

The Maximum Likelihood Degree of Mixtures of Independence Models*

Jose Israel Rodriguez[†] and Botong Wang[‡]

Abstract. The maximum likelihood degree (ML degree) measures the algebraic complexity of a fundamental optimization problem in statistics: maximum likelihood estimation. In this problem, one maximizes the likelihood function over an algebraic statistical model. The ML degree of a model is an upper bound to the number of local extrema of the likelihood function and can be expressed as a weighted sum of Euler characteristics. The independence model (i.e., rank 1 matrices over the probability simplex) is well known to have an ML degree of one, meaning there is a unique local maximum of the likelihood function. However, for mixtures of independence models (i.e., rank 2 matrices over the probability simplex), it was an open question as to how the ML degree behaved. In this paper, we use Euler characteristics to prove an outstanding conjecture by Hauenstein, the first author, and Sturmfels; we give recursions and closed form expressions for the ML degree of mixtures of independence models.

Key words. algebraic statistics, maximum likelihood estimation, mixture models, Euler characteristics

AMS subject classifications. 14M12, 62F10, 90C26

DOI. 10.1137/16M1088843

1. Introduction. Maximum likelihood estimation is a fundamental computational task in statistics. A typical problem encountered in its applications is the occurrence of multiple local maxima. To be certain that a global maximum of the likelihood function has been achieved, one locates all solutions to a system of polynomial equations called *likelihood equations*; every local maxima is a solution to these equations. The number of solutions to these equations is called the maximum likelihood degree (ML degree) of the model. This degree was introduced in [4, 15] and gives a measure of the complexity to the global optimization problem, as it bounds the number of local maxima.

The ML degree has been studied in many contexts. Some of these include Gaussian graphical models [26], variance component models [12], and missing data [16]. In this article, we work in the context of discrete random variables (for a recent survey in this context, see [19]). In our main results, we provide closed form expressions for ML degrees of mixtures of independence models, which are sets of joint probability distributions for two random variables. This resolves Conjecture 4.1 in [14].

1.1. Algebraic statistics preliminaries. We consider a model for two discrete random variables, having m and n states respectively. A joint probability distribution for two such

*Received by the editors August 9, 2016; accepted for publication (in revised form) June 7, 2017; published electronically August 17, 2017.

<http://www.siam.org/journals/siaga/1/M108884.html>

[†]Department of Statistics, The University of Chicago, Chicago, IL 60637 (joisro@uchicago.edu, <http://home.uchicago.edu/~joisro>).

[‡]Department of Mathematics, University of Wisconsin–Madison, Madison, WI 53706 (wang@math.wisc.edu, <http://www.math.wisc.edu/~wang/>).

random variables is written as an $m \times n$ -matrix,

$$(1) \quad P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}.$$

The (i, j) th entry p_{ij} represents the probability that the first variable is in state i and the second variable is in state j . By a statistical model, we mean a subset \mathcal{M} of the probability simplex Δ_{mn-1} of all such P . The $(d - 1)$ -dimensional probability simplex is defined as $\Delta_{d-1} := \{(p_1, p_2, \dots, p_d) \in \mathbb{R}_{\geq 0}^d \text{ such that } \sum p_i = 1\}$.

If independent and identically distributed (i.i.d.) samples are drawn from some distribution P , then the *data* is summarized by the following matrix (2). The entries of u are nonnegative integers, where u_{ij} is the number of samples drawn with state (i, j) :

$$(2) \quad u = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{bmatrix}.$$

The likelihood function corresponding to the data matrix u is given, up to a constant, by

$$(3) \quad \ell_u(p) := p_{11}^{u_{11}} p_{12}^{u_{12}} \cdots p_{mn}^{u_{mn}}.$$

Maximum likelihood estimation is an optimization problem for the likelihood function. This problem consists of determining, for fixed u , the argmax of $\ell_u(p)$ on a statistical model \mathcal{M} . The optimal solution is called the maximum likelihood estimate (mle) and is used to measure the true probability distribution. For the models we consider, the mle is a solution to the likelihood equations. In other words, by solving the likelihood equations, we solve the ML estimation problem. Since the ML degree is the number of solutions to the likelihood equations, it gives a measure on the difficulty of the problem.

The model \mathcal{M}_{mn} in Δ_{mn-1} is said to be the *mixture of independence models* and is defined to be the image of the following map:

$$(4) \quad \begin{array}{ccc} (\Delta_{m-1} \times \Delta_{n-1}) \times (\Delta_{m-1} \times \Delta_{n-1}) \times \Delta_1 & \rightarrow & \mathcal{M}_{mn} \\ (R_1, B_1, R_2, B_2, C) & \mapsto & c_1 R_1 B_1^T + c_2 R_2 B_2^T \end{array},$$

where $R_i, B_j, C := [c_1, c_2]^T$ are $m \times 1$, $n \times 1$, and 2×1 matrices, respectively, with positive entries that sum to one. The Zariski closure of \mathcal{M}_{mn} yields the variety of rank at most 2 matrices over the complex numbers. We will determine the ML degree of the models \mathcal{M}_{mn} by studying the topology of the Zariski closure. Prior to our work, the ML degrees of these models were only known for small values of m and n . In [14], the following table of ML degrees of \mathcal{M}_{mn} was computed:

$n =$	3	4	5	6	7
$m = 3 :$	10	26	58	122	250
$m = 4 :$	26	191	843	3119	
$m = 5 :$	58	843	6776		

Mixtures of independence models appear in many places in science, statistics, and mathematics. In computational biology, the case where (m, n) equals $(4, 4)$ is discussed in Example 1.3 of [22], and the data u consists of a pair of DNA sequences of length $(u_{11} + u_{12} + \cdots + u_{mn})$. The m and n equal four because DNA molecules are composed of four nucleotides. Another interesting case for computational biology is when $(m, n) = (20, 20)$ because there are 20 essential amino acids (Remark 4.3).

Our first main result, Theorem 3.12, proves a formula for the first row of the table:

$$\text{MLdegree } \mathcal{M}_{3n} = 2^{n+1} - 6 \text{ for } n \geq 3.$$

Our second main result is a recursive expression for the ML degree of mixtures of independence models. As a consequence, we are able to calculate a closed form expression for each row of the above table of ML degrees [Corollary 4.2].

