplus_{i=1}^r \mathcal{O}_{D_i} \rightarrow 0. \quad (7)$$

*Proof.* The local sections of the sheaf  $\Omega_X^1(\log D)$  are rational 1-forms which can be written as  $\omega = \sum_{i=1}^r \psi_i \cdot d\log(F_i) + \eta$ , where  $\eta$  is a regular 1-form.

Since the  $D_i$ 's are distinct prime divisors and  $X$  is factorial, the local rings  $\mathcal{O}_{X, D_i}$  are discrete valuation rings with parameter  $F_i$ . Thus  $F_j$  is

invertible in this local ring for  $j \neq i$ , and  $\omega$  is regular if and only if  $F_i$  divides  $\psi_i$ . This implies that the homomorphism which sends  $\omega$  to the vector  $(\psi_i \pmod{F_i})_{i=1, \dots, r}$  is well defined, and it induces an isomorphism from the quotient  $\Omega_X^1(\log D)/\Omega_X^1$  onto  $\bigoplus_{i=1}^r \mathcal{O}_{D_i}$ .  $\square$

Assume now that  $X$  is smooth. Then both sheaves  $\Omega_X^1(D)$  and  $\Omega_X^1$  are locally free of rank  $d = \dim(X)$ . Hence the intermediate sheaf  $\Omega_X^1(\log D)$  is torsion free of the same rank. Our next result shows that  $\Omega_X^1(\log D)$  is locally free if and only if the divisors  $D_i$  are smooth and intersect transversally.

**Proposition 3.** *Let  $x \in X$  be a smooth point,  $x_1, \dots, x_d$  local coordinates at  $x$  and  $D_1, \dots, D_h$  the divisors which contain  $x$ . Then the sheaf  $\Omega_X^1(\log D)$  is locally free at  $x$  if and only if the  $h \times d$ -matrix  $(\partial F_i / \partial x_j)$  has rank  $h$  at  $x$ .*

*Proof.* Any local section of  $\Omega_X^1(\log D)$  can be written in the form

$$\omega = \sum_{i=1}^r \psi_i \cdot d\log(F_i) + \eta = \sum_{i=1}^h \psi_i \cdot d\log(F_i) + \sum_{j=1}^d \eta_j \cdot dx_j. \quad (8)$$

This observation gives rise to a local exact sequence

$$0 \rightarrow \mathcal{O}_{X,x}^h \rightarrow \mathcal{O}_{X,x}^h \oplus \mathcal{O}_{X,x}^d \rightarrow \Omega_{X,x}^1(\log D) \rightarrow 0. \quad (9)$$

The surjective map on the right takes  $((\psi_i), (\eta_j))$  to the sum on the right hand side of (8). The injective map on the left takes the  $h$ -tuple  $(A_1, \dots, A_h)$  to

$$((\psi_i), (\eta_j)) \quad \text{with} \quad \psi_i = F_i A_i \quad \text{and} \quad \eta_j = - \sum_{l=1}^h A_l \frac{\partial F_l}{\partial x_j}.$$

The exactness of the sequence (9) follows from the proof of Lemma 2. If the section  $\omega$  in (8) is identically zero in  $\Omega_{X,x}^1(\log D)$  then  $\omega$  is in particular regular, and so  $F_i$  divides each  $\psi_i$ .

Now, since  $X$  is reduced, a coherent sheaf  $\mathcal{F}$  is locally free of rank  $d$  if and only if  $\dim_{\mathbb{C}} \mathcal{F} \otimes \mathbb{C}_x = d$  for each point  $x$ . Since tensor product is right exact, it follows that this condition is verified for  $\Omega_X^1(\log D)$  if and only if the matrix of  $\mathcal{O}_X^h \rightarrow \mathcal{O}_X^h \oplus \mathcal{O}_X^d$ , evaluated at  $x$ , has rank precisely  $h$ . Since the functions  $F_1, \dots, F_h$  vanish at  $x$ , this is exactly the asserted condition that the Jacobian matrix  $(\partial F_i / \partial x_j)_{i=1, \dots, h, j=1, \dots, d}$  has rank  $h$  at  $x$ .  $\square$

In the above situation where  $X$  is smooth and  $\Omega_X^1(\log D)$  is locally free we shall say that the divisor  $D$  has *global normal crossings* (or GNC).

**Theorem 4.** *Let  $X$  be smooth and assume that  $D$  is a GNC divisor. Then*

1. *the section  $d\log(f)$  of  $\Omega_X^1(\log D)$  does not vanish at any point of  $D$ ,*
2. *if the divisor  $D$  intersects every curve in  $X$  (in particular, if  $D$  is ample) then  $d\log(f)$  vanishes only on a finite subset of  $V = X \setminus D$ ,*
3. *if the above conclusions hold, then the number of critical points of  $f$  on  $V$ , counted with multiplicities, equals the degree of the top Chern class  $c_d(\Omega_X^1(\log D))$ .*

*Proof.* We abbreviate  $\sigma := d\log(f) = \sum_{i=1}^r u_i d\log(F_i)$ . By the proof of Proposition 3 it follows that if  $(\partial F_i / \partial x_j)_{i=1, \dots, h, j=1, \dots, d}$  has rank  $h$  at  $x$ , then  $\Omega_X^1(\log D)$  is locally free of rank  $d$  with generators  $d\log(F_i)$  and some choice of  $d - h$  of the  $dx_j$ . If we write  $\sigma$  in this basis, the coefficients of  $d\log(F_i)$  are the constants  $u_i$  while the coefficients of the  $dx_j$  are some regular functions. The first assertion follows immediately since the exponents  $u_i$  are all nonzero.

The second assertion follows from the first: let  $Z_\sigma$  be the zero set of the section  $\sigma$ . Since  $Z_\sigma$  does not intersect  $D$ , it follows that  $\dim(Z_\sigma) = 0$ .

Thirdly, if  $\mathcal{F}$  is a locally free sheaf of rank  $d$  on a smooth variety  $X$  of dimension  $d$ , and  $\sigma$  is a section of  $H^0(\mathcal{F})$  with a zero scheme  $Z_\sigma$  of dimension 0, then the length of  $Z_\sigma$  equals the degree of the top Chern class  $c_d(\mathcal{F})$ .  $\square$

The *total Chern class* of a sheaf  $\mathcal{F}$  is the sum  $c_{tot}(\mathcal{F}) = \sum_{i=0}^d c_i(\mathcal{F})z^i$ . This is a polynomial in  $z$  whose coefficients are elements in the Chow ring  $A^*(X)$ . Recall that every element in  $A^*(X)$  has a well-defined *degree* which is the image of its degree  $d$  part under the degree homomorphism  $A^d(X) \rightarrow \mathbb{Z}$ .

**Corollary 5.** *Suppose that  $X$  is smooth and  $D$  is a GNC divisor on  $X$  which intersects every curve. Then the number of critical points of  $f$ , counted with multiplicities, is the degree of the coefficient of  $z^d$  in the following polynomial:*

$$c_{tot}(\Omega_X^1) \cdot \prod_{i=1}^r (1 - zD_i)^{-1} \in A^*(X)[z]. \quad (10)$$

*Proof.* The total Chern class  $c_{tot}(\mathcal{F})$  is multiplicative with respect to exact sequences, i.e., if  $0 \rightarrow A \rightarrow B \rightarrow C \rightarrow 0$  is an exact sequence of sheaves, then  $c_{tot}(B) = c_{tot}(A) \cdot c_{tot}(C)$ . Hence the sequence (7) implies the result.  $\square$

In the next section, we apply the formula (10) in the case when  $X$  is a smooth projective toric variety. The Chow group  $A^d(X)$  has rank one and is generated by the class of any point. This canonically identifies  $A^d(X)$  with  $\mathbb{Z}$  and so any top Chern class can be considered to be a number.

**Corollary 6.** *Suppose  $X$  is a smooth toric variety with boundary divisors  $\Delta_1, \dots, \Delta_s$  and  $D$  is GNC and meets every curve. The number of critical points of  $f$ , counted with multiplicity, equals the coefficient of  $z^d$  in*

$$\frac{\prod_{j=1}^s (1 - z\Delta_j)}{\prod_{i=1}^r (1 - zD_i)} \in A^*(X)[z]. \quad (11)$$

*Proof.* By virtue of equation (10) we need only compute the total Chern class  $c_{tot}(\Omega_X^1)$ . For this we use the exact sequence in [12, page 87],

$$0 \rightarrow \Omega_X^1 \rightarrow \Omega_X^1(\log\Delta) \rightarrow \bigoplus_{j=1}^s \mathcal{O}_{\Delta_j} \rightarrow 0,$$

where  $\Delta = \sum_{j=1}^s \Delta_j$ , and the fact that  $\Omega_X^1(\log\Delta)$  is trivial.  $\square$

### 3 Models defined by Generic Polynomials

We now apply the results of the previous section to models  $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^n$ . To illustrate how this works, we first prove Theorem 1 for generic polynomials. The proof of the statement that the ML degree of generic polynomials is an upper bound on the ML degree of special polynomials (when this number is finite) is deferred to Theorem 7 which is a generalization of Theorem 1. See also Theorem 22 where this semi-continuity principle is stated in general.