Our techniques relate ML degrees to Euler characteristics. Previously, Huh showed that the ML degree of a smooth algebraic statistical model \mathcal{M} with Zariski closure X equals the signed topological Euler characteristic of an open subvariety X° , where X° is the set of points of X with nonzero coordinates and coordinate sums [18]. More recent work of Budur and the second author shows that the ML degree of a singular model is a stratified topological invariant. In [2], they show that the Euler characteristic of X° is a sum of ML degrees weighted by Euler obstructions. These Euler obstructions, in a sense, measure the multiplicity of the singular locus.

We conclude this introduction with examples to set notation and definitions.

1.2. Defining the maximum likelihood degree. We will use two notions of ML degree. The first notion is from a computational algebraic geometry perspective, where we define the ML degree for a projective variety. When this projective variety is contained in a hyperplane, the ML degree has an interpretation related to statistics. The second notion is from a topological perspective, where we define the ML degree for a very affine variety, a subvariety of an algebraic torus $(\mathbb{C}^*)^n$.

To \mathbb{P}^{n+1} we associate the coordinates p_0, p_1, \dots, p_n and p_s (where s stands for sum). Consider the *distinguished hyperplane* in \mathbb{P}^{n+1} defined by $p_0 + \cdots + p_n - p_s = 0$ (p_s is the sum of the other coordinates).

Let X be a generically reduced variety contained in the distinguished hyperplane of \mathbb{P}^{n+1} not contained in any coordinate hyperplane. We will be interested in the critical points of the *likelihood function*

$$\ell_u(p) := p_0^{u_0} p_1^{u_1} \cdots p_n^{u_n} p_s^{u_s},$$

where $u_s := -u_0 - \cdots - u_n$ and $u_0, \dots, u_n \in \mathbb{C}$. The likelihood function has the nice property that, up to scaling, its gradient is a rational function $\nabla \ell_u(p) := [\frac{u_0}{p_0} : \frac{u_1}{p_1} : \cdots : \frac{u_n}{p_n} : \frac{u_s}{p_s}]$.

Definition 1.1. *Let u be fixed. A point $p \in X$ is said to be a critical point of the likelihood function on X if p is a regular point of X , each coordinate of p is nonzero, and the gradient $\nabla \ell_u(p)$ at p is orthogonal to the tangent space of X at p .*

Example 1.2. Let X of \mathbb{P}^4 be defined by $p_0 + p_1 + p_2 + p_3 - p_s$ and $p_0 p_3 - p_1 p_2$. For $[u_0 : u_1 : u_2 : u_3 : u_s] = [2 : 8 : 5 : 10 : -25]$ there is a unique critical point for $\ell_u(p)$ on X . This point is $[p_0 : p_1 : p_2 : p_3 : p_s] = [14 : 36 : 21 : 54 : -125]$. Whenever the

coordinates are nonzero, there is a unique critical point $[(u_0 + u_1)(u_0 + u_2) : (u_0 + u_1)(u_1 + u_3) : (u_2 + u_3)(u_0 + u_2) : (u_2 + u_3)(u_1 + u_3) : -(u_0 + u_1 + u_2 + u_3)^2]$.

Definition 1.3. *The ML degree of X is defined as the number of critical points of the likelihood function on X for general u_0, \dots, u_n . The ML degree of X is denoted by $\text{MLdeg}(X)$.*

We say u^* in \mathbb{C}^{n+1} is general whenever there exists a dense Zariski open set \mathcal{U} for which the number of critical points of $\ell_u(p)$ is constant and $u^* \in \mathcal{U}$. In Example 1.2, the Zariski open set \mathcal{U} is the complement of the variety defined by $(u_0 + u_1)(u_0 + u_2)(u_1 + u_3)(u_2 + u_3)u_s = 0$. Determining this Zariski open set explicitly is often quite difficult, but it is not necessary when using reliable probabilistic algorithms to compute ML degrees as done in [15], for example. However, with our results, we compute ML degrees using Euler characteristics and topological arguments.

1.3. Using Euler characteristics. First, we recall that the Euler characteristic of a topological space X is defined as

$$\chi(X) = \sum_{i \geq 0} (-1)^i \dim H^i(X, \mathbb{Q}),$$

where the H^i are the singular cohomology groups.

In the definition of ML degree of a projective variety X , a critical point $p \in X$ must have nonzero coordinates. This means all critical points of the likelihood function are contained in the underlying very affine variety

$$X^\circ := X \setminus \{\text{coordinate hyperplanes}\}.$$

In fact, the ML degree is directly related to the Euler characteristic of smooth X° .

Theorem 1.4 (see [18]). *Suppose X is a smooth projective variety of \mathbb{P}^{n+1} . Then*

$$(5) \quad \chi(X^\circ) = (-1)^{\dim X^\circ} \text{MLdeg } X.$$

The next example will show how to determine the signed Euler characteristic of a very affine variety Y . Recall that the Euler characteristic is a homotopy invariant and satisfies the following properties. The Euler characteristic is *additive* for algebraic varieties. More precisely, $\chi(X) = \chi(Z) + \chi(X \setminus Z)$, where Z is a closed subvariety of X (see [9, section 4.5], for example). The *product property* says $\chi(M \times N) = \chi(M) \cdot \chi(N)$. More generally, the *fibration property* says that if $E \rightarrow B$ is a fibration with fiber F , then $\chi(E) = \chi(F) \cdot \chi(B)$ (see [25, section 9.3], for example).

Example 1.5. Consider X from Example 1.2. The variety X has the following parameterization:

$$\begin{aligned} \mathbb{P}^1 \times \mathbb{P}^1 &\rightarrow X, \\ [x_0 : x_1] \times [y_0 : y_1] &\mapsto [x_0y_0 : x_0y_1 : x_1y_0 : x_1y_1 : x_0y_0 + x_0y_1 + x_1y_0 + x_1y_1]. \end{aligned}$$

Let X° be the underlying very affine variety of X , and consider

$$\mathcal{O} := \mathbb{P}^1 \setminus \{[0 : 1], [1 : 0], [1 : -1]\},$$

i.e., the projective line with three points removed. The very affine variety X^o has a parameterization

$$\begin{aligned} \mathcal{O} \times \mathcal{O} &\rightarrow X^o, \\ [x_0 : x_1] \times [y_0 : y_1] &\mapsto [x_0y_0 : x_0y_1 : x_1y_0 : x_1y_1, (x_0 + x_1)(y_0 + y_1)]. \end{aligned}$$

Since $\chi(\mathbb{P}^1) = 2$, after removing three points, we have $\chi(\mathcal{O}) = -1$. By the product property, $\chi(\mathcal{O} \times \mathcal{O}) = 1$, and hence $\chi(X^o) = 1$. Because X^o is smooth, by Huh's result, we conclude that the ML degree of X^o is 1 as well.