*Proof of Theorem 1 (generic case).* The polynomials  $f_1, \dots, f_n$  are assumed to be generic among all (nonhomogeneous) polynomials of degrees  $b_1, \dots, b_n$  in  $\theta_1, \dots, \theta_d$ , and  $u_1, \dots, u_n$  are positive integers. We take  $X$  to be projective space  $\mathbb{P}^d$  with coordinates  $(\theta_0 : \theta_1 : \dots : \theta_d)$ . Our object of interest is the following rational function on  $X = \mathbb{P}^d$ :

$$F = (f_1^{u_1} f_2^{u_2} \dots f_n^{u_n}) \left( \frac{\theta_1}{\theta_0}, \frac{\theta_2}{\theta_0}, \dots, \frac{\theta_d}{\theta_0} \right).$$

The global factorization (4) of this  $F$  has  $r = n + 1$  prime factors, namely,

$$F_i = \theta_0^{b_i} \cdot f_i \left( \frac{\theta_1}{\theta_0}, \dots, \frac{\theta_d}{\theta_0} \right) \quad \text{for } i = 1, \dots, n,$$

and  $F_{n+1} = \theta_0$  with  $u_{n+1} = -b_1u_1 - b_2u_2 - \dots - b_nu_n$ . The Chow ring of  $X = \mathbb{P}^d$  is  $\mathbb{Z}[H]/\langle H^{d+1} \rangle$ , where  $H$  represents the hyperplane class. By our genericity hypothesis, the  $r = n+1$  prime factors of  $F$  are smooth and global normal crossing. They correspond to the following divisor classes:

$$D_1 = b_1H, D_2 = b_2H, \dots, D_n = b_nH \text{ and } D_{n+1} = H.$$

Projective space  $\mathbb{P}^d$  is a smooth toric variety with  $d+1$  torus-invariant divisors  $\Delta_j$ , each having the same class  $H$ . Hence the formula in (11) specializes to

$$\frac{(1 - zH)^{d+1}}{(1 - zb_1H) \cdots (1 - zb_nH)(1 - zH)} = \frac{(1 - zH)^d}{(1 - zb_1H) \cdots (1 - zb_nH)}.$$

Since we work in the Chow ring of projective space  $\mathbb{P}^d$ , the coefficient of  $(zH)^d$  is the same as the coefficient of  $z^d$  in the generating function in (2).  $\square$

We now generalize our results from polynomials of fixed degrees to Laurent polynomials with fixed Newton polytopes. Recall that the *Newton polytope* of a Laurent polynomial  $f(\theta_1, \dots, \theta_d)$  is the convex hull of the set of exponent vectors of the monomials appearing in  $f$  with nonzero coefficient. Given a convex polytope  $P \subset \mathbb{R}^d$  with vertices in  $\mathbb{Z}^d$ , by a *generic Laurent polynomial with Newton polytope  $P$*  we will mean a sufficiently general  $\mathbb{C}$ -linear combination of monomials with exponent vectors in  $P \cap \mathbb{Z}^d$ .

In the next theorem we consider  $n$  Laurent polynomials  $f_1, f_2, \dots, f_n$  having respective Newton polytopes  $P_1, P_2, \dots, P_n$ . Because the  $f_i$ 's are Laurent polynomials, i.e., their monomials may have negative exponents, we only consider those critical points of  $f = f_1^{u_1} f_2^{u_2} \cdots f_n^{u_n}$  which lie in the algebraic torus  $(\mathbb{C}^*)^d$ . The number of such critical points (counted with multiplicity) will be called the *toric ML degree* of the rational function  $f$ .

Let  $P = P_1 + P_2 + \dots + P_n$  denote the *Minkowski sum* of the given Newton polytopes, and let  $X$  be the projective toric variety defined by  $P$ . Let  $\eta_1, \dots, \eta_s \in \mathbb{Z}^d$  be the primitive inner normal vectors of the facets of  $P$ . They span the rays of the fan of  $X$ . Let  $\Delta_1, \dots, \Delta_s$  denote the corresponding torus-invariant divisors on  $X$ . Each of the Newton polytopes  $P_i$  is the solution set of a system of linear inequalities of the specific form

$$P_i = \{x \in \mathbb{R}^d \mid \langle x, \eta_j \rangle \geq -a_{ij} \text{ for } j = 1, \dots, s\}.$$

The divisor on  $X$  defined by the Laurent polynomial  $f_i$  is linearly equivalent to  $D_i = \sum_{j=1}^s a_{ij} \Delta_j$ . The  $a_{ij}$  are integers which can be positive or negative.

The divisor on  $X$  defined by  $f = f_1^{u_1} f_2^{u_2} \cdots f_n^{u_n}$  is linearly equivalent to

$$\sum_{i=1}^n u_i D_i = \sum_{j=1}^s \left( \sum_{i=1}^n u_i a_{ij} \right) \cdot \Delta_j. \quad (12)$$

We abbreviate the *support* of this divisor by

$$I = \left\{ j \in \{1, \dots, s\} \mid \sum_{i=1}^n u_i a_{ij} \neq 0 \right\}. \quad (13)$$

A toric variety  $X$  is smooth if all the cones in its normal fan are unimodular.

**Theorem 7.** *If the toric variety  $X$  is smooth and the toric ML degree of the rational function  $f$  is finite then it is bounded above by the coefficient of  $z^d$  in the following generating function with coefficients in the Chow ring of  $X$ :*

$$\frac{\prod_{j \notin I} (1 - z \Delta_j)}{\prod_{i=1}^n (1 - z D_i)}. \quad (14)$$

*Equality holds if each  $f_i$  is generic with respect to its Newton polytope  $P_i$ .*

Note that Theorem 1 is the special case of Theorem 7 when  $P_i$  is the standard  $d$ -dimensional simplex  $\text{conv}\{0, e_1, \dots, e_d\}$  scaled by a factor of  $b_i$ .

*Proof.* Let us first assume that  $f_i$  is a generic Laurent polynomial with Newton polytope  $P_i$ . Let  $\mathbb{C}[x_1, \dots, x_s]$  be the homogeneous coordinate ring [8] of  $X$  with one variable for each torus-invariant divisor  $\Delta_j$ . Given a Laurent polynomial  $f_i(\theta)$  with Newton polytope  $P_i$ , the corresponding rational function on  $X$  is  $F_i(x)/x^{D_i}$  where  $D_i$  is as defined above and  $F_i$  is homogeneous of degree  $D_i$ . Therefore the rational function on  $X$  we are interested in is

$$F = x^{-\sum u_i D_i} \prod F_i(x).$$

We next show that the divisor of  $F$  is GNC. Note that  $F_i$  is a generic section of a line bundle on  $X$  that is generated by its sections. This implies, by the Bertini-Sard theorem and by induction on  $n$ , that the divisors  $\{F_i = 0\}$  meet transversally in the dense torus of  $X$ . For points in the boundary of  $X$ , we simply restrict to the torus orbit determined by the corresponding facet where the restricted  $F_i$ 's remain generic sections of the restricted bundles.

The reduced divisor of poles and zeros of  $F$  is  $D = \sum D_i + \sum_{j \in I} \Delta_j$  where  $I$  is defined as in (13). Since  $\sum D_i$  is the divisor corresponding to

$P$  it is ample on  $X$  by construction. So  $\sum D_i$  meets every curve on  $X$  and therefore so does  $D$  and we can apply Corollary 6. A variable  $x_j$  appears as a factor in  $F$  if and only if  $j \in I$ , in which case  $1 - z\Delta_j$  appears in both the numerator and denominator of (11), and we get the expression (14).

Consider now arbitrary Laurent polynomials  $f_1, \dots, f_n$  in  $\theta_1, \dots, \theta_d$  such that  $f = \prod f_i^{u_i}$  has only finitely many critical points in  $(\mathbb{C}^*)^d$ . Let  $\nu$  be the coefficient of  $z^d$  in (14). Let  $\mathbb{C}^m$  be the space of all  $n$ -tuples of Laurent polynomials with the given Newton polytopes. Consider the critical equations of  $f = \prod f_i^{u_i}$  and clear denominators. The resulting collection of  $d$  Laurent polynomials defines an algebraic subset  $\tilde{W}$  in the product space  $\mathbb{C}^m \times (\mathbb{C}^*)^d$ . Saturate  $\tilde{W}$  to remove any components along the hypersurfaces  $\{f_i = 0\}$  and get a new algebraic subset  $W$ . The map from  $W$  onto  $\mathbb{C}^m$  is dominant and generically finite, and the generic fiber of this map consists of  $\nu$  points.

Our given Laurent polynomials  $f_1, \dots, f_n$  represent a point  $\phi$  in  $\mathbb{C}^m$ . Let  $\theta^{(1)}, \dots, \theta^{(\kappa)}$  be the isolated critical points of  $f$ . For each  $i$ , consider any irreducible component  $W^{(i)}$  of  $W$  containing the point  $(\phi, \theta^{(i)})$  in  $W \subset \mathbb{C}^m \times (\mathbb{C}^*)^d$ . By Krull's Principal Ideal Theorem, the component  $W^{(i)}$  of  $W$  has codimension  $\leq d$  and hence it has dimension  $\geq m$ . As the generic fiber is finite, the dimension of  $W^{(i)}$  is exactly  $m$  and the projection to  $\mathbb{C}^m$  is dominant. Since  $\theta^{(i)}$  is an isolated solution of the critical equations, the projection map to  $\mathbb{C}^m$  is open [19, (3.10)], so the intersection of  $W^{(i)}$  with an open neighborhood of  $(\phi, \theta^{(i)})$  maps onto an open neighborhood of  $\phi$ . Hence every generic point  $\tilde{\phi}$  near  $\phi$  has a preimage  $(\tilde{\phi}, \tilde{\theta}^{(i)})$  near  $(\phi, \theta^{(i)})$ , and these preimages are distinct for  $i = 1, \dots, \kappa$ . We conclude that  $\kappa \leq \nu$ . This semicontinuity argument is called the ‘‘specialization principle’’ stated in Mumford's book [19, (3.26)] and also works when the  $\theta^{(i)}$  have multiplicities, as shown in Theorem 22 below.  $\square$

We illustrate Theorem 7 with two examples which we revisit in Section 5.

**Example 8.** Consider  $n$  generic polynomials  $f_1(\theta_1, \theta_2), \dots, f_n(\theta_1, \theta_2)$  where the support of  $f_i$  consists of monomials  $\theta_1^p \theta_2^q$  with  $0 \leq p \leq s_i$  and  $0 \leq q \leq t_i$ , and suppose the  $u_i$ 's are generic. The Newton polytope of  $f_i$  is the rectangle

$$P_i = \text{conv}\{(0, 0), (s_i, 0), (0, t_i), (s_i, t_i)\}.$$

The Minkowski sum of these rectangles is another rectangle, and  $X = \mathbb{P}^1 \times \mathbb{P}^1$ . In the numerator of (14), the contribution of the two torus-invariant divisors  $D$  and  $E$  corresponding to the left and the bottom edge of this

rectangle survives. The denominator comes from the product of the divisors of  $f_1, \dots, f_n$ :

$$\frac{(1 - zD)(1 - zE)}{(1 - (s_1D + t_1E)z)(1 - (s_2D + t_2E)z) \cdots (1 - (s_nD + t_nE)z)}.$$

Now, the coefficient of the term  $z^2$  modulo the Chow ring relations

$$D^2 = 0, \quad E^2 = 0, \quad D \cdot E = 1$$

gives the toric ML degree

$$\left(\sum_{i=1}^n s_i\right)\left(\sum_{j=1}^n t_j\right) + \sum_{k=1}^n s_k t_k - \sum_{i=1}^n (s_i + t_i) + 1. \quad (15)$$

**Example 9.** Let  $f_1, f_2, f_3$  be generic polynomials in  $\theta_1$  and  $\theta_2$  with supports

$$\begin{aligned} A_1 &= \{1, \theta_1, \theta_1\theta_2, \theta_1^2\}, \\ A_2 &= \{1, \theta_1, \theta_2, \theta_1\theta_2, \theta_1^2\}, \\ A_3 &= \{1, \theta_1\theta_2, \theta_1\theta_2^2\}. \end{aligned}$$

The corresponding Newton polytopes  $P_1, P_2, P_3$  are shown in Figure 1.