We call X the variety of 2×2 matrices with rank 1, and restricting to the probability simplex yields the independence model. This example is generalized by considering the map $\mathbb{P}^{m-1} \times \mathbb{P}^{n-1} \rightarrow X$ given by $([x_0 : \cdots : x_{m-1}], [y_0 : \cdots : y_{n-1}]) \rightarrow [x_0y_0 : \cdots : x_{m-1}y_{n-1} : \sum_{i,j} x_iy_j]$. In this case, we can identify X with the variety of $m \times n$ rank 1 matrices, and a similar computation shows that the ML degree is 1 in these cases as well (see Example 1 of [14]).

2. Whitney stratification, Gaussian degree, and Euler obstruction. As we have just seen, for a smooth very affine variety, the ML degree is equal to the Euler characteristic up to a sign. This is not always the case when the very affine variety is not smooth. When the variety is singular, the ML degree is related to the Euler characteristic in a subtle way involving Whitney stratifications. This subtlety is explained in Corollary 2.9 by considering a weighted sum of the ML degrees of each strata. Here, we give a brief introduction to the topological notions of Whitney stratification and Euler obstruction.

2.1. Whitney stratification. Many differential geometric notions do not behave well when a variety has singularities, for instance, tangent bundles and Poincaré duality. This situation is addressed by stratifying the singular variety into finitely many pieces such that along each piece the variety is close to a smooth variety. A naive way to stratify a variety X is by taking the regular locus X_{reg} as the first stratum, then taking the regular part of the singular locus of X , i.e., $(X_{\text{sing}})_{\text{reg}}$, and repeating this procedure. This naive stratification does not always reflect the singular behavior of a variety as seen in the Whitney umbrella. Whitney introduced conditions (now called *Whitney regular*) on a stratification, where many differential geometric results can be generalized to singular varieties. A *Whitney stratification* satisfies the following technical condition of Whitney regular. The definition below is from [11, page 37]. See also [17, E3.7].

Definition 2.1. *Let X be an analytic subvariety of a complex manifold M . Let $X = \bigsqcup_{j \in A} S_j$ be a stratification of X into finitely many locally closed submanifolds of M . This stratification is called a Whitney stratification if all the pairs (S_α, S_β) with $S_\alpha \subset \bar{S}_\beta$ are Whitney regular, which means the following:*

Suppose $x_i \subset S_\beta$ is a sequence of points converging to some point $y \in S_\alpha$. Suppose $y_i \subset S_\beta$ also converges to y , and suppose that (with respect to some local coordinate system on M) the secant lines $l_i = \overline{x_i y_i}$ converge to some limiting line l , and the tangent planes $T_{x_i} S_\beta$ converge to some limiting plane τ . Then $l \subset \tau$.

Example 2.2. Here are some instances of Whitney stratifications.

1. [27, Theorem 19.2] Every complex algebraic and analytic variety admits a Whitney stratification.
2. If X is a smooth variety, then the trivial stratification X itself is a Whitney stratification.
3. If X has isolated singularities at P_1, \dots, P_l , then the stratification $\{P_1\} \sqcup \dots \sqcup \{P_l\} \sqcup X \setminus \{P_1, \dots, P_l\}$ is a Whitney stratification.
4. Since Whitney regularity is a local condition, if $\bigsqcup_{j \in A} S_j$ is a Whitney stratification of an algebraic variety X , and if U is an open subset of X , then $\bigsqcup_{j \in A} (S_j \cap U)$ is a Whitney stratification of U .
5. If X is a smooth algebraic variety and $Z \subset X$ is a closed smooth subvariety, then the pair $\{Z, X \setminus Z\}$ is also a Whitney stratification of X .
6. If $\bigsqcup_{i \in I} S_i$ is a Whitney stratification of X and $\bigsqcup_{j \in J} T_j$ is a Whitney stratification of Y , then $\bigsqcup_{i \in I, j \in J} S_i \times T_j$ is a Whitney stratification of $X \times Y$.

Proposition 2.3. *Let X be an algebraic variety. Suppose an algebraic group \mathcal{G} acts on X with only finitely many orbits O_1, \dots, O_l . Then $O_1 \sqcup \dots \sqcup O_l$ form a Whitney stratification of X .*

Proof. First, we need to show that each O_i is a locally closed smooth subvariety of X . Let $f : \mathcal{G} \times X \rightarrow X$ be the group action map. Then each orbit is of the form $f(\mathcal{G} \times \{point\})$ and hence is a constructible subset of X . On the other hand, since \mathcal{G} acts transitively on each O_i , as constructible subsets of X , O_i are isomorphic to smooth varieties. This implies that if $\dim O_i = \dim X$, then O_i is open in X . Therefore, $X \setminus \bigcup_{\dim O_i = \dim X} O_i$ is a closed (possibly reducible) subvariety of X . Now, \mathcal{G} acts on $X \setminus \bigcup_{\dim O_i = \dim X} O_i$ with finite orbits. Thus, we can use induction to conclude that each O_i is a locally closed smooth subvariety of X .

Without loss of generality, we assume that $O_1 \subset \bar{O}_2$, and we need to show that the pair O_1, O_2 is Whitney regular. Moreover, by replacing X by \bar{O}_2 , we can assume that O_2 is open in X . By [27, Lemma 19.3], there exists an open subvariety U of O_1 such that the pair U, O_2 is Whitney regular. Given any $\tau \in \mathcal{G}$, the action $\tau : \bar{O}_2 \rightarrow \bar{O}_2$ is an algebraic map, which preserves O_2 . Thus, the pair $\tau(U), O_2$ is also Whitney regular. Since \mathcal{G} acts transitively on O_1 ,

$$\bigcup_{\tau \in \mathcal{G}} \tau(U) = O_1.$$

Therefore, the pair O_1, O_2 is Whitney regular. ■

Consider $\mathbb{P}^{mn-1} = \{[a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}\}$ as the projective space of $m \times n$ matrices. The left $Gl(m, \mathbb{C})$ action and the right $Gl(n, \mathbb{C})$ action on \mathbb{P}^{mn-1} both preserve the rank of the matrices. Moreover, the orbits of the total action of $Gl(m, \mathbb{C}) \times Gl(n, \mathbb{C})$ are the matrices with fixed rank. Therefore, the stratification by rank gives a Whitney stratification of the subvariety $X_r := \{[a_{ij}] \mid \text{rank}([a_{ij}]) \leq r\}$. In particular, we have the following corollary.

Corollary 2.4. *Define $X := \{[a_{ij}] \mid \text{rank}([a_{ij}]) \leq 2\}$ and $Z := \{[a_{ij}] \mid \text{rank}([a_{ij}]) \leq 1\}$. Then the stratification $\{Z, X \setminus Z\}$ is a Whitney stratification of X .*

We use Whitney stratifications in Corollary 2.9. We will see that, up to a sign, the ML degree of a singular variety is equal to the Euler characteristic with some correction terms.

The correction terms are linear combinations of the ML degree of smaller dimensional strata of the Whitney stratification, whose coefficients turn out to be the Euler obstructions.

2.2. Gaussian degrees. We have defined the notion of ML degree of a projective variety. Sometimes, it is more convenient to restrict the projective variety to some affine torus and consider the notion of ML degree of a subvariety of the affine torus.

We call a closed irreducible subvariety of an affine torus $(\mathbb{C}^*)^n$ a *very affine variety*. Denote the coordinates of $(\mathbb{C}^*)^n$ by z_1, z_2, \dots, z_n . The likelihood functions in the affine torus $(\mathbb{C}^*)^n$ are of the form

$$l_u := z_1^{u_1} z_2^{u_2} \dots z_n^{u_n}.$$

Definition 2.5. Let $\mathcal{X} \subset (\mathbb{C}^*)^n$ be a very affine variety. Define the ML degree of \mathcal{X} , denoted by $\text{MLdeg}^o(\mathcal{X})$, to be the number of critical points of a likelihood function l_u for general u_1, u_2, \dots, u_n .

Fix an embedding of $\mathbb{P}^n \rightarrow \mathbb{P}^{n+1}$ by $[p_0 : p_1 : \dots : p_n] \mapsto [p_0 : p_1 : \dots : p_n : p_0 + p_1 + \dots + p_n]$. Given a projective variety $X \subset \mathbb{P}^n$, we can consider it as a subvariety of \mathbb{P}^{n+1} by the embedding we defined above. Then as a subvariety of \mathbb{P}^{n+1} , X is contained in the hyperplane $p_0 + p_1 + \dots + p_n - p_s = 0$.

Consider $(\mathbb{C}^*)^{n+1}$ as an open subvariety of \mathbb{P}^{n+1} , given by the open embedding

$$(z_0, z_1, \dots, z_n) \mapsto [z_0 : z_1 : \dots : z_n : 1].$$

Now, for the projective variety $X \subset \mathbb{P}^n$, we can embed X into \mathbb{P}^{n+1} as described above, and then take the intersection with $(\mathbb{C}^*)^{n+1}$. Thus, we obtain a very affine variety, which we denote by X^o .

Lemma 2.6. The ML degree of X as a projective variety is equal to the ML degree of X^o as a very affine variety, i.e.,

$$(6) \quad \text{MLdeg}(X) = \text{MLdeg}^o(X^o).$$

Proof. Fix general $u_0, u_1, \dots, u_n \in \mathbb{C}$. The ML degree of X is defined as the number of critical points of the likelihood function $(p_0/p_s)^{u_0} (p_1/p_s)^{u_1} \dots (p_n/p_s)^{u_n}$. The ML degree of X^o is defined as the number of critical points of $z_0^{u_0} z_1^{u_1} \dots z_n^{u_n}$. The two functions are equal on X^o . Therefore, they have the same number of critical points. ■

For the rest of this section, by ML degree we always mean ML degree of very affine varieties.

As observed in [3], the ML degree is equal to the Gaussian degree defined by Franekci and Kapranov [8]. The main theorem of [8] relates the Gaussian degree with Euler characteristics. In this section, we will review their main result together with the explicit formula from [7] to compute characteristic cycles.

First, we follow the notation in [3]. Fix a positive integer n for the dimension of the ambient space of the very affine variety. Denote the affine torus $(\mathbb{C}^*)^n$ by G and denote its Lie algebra by \mathfrak{g} . Let T^*G be the cotangent bundle of G . Thus, T^*G is a $2n$ dimensionally ambient space with a canonical symplectic structure. For any $\gamma \in \mathfrak{g}^*$, let $\Omega_\gamma \subset T^*G$ be the graph of the corresponding left invariant 1-form on G .

Suppose $\Delta \subset T^*G$ is a Lagrangian subvariety of T^*G . For a generic $\gamma \in \mathfrak{g}^*$, the intersection $\Delta \cap \Omega_\gamma$ is transverse and consists of finitely many points. The number of points in $\Delta \cap \Omega_\gamma$ is constant when γ is contained in a nonempty Zariski open subset of \mathfrak{g}^* . This number is called the Gaussian degree of Δ and denoted by $\text{gdeg}(\Delta)$.

Let $\mathcal{X} \subset G$ be an irreducible closed subvariety. Denote the conormal bundle of \mathcal{X}_{reg} in G by $T_{\mathcal{X}_{\text{reg}}}^*G$, and denote its closure in T^*G by $T_{\mathcal{X}}^*G$. Then $T_{\mathcal{X}}^*G$ is an irreducible conic Lagrangian subvariety of T^*G . In the language of [19], the $T_{\mathcal{X}}^*$ is the likelihood correspondence. Given any $\gamma \in \mathfrak{g}^*$, the left invariant 1-form corresponding to γ degenerates at a point $P \in \mathcal{X}$ if and only if $T_{\mathcal{X}}^*G \cap \Omega_\gamma$ contains a point in T_P^*G . Thus, we have the following lemma.

Lemma 2.7 (see [3]). *Let $\mathcal{X} \subset G$ be an irreducible closed subvariety where $G = (\mathbb{C}^*)^n$. Then*

$$(7) \quad \text{MLdeg}^0(\mathcal{X}) = \text{gdeg}(T_{\mathcal{X}}^*G).$$

Let \mathcal{F} be a bounded constructible complex on G , and let $CC(\mathcal{F})$ be its characteristic cycle. Then $CC(\mathcal{F}) = \sum_{1 \leq j \leq l} n_j \cdot [\Delta_j]$ is a \mathbb{Z} -linear combination of irreducible conic Lagrangian subvarieties in the cotangent bundle T^*G . The Gaussian degree and Euler characteristic are related by the following theorem.

Theorem 2.8 (see [8]). *With the notation above, we have*

$$(8) \quad \chi(G, \mathcal{F}) = \sum_{1 \leq j \leq l} n_j \cdot \text{gdeg}(\Delta_j).$$

2.3. Euler obstructions. An *Euler obstruction* is defined to be a coefficient of a characteristic cycle decomposition (see equation (9)), and by a theorem of Kashiwara it can be computed as the Euler characteristic of a complex link (see Theorem 2.10).