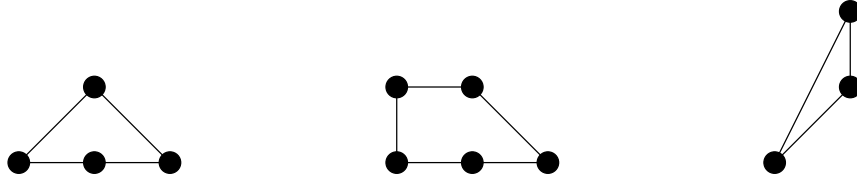


Figure 1: Three Newton polygons

The normal fan of the Minkowski sum has eight rays and is shown in Figure 2.

Theorem 7 applies because the toric surface  $X$  is smooth. We label the eight rays by  $x_1, \dots, x_8$  in counterclockwise order, starting with  $(1, 0)$ . The Chow ring  $A^*(X)$  is the polynomial ring  $\mathbb{Z}[x_1, \dots, x_8]$  modulo the ideal

$$\begin{aligned} &\langle x_1x_3, x_1x_4, x_1x_5, x_1x_6, x_1x_7, x_2x_4, x_2x_5, x_2x_6, x_2x_7, x_2x_8, \\ &x_3x_5, x_3x_6, x_3x_7, x_3x_8, x_4x_6, x_4x_7, x_4x_8, x_5x_7, x_5x_8, x_6x_8, \\ &x_1 - x_3 - x_4 - x_5 + x_7 + 2x_8, x_2 + x_3 - x_5 - x_6 - x_7 - x_8 \rangle. \end{aligned}$$

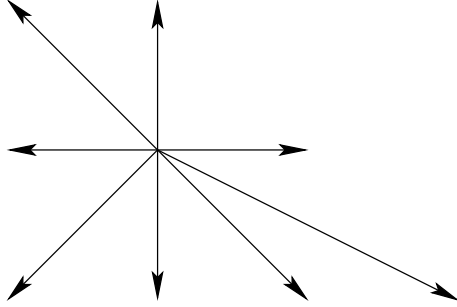


Figure 2: The fan of a smooth projective toric surface

The three divisors corresponding to the polygons  $P_1, P_2, P_3$  in Figure 1 are

$$\begin{aligned} D_1 &= 2x_3 + 2x_4 + 2x_5 + x_6 \\ D_2 &= 2x_3 + 2x_4 + 2x_5 + x_6 + x_7 + x_8 \\ D_3 &= x_4 + 3x_5 + 2x_6 + x_7 \end{aligned}$$

If all  $u_i$  are positive, then the support of the divisor  $u_1D_1 + u_2D_2 + u_3D_3$  is  $I = \{3, \dots, 8\}$ . It follows that the toric ML degree is the coefficient of  $z^2$  in

$$(1 - zx_1)(1 - zx_2)(1 - zD_1)^{-1}(1 - zD_2)^{-1}(1 - zD_3)^{-1}.$$

This coefficient is  $14x_1x_2$ , which means that the toric ML degree is 14.  $\square$

The *toric ML degree of the model  $\mathbf{f}$*  is the toric ML degree defined above for generic  $u$ . In this case, there is no cancellation among the coefficients in (13), and  $I$  is the set of all indices  $j$  such that for some  $P_i$  the supporting hyperplane normal to  $\eta_j$  does not pass through the origin. The toric ML degree of  $\mathbf{f}$  is a numerical invariant of the polytopes  $P_1, \dots, P_n$ . A combinatorial formula for this invariant will be presented in Theorem 15 of Section 5.

## 4 Bounded Regions in Arrangements

As in the Introduction, we consider  $n$  polynomials  $f_1, \dots, f_n$  in  $d$  unknowns  $\theta_1, \dots, \theta_d$ . We now assume that all coefficients of the  $f_i$ 's are real numbers, and we also assume that  $u_1, \dots, u_n$  are positive integers. However, we do not assume that the union of the divisors of the  $f_i$ 's has global normal crossings. This is the case of interest in statistics. Consider the arrangement of

hypersurfaces defined by the  $f_i$ 's and let  $V_{\mathbb{R}} = \mathbb{R}^d \setminus \bigcup_{i=1}^n \{f_i = 0\}$  be the complement of this arrangement. A connected component of  $V_{\mathbb{R}}$  is a *bounded region* if it is bounded as a subset of  $\mathbb{R}^d$ . Then the following observation holds.

**Proposition 10.** *For any polynomial map  $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^n$  and any  $u \in \mathbb{N}_{>0}^n$ ,*

$$\begin{aligned} & \#\{\text{bounded regions of } V_{\mathbb{R}}\} \\ \leq & \#\{\text{critical points of } f_1^{u_1} \cdots f_n^{u_n} \text{ in } \mathbb{R}^d\} \\ \leq & \text{ML degree of } \mathbf{f}. \end{aligned}$$

*Proof.* The function  $f = f_1^{u_1} \cdots f_n^{u_n}$  is continuous, and on the boundary of the closure of each bounded region its value is zero. Hence it has to have at least one (real) critical point in the interior of each region. The second inequality holds trivially, since the ML degree was defined as the number of critical points of  $f_1^{u_1} \cdots f_n^{u_n}$  in  $\mathbb{C}^d$ , counted with multiplicities.  $\square$

This observation raises the question whether the inequalities above could be realized as equalities. We next show that this is the case when  $f_1, \dots, f_n$  are quadrics in the plane. Here the ML degree is  $2n^2 - 2n + 1$  by Theorem 1.

**Proposition 11.** *For each  $n$ , there are  $n$  quadrics  $f_1, \dots, f_n$  in  $\mathbb{R}^2$  such that*

$$\#\{\text{bounded regions of } V_{\mathbb{R}}\} = \text{ML degree of } \mathbf{f} = 2n^2 - 2n + 1.$$

*Hence all critical points are real.*

*Proof.* We will take  $n$  quadrics that define “nested” ellipses with center at the origin, as suggested by Figure 3. The proof follows by induction: assume we have  $2(n-1)^2 - 2(n-1) + 1$  bounded regions with  $n-1$  ellipses. Observe that the  $(n-1)$ st ellipse contains  $2n-3$  bounded regions. Then we add a new long and skinny ellipse which replaces the  $2n-3$  regions with  $3(2n-3) + 2$  regions. The total count comes out to be  $2n^2 - 2n + 1$ .  $\square$

We will see such an equality holding for  $n$  linear hyperplanes in  $\mathbb{R}^d$  below. However, even in the plane  $\mathbb{R}^2$ , the number of critical points and the number of bounded regions of  $V_{\mathbb{R}}$  diverge for curves of degree  $\geq 3$ . Theorem 1 implies that for  $n$  generic plane curves of degrees  $b_1, \dots, b_n$  the ML degree is

$$\sum_{i=1}^n b_i(b_i - 2) + \sum_{i < j} b_i b_j + 1.$$

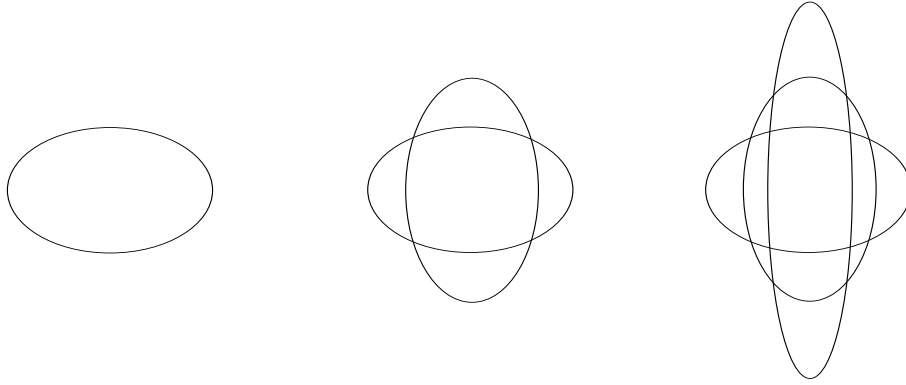


Figure 3: The “nested” ellipse construction

The optimal upper bound for the number of bounded regions of  $V_{\mathbb{R}}$  is smaller than the ML degree, by the following unpublished result due to Oleg Viro.

**Theorem 12.** (Viro) *Let  $f_1, \dots, f_n$  be real plane curves of degrees  $b_1, \dots, b_n$ , and let  $K$  be the number of odd degree curves among them. Then*

$$\#\{\text{bounded regions of } V_{\mathbb{R}}\} \leq \sum_{i=1}^n \frac{(b_i - 1)(b_i - 2)}{2} + \sum_{i < j} b_i b_j + 1 - K,$$

and this bound is optimal.

*Proof.* The proof is by induction. For  $n = 1$  the bound above is Harnack’s inequality [15], and it is optimal. Suppose the statement is proved for  $n - 1$  curves, and  $f_n$  defines a curve of degree  $b_n$ . We will take this new curve so that it has the maximum number of bounded regions allowed by Harnack’s inequality, i.e.  $B_n := (b_n - 1)(b_n - 2)/2 + 1$  if  $b_n$  is even, and one less than that if  $b_n$  is odd. One can achieve  $B_n$  by taking a curve with  $B_n - 1$  unnested ovals and one more distinguished piece (that gives an extra bounded region when  $b_n$  is even) such that some line intersects this distinguished piece in exactly  $b_n$  points. We can arrange this last curve in such a way that when we superimpose the distinguished piece on the arrangement given by  $n - 1$  curves, the last curve will intersect the curve given by  $f_i$  in  $b_i b_n$  points. Now if we trace this last distinguished piece, every time we encounter an intersection point an extra bounded region is created, except the last point in case  $b_n$  is odd. Together with the remaining  $B_n - 1$  ovals we get the bound.  $\square$

In order to get any meaningful lower bound on the number of bounded regions of  $V_{\mathbb{R}}$  one needs to make some assumptions. Without any assumptions the lower bound is zero: for  $f_i$  of even degree we take an empty (real) curve, and for  $f_i$  of odd degree we take the union of an empty curve with a line. If we let all the lines intersect in a single point there will not be any bounded region. If we insist on at least having a GNC configuration, then by the same construction the lower bound we get is the number of bounded regions in a generic arrangement of  $K$  lines where  $K$  is the number of odd degrees  $b_i$ . This idea leads us to studying the ML degree of a hyperplane arrangement.

**Theorem 13.** *Let  $\mathbf{f}$  be given by  $n$  linear polynomials  $f_1, \dots, f_n$  with real coefficients. Then the ML degree of  $\mathbf{f}$  is equal to the number of the bounded regions of  $V_{\mathbb{R}}$ , and all critical points of the optimization problem (1) are real.*

This theorem does not assume any hypothesis such as global normal crossing. Under the GNC hypothesis, the hyperplanes would be in general position and the number of bounded regions equals  $\binom{n-1}{d}$ , as predicted by Theorem 1. Theorem 13 is essentially due to Varchenko [25]. We shall give a new proof.

*Proof.* In light of Proposition 10, we need to show that the number of bounded regions of  $V_{\mathbb{R}}$  equals the number of complex solutions of the ML equations. Let  $f_i = \sum_{j=1}^d a_{ij}\theta_j + c_i$  for  $i = 1, \dots, n$ . The ML equations are

$$\sum_{i=1}^n \frac{u_i a_{i1}}{f_i} = \dots = \sum_{i=1}^n \frac{u_i a_{id}}{f_i} = 0. \quad (16)$$

Consider the map  $\psi : \mathbb{C}^{d+1} \rightarrow \mathbb{C}^n$  given by  $\psi(\theta_0, \dots, \theta_d) = (1/F_1, \dots, 1/F_n)$ . Here  $F_i = c_i\theta_0 + \sum_{j=1}^d a_{ij}\theta_j$  is the homogenization of  $f_i$ . We let  $\bar{\mathcal{H}}$  be the central hyperplane arrangement in  $\mathbb{R}^{d+1}$  given by the  $F_i$ . We assume that the intersection of all the hyperplanes in  $\bar{\mathcal{H}}$  is just the origin; otherwise, the linear forms  $F_i$  depend on fewer than  $d$  coordinates, and then we get infinitely many critical points. The Zariski closure of  $\text{im}(\psi)$  in  $\mathbb{P}^{n-1}$  is a  $d$ -dimensional complex variety  $\mathcal{V}$ . The solution set on  $\mathcal{V}$  of the  $d$  linear equations

$$\sum_{i=1}^n (u_i a_{i1}) y_i = \dots = \sum_{i=1}^n (u_i a_{id}) y_i = 0$$

consists of finitely many points provided  $u_1, \dots, u_n$  are generic. Obviously, the solutions to (16) lift to such complex solutions. In other words, the degree of the projective variety  $\mathcal{V}$  is an upper bound on the ML degree of  $\mathbf{f}$ .

Now we will compute the degree of  $\mathcal{V}$ . This variety is the projective spectrum of the  $\mathbb{N}$ -graded algebra  $R = \mathbb{C}[1/F_i : i = 1, \dots, n]$  where  $\deg(1/F_i) = 1$ . Terao [24, Theorem 1.4] showed that the Hilbert series of  $R$  is equal to

$$\sum_{X \in \mathcal{L}} (-1)^{\text{codim}(X)} \mu(X) \left( \frac{t}{1-t} \right)^{\text{codim}(X)} \quad (17)$$

Here  $\mu$  is the Möbius function of the intersection lattice  $\mathcal{L}$  of the arrangement  $\bar{\mathcal{H}}$ . From this series we shall determine the leading coefficient of its Hilbert polynomial. This coefficient has the form  $e/d!$ , where  $e$  is the degree of  $\mathcal{V}$ .

For large enough  $r$ , the coefficient of  $t^r$  in the Hilbert series (17) equals

$$\sum_{i=0}^{d+1} (-1)^i \binom{r-1}{i-1} \sum_{\text{codim}(X)=i} \mu(X). \quad (18)$$

This is the Hilbert polynomial of the graded algebra  $R$ . Its leading term is

$$(-1)^{d+1} \mu(0) \frac{r^d}{d!}.$$

We conclude that the degree of the projective variety  $\mathcal{V}$  is  $(-1)^{d+1} \mu(0)$ . By Zaslavsky [26], this number equals the number of bounded regions of  $V_{\mathbb{R}}$ .  $\square$

**Example 14.** A family of important statistical models where Theorem 13 applies is the *linear polynomial model* of [22]. Such a model is given by a polynomial in  $r$  unknowns  $x = (x_1, \dots, x_r)$  with indeterminate coefficients,

$$p(x) = \sum_{j=1}^d \theta_j x^{a_j} \quad (\text{with } a_j \in \mathbb{N}^r),$$

together with  $n$  data points  $v_1, \dots, v_n \in \mathbb{R}^r$ . The model is parametrized by

$$f_1(\theta) = \sum_{j=1}^d \theta_j v_1^{a_j}, \dots, f_n(\theta) = \sum_{j=1}^d \theta_j v_n^{a_j}.$$

The ML degree is the number of bounded regions of this arrangement.  $\square$

## 5 Polytopes and Resolution of Singularities

We now return to the setting of Section 3, with the aim of relaxing the restrictive smoothness hypothesis in Theorem 7. Our aim is to derive a combinatorial formula for the toric ML degree of any model  $\mathbf{f}$  defined by generic Laurent polynomials satisfying a mild hypothesis. The derivation of Theorem 15 involves resolution of singularities in the toric category. In the end of the section we shall comment on using resolution of singularities for bounding the ML degree in general.

Given a polytope  $P$  in  $\mathbb{R}^d$  and a linear functional  $v$  on  $\mathbb{R}^d$ , we write

$$P^v = \{ p \in P \mid \forall p' \in P : \langle v, p \rangle \leq \langle v, p' \rangle \}$$

for the face of  $P$  at which  $v$  attains its minimum. Two linear functionals  $v$  and  $v'$  are *equivalent* if  $P^v = P^{v'}$ . The equivalence classes are the relative interiors of cones of the *inner normal fan*  $\Sigma_P$ . If  $\sigma$  is a cone in  $\Sigma_P$ , or  $\sigma$  is a cone in any fan which refines  $\Sigma_P$ , then we write  $P^\sigma = P^v$  for  $v$  in the relative interior of  $\sigma$ . If  $f$  is a polynomial with Newton polytope  $P$  then  $f^\sigma$  denotes the *leading form* consisting of all terms of  $f$  which are supported on  $P^\sigma$ .

As in Section 3, let  $f_1, \dots, f_n$  be Laurent polynomials with Newton polytopes  $P_1, \dots, P_n \subset \mathbb{R}^d$ . Consider any fan  $\Sigma$  which is a common refinement of the inner normal fans  $\Sigma_{P_1}, \dots, \Sigma_{P_n}$ . Suppose  $\tau$  is a cone in  $\Sigma$  and let  $k$  be the dimension of  $(P_1 + \dots + P_n)^\tau$ . There exists a  $k$ -dimensional linear subspace  $L$  of  $\mathbb{R}^d$  and vectors  $q_1, \dots, q_n \in \mathbb{R}^d$  such that  $q_i + P_i^\tau$  lies in  $L$  for all  $i = 1, \dots, n$ . The subspace  $L$  is unique and satisfies  $L \cap \mathbb{Z}^d \simeq \mathbb{Z}^k$ . Let  $V(\cdot, \dots, \cdot)$  denote the *normalized mixed volume* on the subspace  $L$ . Here “normalized” refers to the lattice  $L \cap \mathbb{Z}^d$ , as is customary in toric geometry [12]. For any  $k$ -element subset  $\{i_1, \dots, i_k\}$  of  $\{1, 2, \dots, n\}$  we abbreviate

$$V(P_{i_1}, \dots, P_{i_k}; \tau) = V(q_{i_1} + P_{i_1}^\tau, \dots, q_{i_k} + P_{i_k}^\tau) \quad \text{if } \text{codim}(\tau) = k, \quad (19)$$

and  $V(P_{i_1}, \dots, P_{i_k}; \tau) = 0$  if  $\text{codim}(\tau) > k$ . If  $k = d$  and  $\tau = \{0\}$  then we simply write  $V(P_{i_1}, \dots, P_{i_d})$  for the mixed volume in (19). If  $k = 0$  and  $\tau$  is full-dimensional then (19) equals 1; this happens in the last sum of (20).

We are now ready to state our more general toric ML degree formula. As in Section 3, let  $X$  be the toric variety corresponding to the Minkowski sum  $P = P_1 + \dots + P_n$  and  $\Sigma_X$  the normal fan with rays  $\eta_1, \dots, \eta_s$ . We consider the function  $f = f_1^{u_1} \dots f_n^{u_n}$ . Each polytope  $P_i$  corresponds to a divisor  $D_i$  so the divisor of  $f$  is  $D = \sum u_i D_i$ . Let  $I$  be the support of  $D$  as in (13). Label the rays of  $\Sigma_X$  so that  $\{1, \dots, r\}$  are the indices not in  $I$ .

For each subset  $J$  of  $\{1, \dots, r\}$  let  $\tau_J$  denote the smallest cone of  $\Sigma$  which contains the vectors  $\eta_j$  for  $j \in J$ . If no such cone exists then  $\tau_J$  is just a formal symbol and the expression (19) is declared to be zero for  $\tau = \tau_J$ . The mild smoothness hypothesis we need is that every singular cone of  $\Sigma$  contains at least one ray from  $I$ . Equivalently all cones  $\tau_J$  are smooth.