Let $\bigsqcup_{j=1}^k S_j$ be a Whitney stratification of the very affine variety \mathcal{X} with $S_1 = Y_{\text{reg}}$. Let e_{j1} be the Euler obstruction of the pair $(S_j, \mathcal{X}_{\text{reg}})$, which measures the singular behavior of \mathcal{X} along S_j . More precisely, e_{j1} are defined such that the following equality holds (see [7, section 1.1] for more details):

$$(9) \quad CC(\mathbb{C}_{S_1}) = e_{11}[T_{\mathcal{X}_{\text{reg}}}^*G] + \sum_{2 \leq j \leq k} e_{j1}[T_{S_j}^*G].$$

For example, $e_{11} = (-1)^{\dim \mathcal{X}}$.

Since $\chi(S_1) = \chi(G, \mathbb{C}_{S_1})$, combining (7)–(9), we have the following corollary that expresses $\chi(S_1)$ as a weighted sum of ML degrees.

Corollary 2.9. *Let $\bigsqcup_{j=1}^k S_j$ be a Whitney stratification of \mathcal{X} with $S_1 = \mathcal{X}_{\text{reg}}$. Then*

$$\chi(\mathcal{X}_{\text{reg}}) = e_{11} \text{MLdeg}^0(\mathcal{X}) + \sum_{2 \leq j \leq k} e_{j1} \text{MLdeg}^0(\bar{S}_j),$$

where \bar{S}_j denotes the closure of S_j in \mathcal{X} and e_{j1} is the Euler obstruction of the pair (S_j, S_1) .

Even though the abstract definition of Euler obstruction uses characteristic cycles, there is a concrete topological formula computing Euler obstructions due to Kashiwara. Here we recall that the Euler characteristic with compact support of a topological space M is defined by

$$\chi_c(M) = \sum_{i \geq 0} \dim H_c^i(M, \mathbb{Q}),$$

where H_c^i is the i th cohomology with compact support. When M is a d -dimensional (possibly noncompact) orientable manifold, the Poincaré duality of M implies that

$$\dim H^i(M, \mathbb{Q}) = \dim H_c^{d-i}(M, \mathbb{Q}),$$

and hence we have $\chi_c(M) = (-1)^d \chi(M)$.

Theorem 2.10 (Kashiwara).¹ *Under the above notation, fix a point $z \in S_j$, and fix an embedding of a neighborhood of z in X to an affine space W . Let V be an affine subspace of W with $\dim V + \dim S_j = \dim W$, which intersects S_j transversally at z .*

Then the Euler obstruction can be computed using Euler characteristic with compact support:

$$(10) \quad e_{j1} = (-1)^{\dim S_j + 1} \chi_c(B \cap Y_{\text{reg}} \cap \phi^{-1}(\epsilon)),$$

where B is a ball of radius δ in W centered at z , ϕ is a general linear function on V vanishing at z , and $0 < |\epsilon| \ll \delta \ll 1$.

The affine subspace V is called a “normal slice” of S_j , and the intersection $B \cap Y_{\text{reg}} \cap \phi^{-1}(\epsilon)$ is called a “complex link” of the pair S_j, Y_{reg} . By the formula above, every Euler obstruction we consider is computable.

Example 2.11. Consider the variety of rank at most 2 matrices X in \mathbb{P}^9 defined by

$$\det[p_{ij}]_{3 \times 3} = p_{11} + p_{12} + \cdots + p_{33} - p_s = 0.$$

Denote the very affine subvariety of X by X^o , that is,

$$X^o := \{p \in X \mid p_{ij} \neq 0 \text{ for all } i, j \text{ and } p_s \neq 0\}.$$

The Whitney stratification of X^o consists of $S_1 = X_{\text{reg}}^o$, and S_2 , the singular points of X^o , which are the rank 1 matrices. By Corollary 2.9, we have

$$(11) \quad \chi(X_{\text{reg}}^o) = e_{11} \text{MLdeg}^o(X^o) + e_{21} \text{MLdeg}^o(\bar{S}_2).$$

With this equation, we determine the ML degree of X . The rank 1 matrices are known to have ML degree one, so $\text{MLdeg}^o(\bar{S}_2) = 1$. The Euler obstruction e_{21} is equal to the Euler characteristic of the complex link as in Theorem 2.10. In this case, the complex link turns out to be homeomorphic to a vector bundle over \mathbb{P}^1 (see Lemma 3.2 for more details). The sign in front of the Euler characteristic is negative, and hence $e_{21} = -2$.

The Euler obstruction e_{11} is much easier to determine because this always equals $(-1)^{\dim X}$. So here $e_{11} = -1$. In subsection 3.2, we will calculate the Euler characteristic of $\chi(X_{\text{reg}}^o)$ to be -12 . Therefore, (11) implies that $\text{MLdeg}(X) = 10$, concluding the example.

¹Here the formula is in the form of [7, Theorem 1.1]; see also [5, page 100] and [10, section 8.1].

In Example 2.11, we used Corollary 2.9 and topological computations to determine the ML degree of a singular variety. In the next section we will again use Corollary 2.9 and topological computations to determine ML degrees.

3. The ML degree for rank 2 matrices. To the projective space \mathbb{P}^{mn} we associate the coordinates $p_{11}, p_{12}, \dots, p_{mn}, p_s$. Let X_{mn} denote the variety defined by $p_{11} + \dots + p_{mn} - p_s = 0$ and the vanishing of the 3×3 minors of the matrix

$$(12) \quad \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{(m-1)1} & p_{(m-1)2} & \cdots & p_{(m-1)n} \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}.$$

We think of X_{mn} as the Zariski closure of the set of rank 2 matrices in the distinguished hyperplane of \mathbb{P}^{mn} . Let Z_{mn} be the subvariety of X_{mn} defined by the vanishing of the 2×2 minors of the matrix (12). Then Z_{mn} is the singular locus of X_{mn} for $m, n \geq 3$.

Denote the very affine varieties associated to X_{mn} and Z_{mn} by X_{mn}^o and Z_{mn}^o , respectively.

We will make topological computations to determine the ML degree of X_{mn} . Summarizing, according to Lemma 2.6, $\text{MLdeg}(X_{mn}) = \text{MLdeg}^o(X_{mn}^o)$. Proposition 3.1 gives a Whitney decomposition of X_{mn} and determines the Euler obstructions for this stratification. Since being Whitney regular and Euler obstructions are local invariants, we obtain a Whitney stratification of X_{mn}^o and the Euler obstructions. Using these computations, we reduce our problem to determining a single Euler characteristic $\chi(X_{mn}^o \setminus Z_{mn}^o)$ by Corollary 3.3. Next, Theorem 3.4 provides a closed form expression of $\chi(X_{mn}^o \setminus Z_{mn}^o)$, for fixed m , in terms of the elements of a finite sequence Λ_m of integers. We conclude this section by computing Λ_m for $m = 3$, thereby proving Theorem 3.12.

3.1. Calculating Euler obstructions. We will start by proving general results for $m, n \geq 2$.

At the end of this section, we will specialize to the case where $m = 3$. With some topological computations, we determine λ_1, λ_2 of Λ_3 , thereby giving a closed form expression of the ML degree of X_{3n} [Theorem 3.12].

To ease notation, we let $e_{(mn)}$ denote the Euler obstruction e_{21} of the pair $\{Z_{mn}, X_{mn} \setminus Z_{mn}\}$, which is a Whitney stratification of X_{mn} by Corollary 2.4, and we have

$$Y_{mn} := X_{mn} \setminus Z_{mn}.$$

Proposition 3.1. *Denote the Euler obstruction of the pair of strata (Z_{mn}, Y_{mn}) by $e_{(mn)}$. Then*

$$(13) \quad e_{(mn)} = (-1)^{m+n-1}(\min\{m, n\} - 1).$$

Proof. Without loss of generality, we assume that $m \leq n$. When $m = 2$, X_{mn} is the distinguished hyperplane defined by $p_{11} + p_{12} + \dots + p_{mn} = p_s$, and $Z_{mn} \subset X_{mn}$ is a smooth subvariety. It follows from the definition of Euler obstruction that $e_{(mn)} = (-1)^{\dim Z_{mn} + 1}$. Since $\dim Z_{mn} = m + n - 2$, the proposition follows. Therefore, we can assume $m \geq 3$.

We will compute the Euler obstructions using Theorem 2.10. Since X_{mn} is contained in the distinguished hyperplane, consider X_{mn} as a subvariety of \mathbb{P}^{mn-1} , the projective space of

all $m \times n$ matrices. To simplify notation, we will write X and Z instead of X_{mn} and Z_{mn} . Next, we will compute the Euler obstruction $e_{(mn)}$ of the pair $Z, X \setminus Z$ using Theorem 2.10.

One can easily compute that $\dim Z = m + n - 2$. Under the notation of Theorem 2.10, since $B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)$ is a complex manifold, it is orientable and even-dimensional, and hence

$$\chi_c(B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)) = \chi(B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon))$$

by Poincaré duality. Since the Euler characteristic is homotopy invariant, and since $\chi(\mathbb{P}^{m-2}) = m - 1$, the second part of the proposition follows from the following lemma, where we show that the link is homotopy equivalent to \mathbb{P}^{m-2} . ■

Lemma 3.2. *Suppose $m \leq n$. With the above notation, we have that $B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)$ is homotopy equivalent to \mathbb{P}^{m-2} .*

Proof. First, we give a concrete description of the normal slice V . Notice that $X \subset \mathbb{P}^{mn}$ is contained in the distinguished hyperplane $p_{11} + \dots + p_{mn} - p_s = 0$. In this proof, we will consider X as a subvariety of \mathbb{P}^{mn-1} with homogeneous coordinates p_{11}, \dots, p_{mn} . Denote the affine chart $p_{11} \neq 0$ of \mathbb{P}^{mn-1} by U_{11} . Let $a_{ij} = \frac{p_{ij}}{p_{11}}$ ($(i, j) \neq (1, 1)$) be the affine coordinates of U_{11} , and let $a_{11} = 1$. Denote the origin of U_{11} by O . Let O be the fixed point z in Theorem 2.10, and let U_{11} be the affine space W .

Now, we define a projection $\pi : U_{11} \rightarrow Z \cap U_{11}$ by $[a_{ij}] \mapsto [b_{ij}]$, where $b_{ij} = a_{i1} \cdot a_{1j}$ and $a_{11} = 1$. Then U_{11} becomes a vector bundle over $Z \cap U_{11}$ via $\pi : U_{11} \rightarrow Z \cap U_{11}$. The preimage of O is the vector space parameterized by a_{ij} with $2 \leq i \leq m, 2 \leq j \leq n$.

In terms of matrices, we can think of π as the following map:

$$\begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \xrightarrow{\pi} \begin{bmatrix} 1 \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \begin{bmatrix} 1 \\ a_{12} \\ \vdots \\ a_{1n} \end{bmatrix}^T = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{21}a_{12} & \cdots & a_{21}a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m1}a_{12} & \cdots & a_{m1}a_{1n} \end{bmatrix},$$

and we think of the preimage of O as

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \pi^{-1} \left(\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \right).$$

By the above construction, we can take the normal slice V at O to be the fiber $\pi^{-1}(O)$. The intersection $V \cap X$ is clearly isomorphic to the affine variety $\{[a_{ij}]_{2 \leq i \leq m, 2 \leq j \leq n} \mid \text{rank} \leq 1\}$. Thus, we can define a map $\rho : V \cap (X \setminus Z) \rightarrow \mathbb{P}^{m-2}$, which maps the matrix $\{[a_{ij}]_{2 \leq i \leq m, 2 \leq j \leq n}\}$ to one of its nonzero column vectors, as an element in \mathbb{P}^{m-2} . Since the rank of $\{[a_{ij}]_{2 \leq i \leq m, 2 \leq j \leq n}\}$ is 1, the map does not depend on which nonzero column vector we choose. Using basic linear algebra, it is straightforward to check the following two statements about ρ :

- The restriction of ρ to $B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)$ is surjective.
- The restriction of ρ to $B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)$ has convex fibers; that is, the fibers are convex in V with respect to its natural complex vector space structure.

Thus, ρ is a fiber bundle with contractible fibers. By choosing a section, one can fiber-wise contract the total space of the fiber bundle to the chosen section. Hence ρ induces a homotopy equivalence between $B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)$ and \mathbb{P}^{m-2} . For a more general and technical topological statement, see the main theorem of [24]. ■

Corollary 3.3. *Let X_{mn} and Y_{mn}^o be defined as in the beginning of this section. Then*

$$(14) \quad \chi(Y_{mn}^o) = -\text{MLdeg}(X_{mn}) + (-1)^{m+n-1}(\min\{m, n\} - 1).$$

Proof. By Example 1.5, we know that $\text{MLdeg}(Z_{mn}) = 1$. One can easily compute that $\dim X_{mn}^o = 2m + 2n - 3$. Now, the corollary follows from (13) and Corollary 2.9. ■

3.2. Calculating Euler characteristics and ML degrees. In this subsection, an expression for the Euler characteristic $\chi(Y_{mn}^o)$ is given to determine formulas for ML degrees.