**Theorem 15.** *Suppose every singular cone of  $\Sigma_X$  contains some ray in the support of the divisor  $D$ . Then, the toric ML degree of the rational function  $f$  is bounded above by the following alternating sum of mixed volumes:*

$$\begin{aligned} & \sum_{1 \leq i_1 \leq \dots \leq i_d \leq n} V(P_{i_1}, \dots, P_{i_d}) - \sum_{\substack{j \in \{1, \dots, r\} \\ 1 \leq i_1 \leq \dots \leq i_{d-1} \leq n}} V(P_{i_1}, \dots, P_{i_{d-1}}; \tau_{\{j\}}) \\ & + \sum_{\substack{\{j_1, j_2\} \subset \{1, \dots, r\} \\ 1 \leq i_1 \leq \dots \leq i_{d-2} \leq n}} V(P_{i_1}, \dots, P_{i_{d-2}}; \tau_{\{j_1, j_2\}}) + \dots + \\ & (-1)^d \sum_{\{j_1, \dots, j_d\} \subset \{1, \dots, r\}} V(\emptyset; \tau_{\{j_1, \dots, j_d\}}). \end{aligned} \quad (20)$$

Equality holds if each  $f_i$  is generic relative to its Newton polytope  $P_i$ .

*Proof.* In order to apply Corollary 6 we must resolve the singularities of  $X$ . For toric varieties this is done in two steps. First we get a simplicial toric variety without adding any new rays to the fan. Second we resolve the remaining singular (but simplicial) cones by adding new rays. This procedure is described in detail in [12]. Typically the first step involves taking the pulling subdivision at each ray in the fan. However, under the given hypothesis it is enough to perform pulling subdivisions only at the rays in the support of  $D$  to obtain a simplicial fan  $\Sigma_{\tilde{X}}$ . This fine detail will be important below. Our hypothesis holds for this intermediate fan as well, and subsequently we take a smooth refinement  $\Sigma_{X'}$  of  $\Sigma_{\tilde{X}}$  by adding new rays in the relative interiors of each of the singular cones. Let  $\pi : X' \rightarrow X$  be the induced map.

We will show that we get no new critical points under the resolution. Hence the number of critical points can be computed on  $X'$ . We finally claim that the Chern class formula expands into the given combinatorial formula.

We investigate critical points of the pullback of our rational function:

$$F' = \pi^*(F) = (x^{-\sum u_i \pi^*(D_i)}) \prod \pi^*(F_i(x)).$$

For generic  $f_i$ , the same argument as in the proof of Theorem 7 shows that the reduced divisor of poles and zeros  $D'$  of  $F'$  is GNC. What we must show is that all critical points of  $F'$  on  $X'$  are off the exceptional locus hence actually critical points of  $F$  on  $X$ .

There are two types of new cones in  $\Sigma_{X'}$ . The first come from the triangulation step. By our construction, any such cone must contain a ray  $\eta_j$  in the support of  $D$ . This ray corresponding to the strict transform under  $\pi$  is in the support of  $\sum u_i \pi^*(D_i)$ , and its variable appears as a factor in  $F'$ . By part 1 of Theorem 4,  $F'$  has no critical points along the torus-invariant divisor  $\Delta'_j$ , hence no critical points on any torus orbit contained in  $\Delta'_j$ .

The second type of new cone comes from the desingularization step. These cones all contain at least one new ray  $\eta_E$  corresponding to an exceptional divisor  $\Delta_E$  in  $X'$ . We will show there are no critical points on  $\Delta_E$ . Equivalently we show that there are no critical points on each torus orbit contained in  $\Delta_E$ .

Given a torus orbit let  $\tau_E$  be the corresponding cone of  $\Sigma'$  containing  $\eta_E$ . There is some minimal cone  $\tau$  of  $\Sigma$  containing  $\tau_E$ . Let  $\tau'$  be any cone of  $\Sigma'$  containing  $\tau_E$  that is maximal with respect to being contained in  $\tau$ . Since  $\tau$  is refined in  $\Sigma'$  it must be a singular cone, and so by the hypothesis it has some generating ray in the support of  $\sum u_i D_i$ , or equivalently the linear function of this Cartier divisor is not identically zero on  $\tau$ . The pullback keeps the same linear functional which cannot be zero on the subset  $\tau'$  of  $\tau$ . As a consequence,  $\tau'$  contains a ray  $\eta_j$  in the support of  $\sum u_i \pi^*(D_i)$ . This means  $x_j^c$  appears as a factor in  $F'$  for some nonzero integer  $c$ .

If  $\eta_j$  is a generator of  $\tau_E$ , then as above there are no critical points on  $\Delta'_j$  and thus no critical points on the orbit corresponding to  $\tau_E$ . Suppose on the other hand  $\eta_j$  is not a generator of  $\tau_E$ . Let  $x_{E_1}, \dots, x_{E_k}$  be the variables corresponding to the generators of  $\tau_E$  in an affine chart of  $X'$  that contains  $\tau'$ . Note that the variable  $x_j$  corresponding to  $\eta_j$  is not among these variables. Because  $\tau$  is the minimal cone of  $\Sigma$  containing  $\tau_E$ , the face  $P_i^{\tau_E}$  is contained in the face  $P_i^\tau$ , and hence it is contained in the face  $P_i^{\tau'}$ . So  $(F'_i)^{\tau_E}$ , obtained by setting  $x_{E_1}, \dots, x_{E_k}$  to zero, does not contain any of the variables corresponding to  $\tau'$ ; in particular it doesn't contain  $x_j$ . On the other hand,  $F'_i = (F'_i)^{\tau_E} + G'_i$  where  $G'_i$  is in the ideal generated by  $x_{E_1}, \dots, x_{E_k}$ . Since  $x_j$  is not among the  $x_{E_i}$  we have

$$\left(\frac{\partial F'_i}{\partial x_j}\right)\Big|_{x_E=0} = \frac{\partial((F'_i)^{\tau_E})}{\partial x_j} = 0,$$

where  $x_E = 0$  means  $x_{E_1} = \cdots = x_{E_k} = 0$ . Hence we have

$$\left(\frac{\partial(\log F')}{\partial x_j}\right)|_{x_E=0} = \frac{c}{x_j} + \sum_i \left(\frac{\partial F'_i}{\partial x_j}\right)|_{x_E=0} \neq 0.$$

We conclude that  $F'$  has no critical points on the torus orbit corresponding to  $\tau_E$  as desired.

Thus the toric ML degree of  $f$  on  $X'$  is the same as that on  $f$ . Since  $F'$  has no critical points on the exceptional locus and  $D$  is ample on  $X$ ,  $\pi^*(D)$  meets any curve off the exceptional locus and therefore the ML degree must be finite. It is computed in the Chow ring of  $X'$  as the coefficient of  $z^d$  in

$$\frac{\prod_{j \notin I} (1 - z\Delta'_j) \prod_k (1 - z\Delta_{E_k})}{\prod_{i=1}^n (1 - zD'_i)}.$$

Here  $\Delta'_j$  are the strict transforms of the  $\Delta_j$  not in the support of  $\sum u_i D_i$ . The  $\Delta_{E_k}$  are those exceptional divisors which are not in the support of  $\sum u_i \pi_i^*(D_i)$ , and  $D'_i$  are the proper transforms of the divisor classes of the  $F_i$ 's.

We can now expand our Chern class product replacing  $(1 - zD'_i)$  in the denominator by  $1 + zD'_i + z^2(D'_i)^2 + \cdots + z^d(D'_i)^d$  in the numerator. The intersection product of any collection of prime torus-invariant divisors is the cycle of the cone they span or 0 if there is no such cone. Hence the coefficient of  $z^d$  is the sum of all terms of the form

$$(-1)^k D'_{i_1} D'_{i_2} \cdots D'_{i_{d-c}} \cdot \tau^c.$$

Here  $1 \leq i_1 \leq \cdots \leq i_{d-c} \leq n$  and  $\tau^c$  ranges over all dimension  $c$  cones of  $\Sigma'$  spanned by rays not in the support of  $\sum u_i D'_i$ . This product is exactly the mixed volume  $V(P_{i_1}, \dots, P_{i_{d-c}}; \tau^c)$ .

To finish we note that if  $\tau^c$  contains an exceptional divisor  $\Delta_{E_k}$ , the minimal cone  $\tau$  of  $\Sigma$  containing  $\tau^c$  must have dimension strictly larger than  $c$ . This is because  $\tau^c$  does not have any rays in the support of  $\sum u_i D'_i$  hence is not maximal in  $\tau$ . As a consequence all of the faces  $P_{i_j}^{\tau^c}$  have a translate that lies in a subspace of dimension strictly less than  $d - c$  and the corresponding mixed volume is 0. In conclusion, the exceptional divisors do not contribute to the top Chern class product and the formula reduces to the stated one.  $\square$

In two variables we recover a particularly simple formula:

**Corollary 16.** *Let  $f_1, \dots, f_n$  be generic Laurent polynomials in two variables  $(\theta_1, \theta_2)$  with Newton polygons  $P_1, \dots, P_n$ . If the origin lies on none of the lines spanned by edges of their Minkowski sum  $P = P_1 + \dots + P_n$  then the toric ML degree equals the area of  $P$  plus the areas of each of the  $P_i$ .*

*Proof.* This is a special case of Theorem 15 when no facets pass through the origin. Therefore the only term is

$$\sum_{i=1}^n \sum_{j=i}^n V(P_i, P_j).$$

The Euclidean area of each  $P_i$  is  $V(P_i, P_i)/2$ . The Euclidean area of the Minkowski sum  $P = \sum P_i$  equals  $\frac{1}{2} \sum_i V(P_i, P_i) + \sum_{i < j} V(P_i, P_j)$ . The stated formula is the sum of these expressions.  $\square$

Now that we are equipped with the volume formulas in Theorem 15 and Corollary 16, let us revisit the two-dimensional examples from Section 3.

**Example 17.** The Newton polygons  $P_1, \dots, P_n$  in Example 8 are axis-parallel parallelograms. The first term in the formula (15) is the area of their Minkowski sum  $P_1 + \dots + P_n$ , and the second term is the sum of the areas of the  $P_i$ , as in Corollary 16. The third and fourth term are the two correction terms stemming from the fact that the origin is a vertex of each Newton polytope. These terms disappear if we replace one  $f_i$  by  $\theta_1 \theta_2 f_i$ .