Recall that X_{mn}^o is the complement of all the coordinate hyperplanes in X_{mn} , and that $Z_{mn} \subset X_{mn}$ is the subvariety corresponding to rank 1 matrices of size $m \times n$.

Theorem 3.4. *Fix m to be an integer greater than two. Then there exists a sequence, denoted Λ_m , of integers $\lambda_1, \lambda_2, \dots, \lambda_{m-1}$ such that*

$$(15) \quad \chi(Y_{mn}^o) = (-1)^{n-1} \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} - \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} \cdot i^{n-1} \quad \text{for } n \geq 2.$$

Before proving Theorem 3.4, we quote a hyperplane arrangement result, which follows immediately from the theorem of Orlik and Solomon (see, e.g., [21, Theorem 5.90]).

Lemma 3.5. *Let L_1, \dots, L_r be distinct hyperplanes in \mathbb{C}^s . Suppose they are in general position, that is, the intersection of any t hyperplanes from $\{L_1, \dots, L_r\}$ has codimension t , for any $1 \leq t \leq s$. Denote the complement of $L_1 \cup \dots \cup L_r$ in \mathbb{C}^s by M . Then*

- if $r = s + 1$, then $\chi(M) = (-1)^s$;
- if $r = s + 2$, then $\chi(M) = (-1)^s(s + 1)$.

We will use the preceding lemma to compute Euler characteristics of strata of the following stratifications. Throughout, we assume that m is fixed and, to simplify indices, have \mathcal{Y}_n denoting Y_{mn}^o , the rank 2 matrices with nonzero coordinates whose entries sum to 1. In other words,

$$\mathcal{Y}_n = \left\{ [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n} \mid a_{ij} \in \mathbb{C}^*, \sum_{1 \leq i \leq m, 1 \leq j \leq n} a_{ij} = 1, \text{rank}[a_{ij}] = 2 \right\}.$$

A stratification of \mathcal{Y}_n is given by the number of columns which a matrix has summing to zero. Defining $\mathcal{Y}_n^{(l)}$ as

$$\mathcal{Y}_n^{(l)} := \left\{ [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathcal{Y}_n \mid l = \# \left\{ j \mid \sum_{1 \leq i \leq m} a_{ij} = 0 \right\} \right\}$$

yields the stratification

$$\mathcal{Y}_n = \mathcal{Y}_n^{(0)} \sqcup \dots \sqcup \mathcal{Y}_n^{(n-1)},$$

where each $\mathcal{Y}_n^{(l)}$ is a locally closed subvariety of \mathcal{Y}_n . Note that by definition, $\mathcal{Y}_n^{(0)}$ is the set of rank 2 matrices with nonzero column sums, i.e.,

$$\mathcal{Y}_n^{(0)} := \left\{ [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathcal{Y}_n \mid \sum_{1 \leq i \leq m} a_{ij} \neq 0 \text{ for each } 1 \leq j \leq n \right\}.$$

The following lemma will show $\chi(\mathcal{Y}_n^{(0)}) = \chi(\mathcal{Y}_n)$ by proving $\chi(\mathcal{Y}_n^{(l)}) = 0$ for $l \geq 1$.

Lemma 3.6. *The Euler characteristic of rank 2 matrices that sum to one equals the Euler characteristic of rank 2 matrices with nonzero column sums that sum to one. In other words, with notation as above,*

$$(16) \quad \chi(\mathcal{Y}_n) = \chi(\mathcal{Y}_n^{(0)}) \text{ and } \chi(\mathcal{Y}_n^{(1)}) = \chi(\mathcal{Y}_n^{(2)}) = \dots = \chi(\mathcal{Y}_n^{(n-1)}) = 0.$$

Proof. We define a \mathbb{C}^* action on \mathcal{Y}_n by putting $t \cdot (a_{ij}) = a'_{ij}$, where

$$a'_{ij} = \begin{cases} a_{ij} & \text{if } a_{1j} + \dots + a_{mj} \neq 0, \\ t \times a_{ij} & \text{if } a_{1j} + \dots + a_{mj} = 0. \end{cases}$$

It is straightforward to check that the action preserves each $\mathcal{Y}_n^{(l)}$. The action is transitive and continuous on $\mathcal{Y}_n^{(l)}$ for any $l \geq 1$. Recall that when an algebraic variety M admits a \mathbb{C}^* action, $\chi(M) = \chi(M^{\mathbb{C}^*})$, where $M^{\mathbb{C}^*}$ is the fixed locus of the action (see [7, Proposition 1.2]). Therefore, $\chi(\mathcal{Y}_n^{(l)}) = 0$ for any $l \geq 1$, and hence

$$\chi(\mathcal{Y}_n) = \chi(\mathcal{Y}_n^{(0)}). \quad \blacksquare$$

A column sum of a matrix in $\mathcal{Y}_n^{(0)}$ can be any element of \mathbb{C}^* . Now we consider a subset of $\mathcal{Y}_n^{(0)}$, where the column sums are all exactly 1. Let V_n denote the set of rank 2 matrices with column sums equal to 1, i.e.,

$$V_n := \left\{ [b_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n} \mid b_{ij} \in \mathbb{C}^*, \sum_{1 \leq i \leq m} b_{ij} = 1 \text{ for each } 1 \leq j \leq n, \text{rank}([b_{ij}]) = 2 \right\}.$$

In Lemma 3.7, we express the Euler characteristic of $\mathcal{Y}_n^{(0)}$ in terms of $\chi(V_n)$.