The number 14 in Example 9 can also be derived using Theorem 15. The three polygons in Figure 1 have areas  $1, \frac{3}{2},$  and  $\frac{1}{2}$  respectively. The area of their sum is 15. The two divisors corresponding to  $x_1$  and  $x_2$  pass through the origin and yield correction terms of 1 and 4 respectively. Finally we add back 1 for the vertex at the origin. Altogether  $1 + \frac{3}{2} + \frac{1}{2} + 15 - 1 - 4 + 1 = 14$ .  $\square$

Our discussion so far indicates that we get the sharpest results when  $X$  is smooth and  $D$  is GNC. In the toric case the smoothness hypothesis could be largely removed as in Theorem 15. The GNC condition can also be relaxed for certain other cases as we saw in the previous section. In general, if the pair  $(X, D)$  does not satisfy the smoothness and GNC hypotheses then we must appeal to Hironaka's theorem on resolution of singularities (see e.g. [16]). This furnishes a proper projective morphism  $\pi : X' \rightarrow X$  such that  $X'$  is smooth and  $\pi^{-1}(D)$  has GNC. We need to compute the divisor  $\pi^*(\text{div}(f))$  of

the pullback of the function  $f$ . If  $D'_i$  is the proper transform of the divisor  $D_i$  and  $E_1, \dots, E_k$  are the exceptional divisors of  $\pi$  then

$$\pi^*(\text{div}(f)) = \sum_{i=1}^r u_i D'_i + \sum_{j=1}^k \mu_j E_j,$$

where  $\mu_j$  are certain (possibly negative) integers. These integers are  $\mu_j = \sum_{i=1}^r u_i m_{ij}$  where  $m_{ij}$  is the multiplicity of the full transform of  $D_i$  along  $E_j$ . The underlying reduced divisor is  $D' := \sum_{i=1}^r D'_i + \sum_{j:\mu_j \neq 0} E_j$ . The number of critical points is now gotten by applying Theorem 4 to  $(X', D')$  instead of  $(X, D)$ . This procedure can be very complicated in practice. We illustrate it with a simple example.

**Example 18.** Let  $d = 2$ ,  $n = 4$ ,  $f_1 = x$ ,  $f_2 = y$ ,  $f_3 = (x-1)^2 + (y-1)^2 - 2$ , and  $f_4 = (x+1)^2 + 2(y-2)^2 - 9$ . The divisor  $D$  is not a GNC divisor since at the origin all the four curves defined by  $f_1, \dots, f_4$  meet. In order to resolve this singularity we blow up  $X = \mathbb{P}^2$  at  $(0 : 0 : 1)$  to obtain  $X'$  which is smooth. We note that  $c_{\text{tot}}(\Omega_{X'}^1)$  is  $(1-zE)(1-zH)(1-zH')^2$  where  $H$  is the proper transform of the generic hyperplane section,  $E$  is the exceptional divisor, and  $H'$  is the proper transform of a line through the origin.

We have four cases: we consider first the general case where  $u_1 + u_2 + u_3 + u_4 \neq 0$  and  $u_1 + u_2 + 2u_3 + 2u_4 \neq 0$ . In this case  $D'$  consists of the proper transforms of the four original curves, the exceptional divisor, and the pullback of the line at infinity. After cancellations, we just need to compute the coefficient of  $z^2$  in  $\frac{1}{(1-C_1z)(1-C_2z)}$  where  $C_1$  and  $C_2$  are the irreducible divisors corresponding respectively to the circle and the ellipse. This coefficient is  $C_1^2 + C_2^2 + C_1 \cdot C_2$ . In  $X'$  the two curves intersect in three points, and their self-intersection also yields three points. Hence the ML degree is nine.

In the special case where  $u_1 + u_2 + u_3 + u_4 = 0$ , we need to compute the coefficient of  $z^2$  in  $\frac{1-zE}{(1-C_1z)(1-C_2z)}$ , which is  $C_1^2 + C_2^2 + C_1 \cdot C_2 - E \cdot C_1 - E \cdot C_2$ . Since  $E \cdot C_i = 1$  the ML degree drops down to seven. If  $u_1 + u_2 + 2u_3 + 2u_4 = 0$  then the coefficient of  $z^2$  in  $\frac{1-zH}{(1-C_1z)(1-C_2z)}$  is  $C_1^2 + C_2^2 + C_1 \cdot C_2 - H \cdot C_1 - H \cdot C_2$ . Since  $H \cdot C_i = 2$ , the ML degree is five. Finally, if both  $u_1 + u_2 + u_3 + u_4$  and  $u_1 + u_2 + 2u_3 + 2u_4$  are zero, then the coefficient of  $z^2$  in  $\frac{(1-zH)(1-zE)}{(1-C_1z)(1-C_2z)}$  is  $C_1^2 + C_2^2 + C_1 \cdot C_2 - H \cdot C_1 - H \cdot C_2 - E \cdot C_1 - E \cdot C_2 + H \cdot E$ : since  $H \cdot E = 0$ , the ML degree further drops down to three.

The number of bounded regions of the complement of the four curves in  $\mathbb{R}^2$  is seven. By Proposition 10, this is a lower bound on the ML degree when

all  $u_i$  are positive. This example shows that, for specific *negative* values of the  $u_i$ 's, the number of critical points may be smaller than this lower bound.  $\square$

## 6 ML degree and Euler characteristic

A well-known result in the theory of hyperplane arrangements [20] states that the number of bounded regions of a real arrangement equals the Euler characteristic of the complement of its complexification. The Euler characteristic is  $(-1)^{d+1}\mu(0)$  where  $\mu(0)$  is the Möbius number of the intersection lattice, and, by Theorem 13, this is precisely the ML degree of the associated linear model. Here we extend this relationship between topology and the ML degree to statistical models which are given by nonlinear polynomials  $f_i$ . Working in the general setting of Section 2, we shall prove the following:

**Theorem 19.** *Let  $X$  be a smooth complete algebraic variety over  $\mathbb{C}$  of dimension  $d$ , and let  $D$  be the reduced divisor associated to  $f = f_1^{u_1} \cdots f_n^{u_n}$ . Assume that the hypotheses (a), (b) and (c) below hold. Then the ML degree equals  $(-1)^d e_{top}(X \setminus D)$  where  $e_{top}$  is the topological Euler characteristic.*

Invoking Hironaka's theorem on resolution of singularities, we fix a blow up  $\pi : \tilde{X} \rightarrow X$  such that  $\tilde{X}$  is smooth, and the rational function  $f$  pulls back to a proper morphism  $\tilde{f}$  onto  $\mathbb{P}_{\mathbb{C}}^1$ . The three **hypotheses** are as follows:

- (a) The inverse image  $D' := \pi^{-1}(D)$  of the divisor  $D$  can be written as  $D' = \tilde{D} + D_H$  where  $\tilde{D}$  is the support of the divisor  $\text{div}(\tilde{f})$  of the pullback  $\tilde{f}$  of the rational function  $f$ , while  $D_H$  is the horizontal divisor consisting of the sum of all components of  $D'$  which map onto  $\mathbb{P}_{\mathbb{C}}^1$ .
- (b) The restriction of  $\tilde{f}$  to  $D_H \setminus \tilde{D}$  is a topological fiber bundle over  $\mathbb{C}^* = \mathbb{P}_{\mathbb{C}}^1 \setminus \{0, \infty\}$ .
- (c) The number of critical points of  $\tilde{f}$  on  $\tilde{X} \setminus \tilde{D}$  is finite.

**Remark 20.** Hypothesis (a) is crucial and depends on the exponents  $u_i$ . For instance, consider the rational function  $f = (y - x^2)(y + x^2)^{-1}$  on  $X = \mathbb{P}_{\mathbb{C}}^2$ . We get  $\tilde{X}$  by blowing up the origin twice. The exceptional curve of the first blow-up belongs to the fiber  $\{\tilde{f} = 1\}$  and hence is not supported on  $\text{div}(\tilde{f})$ . Hypothesis (a) is not satisfied for this example. If we take instead

$f = (y - x^2)(y + x^2)^{-2}$  then hypothesis (a) is satisfied because the exceptional curve belongs to the fiber  $\{f = \infty\}$  and is hence supported on  $\text{div}(\tilde{f})$ ,

Hypothesis (b) implies that the cohomology ranks of  $D_H \setminus \tilde{D}$  can be computed from that of  $\mathbb{C}^*$  and the fibers using Künneth's formula. The alternating sum of the ranks is zero for the fibers, and we get  $e_{top}(D_H \setminus \tilde{D}) = 0$ . In fact, the hypothesis (b) could be replaced by the more general condition  $e_{top}(D_H \setminus \tilde{D}) = 0$ .

Any proper map  $f : X \rightarrow \mathbb{C}^*$  for  $X$  smooth is a topological fiber bundle if it is a submersion, i.e.  $df \neq 0$  for all points in  $X$ . Therefore to check hypothesis (b), we need only find a controlled stratification (in the sense of Thom-Mather theory [17]) of  $D_H$  into locally closed smooth sets such that for all points on each strata  $S$ ,  $d(\tilde{f}|_S) \neq 0$ . In particular, this last condition will imply that the critical points of  $\tilde{f}$  on  $\tilde{X} \setminus \tilde{D}$  are the same as the critical points of  $f$  on  $X \setminus D$ .

*Proof of Theorem 19:* Our method follows closely the proof of the Lefschetz hyperplane section theorem (cf. [1, 2]). Moreover, since the complement is not necessarily compact we shall use Borel-Moore homology [4] (see also [6]). We note that for compact spaces  $X$  the ordinary homology groups coincide with the Borel-Moore homology groups. In the Borel-Moore homology theory we have the following useful exact sequence to be used below: if  $X$  is locally compact,  $F$  is closed in  $X$ , and  $U := X \setminus F$ , we have

$$\dots \rightarrow H_i(F) \rightarrow H_i(X) \rightarrow H_i(U) \rightarrow \dots \quad (21)$$

Thus in this situation the Borel-Moore Euler characteristic is additive:

$$e_{BM}(X) = e_{BM}(F) + e_{BM}(U). \quad (22)$$

Finally, if  $U$  is an even-dimensional orientable manifold then Poincaré duality holds between Borel Moore homology and ordinary cohomology, and  $e_{BM}(U)$  coincides with the topological Euler number  $e_{top}(U)$ . In our situation we get

$$e_{top}(X \setminus D) = e_{top}(\tilde{X} \setminus D') = e_{BM}(\tilde{X} \setminus D') = e_{BM}(\tilde{X} \setminus \tilde{D}) - e_{BM}(D' \setminus \tilde{D}).$$

The last equation follows from (22). Hypothesis (b) implies  $e_{BM}(D' \setminus \tilde{D}) = 0$  (see Remark 20), and hence it suffices to show that the ML degree equals

$$e_{top}(X \setminus D) = e_{BM}(\tilde{X} \setminus \tilde{D}) = e_{top}(\tilde{X} \setminus \tilde{D}).$$

In other words, we may now simply erase the tilde and consider the case when  $X$  is smooth and  $f$  defines a proper morphism  $X \setminus D \rightarrow \mathbb{C}^*$ .

Let  $\mathcal{C}$  denote the set of critical points of  $f$  on  $X$ . By hypothesis (c), this set is finite and the ML degree equals its cardinality counting multiplicities:

$$\mu = \sum_{p \in \mathcal{C}} \mu_p.$$

The multiplicity  $\mu_p$  of a critical point  $p$  of  $f$  is known as the *Milnor number* at  $p$  of the hypersurface  $F_p = \{x \in X : f(x) = f(p)\}$ . Milnor [18] showed that this algebraic invariant of a singularity has the following topological interpretation. Consider a coordinate chart around the point  $p$  and intersect the fiber  $F_\epsilon := \{x \mid f(x) = f(p) + \epsilon\}$  with a ball of radius  $\delta$  around  $p$ . For  $\epsilon \ll \delta$  this intersection is the *Milnor fiber*. Milnor [18] showed that the Milnor fiber is homotopy equivalent to a bouquet of  $\mu_p$  spheres of dimension  $d - 1$ .

Each singular fiber is obtained (cf. [3, 18]) from a smooth fiber by replacing the Milnor fiber by a contractible set. The Borel-Moore exact sequence (21) implies that the Euler number of a singular fiber  $F'$  is obtained from the Euler number of a smooth fiber  $F$  by adding  $-(-1)^{d-1} \sum_{p \in F'} \mu_p$ .

Then the Euler number of the union of the singular fibers equals  $|\mathcal{C}|$  times the Euler number of a smooth fiber  $F$  plus the correction  $-(-1)^{d-1} \mu$ . Applying Künneth's formula to the fiber bundle defined by  $f$  on  $X \setminus D$  minus the union of the singular fibers, and then applying the additivity formula (22), we conclude that  $e_{BM}(X \setminus D) = e_{top}(X \setminus D) = (-1)^d \mu$  as desired.  $\square$

**Example 21.** For another illustration consider Example 18 with  $X = \mathbb{P}_{\mathbb{C}}^2$ . The generic ML degree was 9 but it decreased by 4 when  $u = (u_1, u_2, u_3, u_4)$  is a general solution of  $u_1 + u_2 + 2u_3 + 2u_4 = 0$ . This is consistent with Theorem 19, because for such  $u$  the divisor  $D$  loses the component at infinity. The difference is a projective line minus 6 points, which has Euler number  $-4$ . Consider our hypotheses when  $\tilde{X}$  is the blow-up of  $X$  at the origin. If  $u_1 + u_2 + u_3 + u_4 \neq 0$  then the exceptional curve is part of  $\tilde{D}$  and Theorem 19 is valid. On other hand, if  $u_1 + u_2 + u_3 + u_4 = 0$  then it maps to  $\mathbb{P}^1$  under a rational map of degree  $\geq 2$ , so hypothesis (a) does not hold. The philosophy of this example is that, even if the divisor  $D$  is locally biholomorphic to an arrangement of hyperplanes, genericity of the exponents  $u_i$  may be necessary for the topological formula of Theorem 19 to hold.  $\square$

Theorems 1, 7 and 15 offer combinatorial formulae for the ML degree and hence (using Theorem 19) for the Euler number of the complement of

an arrangement of generic hypersurfaces  $\{f_i = 0\}$  in  $X = \mathbb{P}_{\mathbb{C}}^d$  or in a toric manifold. In each of these theorems, the combinatorial number becomes an upper bound for the ML degree when the coefficients of the  $f_i$ 's are special. This semi-continuity principle will be explained by the following general topological result, in which also the underlying manifold is allowed to vary.

**Theorem 22.** *Assume we are given a one-parameter smooth proper family  $X_t$  of complex manifolds over the unit disk  $B := \{t \in \mathbb{C} : |t| < 1\}$ , and a family of rational functions  $f_t$  on  $X_t$ , such that*

1. *for  $t \neq 0$  the divisor  $D(t)$  defined by  $f_t$  has GNC, and*
2. *for  $t = 0$  the divisor  $D(0)$  defined by  $f_0$  has the same homology class as  $D(t)$  for the natural differentiable trivialization of the family  $X_t$ .*

*Then the ML degree of  $f_0$  is less than or equal to the ML degree of  $f_t$ .*

In order to understand the second hypothesis, let us recall Ehresmann's Theorem [11]: any proper submersion  $\phi : \mathcal{X} \rightarrow B$  of differentiable manifolds is a differentiable fiber bundle, i.e., if  $U$  is a sufficiently small open set in  $B$ , there is a local diffeomorphism between  $\phi^{-1}(U)$  and  $U \times F$ , for a fiber  $F$ , and this diffeomorphism is compatible with the two projections to  $U$ .

*Sketch of Proof.* Let  $\mathcal{X} \rightarrow B$  be the total space of the family  $\{X_t\}_{t \in B}$ . We consider the function  $\phi : \mathcal{X} \rightarrow B \times \mathbb{P}^1$ , given by  $\phi(x) = (t(x), f(x))$ .

Consider the locus  $\Xi \subset \mathcal{X}$  given by the vanishing of the vertical differential of  $\phi$ : this is the local complete intersection defined by the  $d$  partials  $\partial f / \partial x_i = 0$ , where  $x_1, \dots, x_d$  are local coordinates on the fibers provided by the Implicit Function Theorem. At each critical point  $p$  of  $f_0$ , the locus  $\Xi$  has dimension 1, and thus, in a neighborhood of  $p$ , the morphism  $\Xi \rightarrow B$  is finite, whence its degree is locally constant. This establishes the desired semi-continuity.  $\square$

In general, it might be difficult to show that a given rational function  $f_0$  has a perturbation as above, or we might want to calculate the ML degree with more algebraic precision. The results in the next section may help.

## 7 Logarithmic vector fields

In this section we will show that the formula for the ML degree given by the logarithmic Chern number (Theorem 4) holds in greater generality. Returning to the setting of Chapter 2, we consider logarithmic vector fields along  $D$ . Again, our definition differs slightly from the one given by Saito [23].

**Definition 23.** If  $D$  is a reduced divisor on a factorial variety  $X$ , the *sheaf of logarithmic vector fields*  $\Theta_X(-\log D)$  is the dual  $\mathcal{H}om_{\mathcal{O}_X}(\Omega_X^1(\log D), \mathcal{O}_X)$ .

Recall that the *tangent sheaf*  $\Theta_X$  is  $\mathcal{H}om_{\mathcal{O}_X}(\Omega_X^1, \mathcal{O}_X)$ , the dual of the 1-forms on  $X$ . The inclusion of  $\Omega_X^1$  into  $\Omega_X^1(\log D)$ , studied in Lemma 2, dualizes to an inclusion of the logarithmic vector fields into the tangent sheaf.

**Proposition 24.** *We have the following exact sequence of sheaves on  $X$ :*

$$0 \rightarrow \Theta_X(-\log D) \rightarrow \Theta_X \rightarrow \bigoplus_{i=1}^r \mathcal{O}_{D_i}(D_i) \rightarrow \mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1(\log D), \mathcal{O}_X).$$

*If  $X$  is smooth, then the rightmost homomorphism is onto, and the total Chern class of the sheaf of logarithmic vector fields equals*

$$c_{tot}(\Theta_X(-\log D)) = \frac{c_{tot}(\Theta_X) \cdot c_{tot}(\mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1(\log D), \mathcal{O}_X))}{\prod_{i=1}^r (1 + zD_i)}. \quad (23)$$

*Proof.* Dualizing the sequence  $0 \rightarrow \mathcal{O}_X(-D_i) \rightarrow \mathcal{O}_X \rightarrow \mathcal{O}_{D_i} \rightarrow 0$ , we get

$$\mathcal{H}om_{\mathcal{O}_X}(\mathcal{O}_{D_i}, \mathcal{O}_X) = 0 \quad \text{and} \quad \mathcal{E}xt_{\mathcal{O}_X}^1(\mathcal{O}_{D_i}, \mathcal{O}_X) \cong \mathcal{O}_{D_i}(D_i).$$

Hence the  $\mathcal{E}xt_{\mathcal{O}_X}$ -sequence gotten by dualizing the sequence (7) has the form

$$0 \rightarrow 0 \rightarrow \Theta_X(-\log D) \rightarrow \Theta_X \rightarrow \bigoplus_{i=1}^r \mathcal{O}_{D_i}(D_i) \rightarrow \mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1(\log D), \mathcal{O}_X) \rightarrow \mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1, \mathcal{O}_X) \rightarrow \dots$$

This is the first statement of Proposition 24. If  $X$  is smooth then the cotangent sheaf  $\Omega_X^1$  is free, and we have  $\mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1, \mathcal{O}_X) = 0$ . The formula (23) follows from the multiplicativity of the total Chern class.  $\square$

**Remark 25.** The two leftmost maps in the exact sequence of Proposition 24 characterize the logarithmic vector fields on  $X$  along  $D$  as those vector fields  $\xi \in \Theta_X$  which satisfy  $\xi(F_i) \equiv 0 \pmod{F_i}$  for all  $i$ . In other words, for each  $i = 1, \dots, h$ , the vector field  $\xi = \sum_{j=1}^d \xi_j \partial/\partial x_j$  has the property that there exist functions  $\psi_i$  such that  $\xi(F_i) := \sum_{j=1}^d \xi_j \partial F_i / \partial x_j = \psi_i F_i$ .

**Remark 26.** An interesting case will be the one where  $X$  is smooth and the sheaf  $\mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1(\log D), \mathcal{O}_X)$  has zero-dimensional support and length  $p$ . Here the top Chern class of  $\Theta_X(-\log D)$  and the top Chern class of  $\Omega_X^1(\log D)$  differ by  $(-1)^{d-1}p$ , and the count in part 3 of Theorem 4 changes accordingly.

As in the proof of Theorem 4, set  $\sigma := d\log(f)$  and let  $Z_\sigma$  be the subscheme of  $X$  defined by the vanishing of  $\sigma$ . The restriction of  $Z_\sigma$  to the open set  $V = X \setminus D$  is the critical locus of  $f$ . The reason why the sheaf  $\Theta_X(-\log D)$  is important is that it enters directly into the algebraic description of  $Z_\sigma$ .

Indeed, the section  $\sigma$  of  $\Omega_X^1(\log D)$  corresponds to an exact sequence

$$0 \rightarrow \mathcal{O}_X \rightarrow \Omega_X^1(\log D) \rightarrow \mathcal{E} \rightarrow 0,$$

where  $\mathcal{O}_{Z_\sigma}$  is the kernel of  $\mathcal{E}xt_{\mathcal{O}_X}^1(\mathcal{E}, \mathcal{O}_X) \rightarrow \mathcal{E}xt_{\mathcal{O}_X}^1(\Omega_X^1(\log D), \mathcal{O}_X)$ . Hence the  $\mathcal{E}xt_{\mathcal{O}_X}$ -sequence gotten by dualizing the previous sequence has the form

$$0 \rightarrow \mathcal{H}om_{\mathcal{O}_X}(\mathcal{E}, \mathcal{O}_X) \rightarrow \Theta_X(-\log D) \rightarrow \mathcal{O}_X \rightarrow \mathcal{O}_{Z_\sigma} \rightarrow 0. \quad (24)$$

This leads to the following more refined formula for the ML degree.

**Theorem 27.** *Assume that  $X$  is smooth,  $D$  meets every curve, the sheaf  $\Theta_X(-\log D)$  is locally free, and  $Z_\sigma$  does not intersect the divisor  $D$ . Then the number of critical points of  $f$  on  $V = X \setminus D$  equals  $(-1)^d c_d(\Theta_X(-\log D))$ .*

*Proof.* Using (24), we can view  $Z_\sigma$  as the locus of zeros of the locally free sheaf dual to  $\Theta_X(-\log D)$ , whose top Chern class is  $(-1)^d c_d(\Theta_X(-\log D))$ .  $\square$

**Remark 28.** The sequence (24) gives the following description of the equations defining the critical points. The ideal sheaf of the subscheme  $Z_\sigma$  is generated by the functions  $\sum_{i=1}^r u_i \psi_i$ , where  $\xi$  varies among the logarithmic vector fields and  $(\psi_1, \dots, \psi_r)$  is derived from  $\xi$  as in Remark 25. This holds because the section  $\sigma = d\log(f)$  factors through the homomorphism  $\mathcal{O}_X^r \rightarrow \Omega_X^1(\log D)$ , thus the homomorphism dual to the section  $\sigma$  also factors through the dual homomorphism  $\Theta_X(-\log D) \rightarrow \mathcal{O}_X^r$ ,  $\xi \mapsto (\psi_1, \dots, \psi_r)$ .

We next present two examples which illustrate the hypotheses of Theorem 27. Here  $X$  is a smooth surface (i.e.  $d = 2$ ) with local coordinates  $x$  and  $y$ .

**Example 29.** Let  $h = 2$  with  $F_1 = x$  and  $F_2 = x^v - y^u$ . A vector field  $\xi = a(x, y) \cdot \partial/\partial x + b(x, y) \cdot \partial/\partial y$  is logarithmic if and only if  $a = x\psi_1$  and  $b = (v/u) y \psi_1 + \lambda(x^v - y^u)$ , for some function  $\lambda$ . Observe that  $\psi_2 = v\psi_1 - y^{u-1} \lambda$ . We conclude that the sheaf  $\Theta_X(-\log D)$  is locally free of rank 2.

The origin does not belong to the subscheme  $Z_\sigma$  if there is a function  $u_1\psi_1 + u_2\psi_2 = (u_1 + v \cdot u_2)\psi_1 - u_2 y^{u-1} \lambda$ , for some  $\psi_1$  and  $\lambda$ , which does not vanish at the origin. This holds if and only if  $u_1 + v \cdot u_2 \neq 0$ .  $\square$

**Example 30.** Let  $h = 3$  with  $F_1 = x, F_2 = y$  and  $F_3 = x - y$ . Logarithmic vector fields  $\xi$  have the form  $x\psi_1 \cdot \partial/\partial x + y\psi_2 \cdot \partial/\partial y$  with  $x\psi_1 - y\psi_2$  divisible by  $x - y$ . This implies  $\psi_1 = \lambda + y \cdot \mu, \psi_2 = \lambda + x \cdot \mu$ , and  $\psi_3 = \lambda$ , for any functions  $\lambda, \mu$ . Thus  $\Theta_X(-\log D)$  is locally free of rank two. The origin does not belong to the subscheme  $Z_\sigma$  if and only if  $u_1 + u_2 + u_3 \neq 0$ .

**Acknowledgements:** This project grew out of discussions with Lior Pachter at the December 2003 workshop on *Computational Algebraic Statistics* at the American Institute for Mathematics (AIM) at Palo Alto. Many thanks to both Lior and AIM for getting us started, and many thanks to MSRI Berkeley and the organizers of the Spring 2004 MSRI program *Topology of Real Algebraic Varieties*, which gave the four of us the opportunity to collaborate. We also thank Oleg Viro for letting us include Theorem 12. Amit Khetan was supported by an NSF postdoctoral fellowship (DMS-0303292). Bernd Sturmfels was supported by the Hewlett Packard Visiting Research Professorship 2003/2004 at MSRI Berkeley and in part by the NSF (DMS-0200729).

## References

- [1] A. Andreotti and T. Frankel. The Lefschetz theorem on hyperplane sections, *Annals of Math.* (2) **69** (1959), 713–717.
- [2] A. Andreotti and T. Frankel. The second Lefschetz theorem on hyperplane sections, *Global Analysis* (Papers in Honor of K. Kodaira), Univ. Tokyo Press, Tokyo (1969), 1–20.
- [3] W. Barth, C. Peters and A. Van de Ven. *Compact Complex Surfaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Folge 3, Band 4, Springer Verlag, 1984.
- [4] A. Borel and J.C. Moore. Homology theory for locally compact spaces, *Michigan Math. J.* **7** (1960) 137–159.
- [5] M. Bout and D. Richards. Counting and locating the solutions of polynomial systems of maximum likelihood equations, Manuscript, 2004.
- [6] G.E. Bredon. *Sheaf theory*, McGraw-Hill Book Co., New York-Toronto, Ont.-London 1967 xi+272 pp. Second edition: Graduate Texts in Mathematics, 170, Springer-Verlag, New York, 1997.

- [7] B. Chor, A. Khetan and S. Snir. Maximum likelihood on four taxa phylogenetic trees: analytic solutions, *The 7th Annual Conference on Research in Computational Molecular Biology – RECOMB 2003*, Berlin, April 2003, pp. 76–83.
- [8] D. Cox. The homogeneous coordinate ring of a toric variety. *Journal of Algebraic Geometry*, 4(1):17–50, 1995.
- [9] P. Déglise. Équations différentielles à points singuliers réguliers, Lecture Notes in Mathematics **163**, Springer Verlag, 1970.
- [10] M. Drton and T.S. Richardson. Multimodality of the likelihood in the bivariate seemingly unrelated regressions model, *Biometrika* **91** (2004) 383–392.
- [11] C. Ehresmann. Sur les espaces fibrés différentiables, *C. R. Acad. Sci. Paris* **224** (1947) 1611–1612.
- [12] W. Fulton. *Introduction to Toric Varieties*. Annals of Mathematics Studies, Vol. 131 Princeton University Press, Princeton, NJ, 1993.
- [13] L. Garcia, M. Stillman and B. Sturmfels. Algebraic geometry of Bayesian networks, [math.AG/0301255](https://arxiv.org/abs/math/0301255), *J. Symbolic Comput.*, to appear.
- [14] D. Geiger, C. Meek and B. Sturmfels. The toric algebra of graphical models, Preprint, Microsoft Research, February 2002, posted at <http://research.microsoft.com/research/pubs/>.
- [15] A. Harnack. Über die Vieltheiligkeit der ebenen algebraischen Curven, *Mathematische Annalen* **10** (1876) 189–198.
- [16] H. Hauser. The Hironaka theorem on resolution of singularities (or: A proof we always wanted to understand). *Bull. Amer. Math. Soc.* **40** (2003), no. 3, 323–403.
- [17] J. Mather, *Notes on topological stability*, Harvard University Math. Notes (1970).
- [18] J. Milnor. *Singular points of complex hypersurfaces*, Annals of Mathematics, Studies, No. 61, Princeton University Press, 1968.

- [19] D. Mumford. *Algebraic Geometry. I*, Grundlehren der mathematischen Wissenschaften, 221. Springer-Verlag, Berlin-New York, 1981.
- [20] P. Orlik and H. Terao. *Arrangements of Hyperplanes*, Grundlehren der mathem. Wissenschaften, 300, Springer Verlag, Berlin-New York, 1992.
- [21] L. Pachter and B. Sturmfels. Tropical geometry of statistical models, q-bio.QM/0311009, *Proc. Natl. Acad. Sci. USA*, to appear.
- [22] G. Pistone, E. Riccomagno, and H.P. Wynn. *Algebraic Statistics: Computational Commutative Algebra in Statistics*, Chapman and Hall, Boca Raton, Florida, 2001.
- [23] K. Saito. Theory of logarithmic differential forms and logarithmic vector fields. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **27** (1980), no. 2, 265–291.
- [24] H. Terao. Algebras generated by reciprocals of linear forms, *Journal of Algebra* **250** (2002), no. 2, 549–558.
- [25] A. Varchenko. Critical points of the product of powers of linear functions and families of bases of singular vectors, *Compositio Math.* **97** (1995) 385–401.
- [26] T. Zaslavsky. Facing up to arrangements: face count formulas for partitions of space by hyperplanes, *Memoirs AMS* **1**, no. 154, 1975.

**Authors' addresses:**

Fabrizio Catanese, Lehrstuhl Mathematik VIII, Universität Bayreuth, NWII, D-95440 Bayreuth, Germany, [Fabrizio.Catanese@uni-bayreuth.de](mailto:Fabrizio.Catanese@uni-bayreuth.de)

Serkan Hoşten, Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA, [serkan@math.sfsu.edu](mailto:serkan@math.sfsu.edu)

Amit Khetan, Department of Mathematics, University of Massachusetts, Amherst, MA 01002, USA, [khetan@math.umass.edu](mailto:khetan@math.umass.edu)

Bernd Sturmfels, Department of Mathematics, University of California, Berkeley, CA 94720, USA, [bernd@math.berkeley.edu](mailto:bernd@math.berkeley.edu)