Lemma 3.7.

$$(17) \quad \chi(\mathcal{Y}_n^{(0)}) = (-1)^{n-1} \chi(V_n).$$

Proof. Let $T_n = \{(t_j)_{1 \leq j \leq n} \in (\mathbb{C}^*)^n \mid \sum_j t_j = 1\}$. Define a map $F : T_n \times V_n \rightarrow \mathcal{Y}_n^{(0)}$ by putting $a_{ij} = t_j b_{ij}$; we think of the j th element of T_n as scaling the j th column of V_n . Clearly, F is an isomorphism. Therefore, $\chi(\mathcal{Y}_n^{(0)}) = \chi(T_n) \cdot \chi(V_n)$. T_n can be considered as \mathbb{C}^{n-1} removing n hyperplanes in a general position. By Lemma 3.5, $\chi(T_n) = (-1)^{n-1}$, and hence $\chi(\mathcal{Y}_n^{(0)}) = (-1)^{n-1} \chi(V_n)$. \blacksquare

A stratification for V_n can be given by the minimal j_0 such that the column vector $[b_{ij_0}]$ is linearly independent from $[b_{i1}]$. Since $\sum_i b_{ij} = 1$ for all j , if two column vectors are linearly dependent, they must be equal. Defining $V_n^{(l)}$ as

$$V_n^{(l)} := \{[b_{ij}] \in V_n \mid \text{the column vectors satisfy } [b_{i1}] = [b_{i2}] = \dots = [b_{i(n-l+1)}] \neq [b_{i(n-l+2)}]\}$$

yields the following stratification of V_n by locally closed subvarieties:

$$V_n = \bigsqcup_{l=2}^n V_n^{(l)} = V_n^{(2)} \sqcup V_n^{(3)} \sqcup \dots \sqcup V_n^{(n)}.$$

Hence

$$\chi(V_n) = \chi(V_n^{(2)}) + \chi(V_n^{(3)}) + \dots + \chi(V_n^{(n)}).$$

We use W_l to denote the $m \times l$ rank 2 matrices with column sums equal to 1 such that the first two column vectors are linearly independent. Then we have the isomorphism $V_n^{(l)} \cong W_l$ by the following map:

$$\underbrace{[\begin{matrix} b_{i1} & b_{i1} & \dots & b_{i1} \end{matrix}]}_{(n-l+1) \text{ copies of } b_{i1}} \quad [\begin{matrix} b_{i(n-l+2)} & \dots & b_{in} \end{matrix}] \mapsto [\underbrace{ \begin{matrix} b_{i1} & b_{i(n-l+2)} & \dots & b_{in} \end{matrix} }_{l \text{ columns}}].$$

Therefore,

$$(18) \quad \chi(V_n) = \chi(W_2) + \chi(W_3) + \dots + \chi(W_n).$$

For any $l \geq 2$, we can define a map $\pi_l : W_l \rightarrow W_2$ by taking the first two columns. Thus, we can consider all W_l as varieties over W_2 . The following lemma gives a topological description of W_l , which will be useful to compute its Euler characteristics.

Lemma 3.8. *For any $l \geq 2$,*

$$W_l \cong W_3 \times_{W_2} W_3 \times_{W_2} \dots \times_{W_2} W_3,$$

where there are $l - 2$ copies of W_3 on the right-hand side and the product is the topological fiber product. In other words, given any point $x \in W_2$, the fiber of $\pi_l : W_l \rightarrow W_2$ over x is equal to the $(l - 2)$ th power of the fiber of $\pi_3 : W_3 \rightarrow W_2$ over x .

Proof. Given $l - 2$ elements in W_3 , suppose they all belong to the same fiber of $\pi_3 : W_3 \rightarrow W_2$. This means that we have $(l - 2)$ size $m \times 3$ matrices of rank 2, which all have the same first two columns. Then we can collect the third columns of all the matrices and put them after the same first two columns. Thus, we obtain an $m \times l$ matrix, whose rank is still two. In this way, we obtain a map $W_3 \times_{W_2} W_3 \times_{W_2} \dots \times_{W_2} W_3 \rightarrow W_l$,

$$\overbrace{[\begin{matrix} b_{i1} & b_{i2} & b_{i3} \end{matrix}], [\begin{matrix} b_{i1} & b_{i2} & b_{i4} \end{matrix}], \dots, [\begin{matrix} b_{i1} & b_{i2} & b_{il} \end{matrix}]}_{(l-2) \text{ } m \times 3 \text{ matrices}} \mapsto [\begin{matrix} b_{i1} & b_{i2} & b_{i3} & \dots & b_{il} \end{matrix}],$$

which is clearly an isomorphism. ■

Given a point $x \in W_2$, we let F_x denote the fiber of $\pi_3 : W_3 \rightarrow W_2$ over x . Since the Euler characteristic satisfies the product formula for fiber bundles, to compute the Euler characteristic of W_l , it suffices to study the stratification of W_2 by the Euler characteristic of F_x and compute the Euler characteristic of each stratum. More precisely, let

$$W_2^{(k)} := \{x \in W_2 \mid \chi(F_x) = -k\}, \text{ so } W_2 = \bigsqcup_{k \in \mathbb{Z}} W_2^{(k)}.$$

Then to compute $\chi(W_l)$, it suffices to compute $\chi(W_2^{(k)})$ for all $k \in \mathbb{Z}$. According to Lemma 3.9, it suffices to consider $k \in \{0, 1, \dots, m - 1\}$. For this reason, let Λ_m denote the sequence $\lambda_0, \lambda_1, \dots, \lambda_{m-1}$, where λ_k are defined by

$$(19) \quad \lambda_k := \chi(W_2^{(k)}) \text{ for } 0 \leq k \leq m - 1.$$

As we will soon see in the proofs, we can consider F_x as the complement of some point arrangement in \mathbb{C} . The arrangement is parameterized by the point $x \in W_2$. In other words, W_2 is naturally a parameter space of point arrangement in \mathbb{C} . According to the Euler characteristic of the corresponding point arrangement, W_2 is canonically stratified. Our problem is to compute the Euler characteristic of each stratum. The main difficulty in generalizing our method to compute the ML degree of higher rank matrices is to solve the corresponding problem for higher dimensional hyperplane arrangements.

Lemma 3.9. *For any $x \in W_2$,*

$$0 \geq \chi(F_x) \geq -(m - 1).$$

Moreover, the map $W_2 \rightarrow \mathbb{Z}$ defined by $x \mapsto \chi(F_x)$ is a semicontinuous function. In other words, for any integer k , $\{x \in W_2 \mid \chi(F_x) \geq -k\}$ is a closed algebraic subset of W_2 . In particular, the subsets $W_2^{(k)} = \{x \in W_2 \mid \chi(F_x) = -k\}$, $0 \leq k \leq (m - 1)$, give a stratification of W_2 into locally closed subsets.

Proof. By definition,

$$W_2 = \left\{ [b_{ij}]_{1 \leq i \leq m, j=1,2} \mid b_{ij} \in \mathbb{C}^*, \sum_{1 \leq i \leq m} b_{i1} = \sum_{1 \leq i \leq m} b_{i2} = 1, \text{rank}([b_{ij}]) = 2 \right\}.$$

Fix an element $x = [b_{i1} \ b_{i2}] \in W_2$. By definition, the fiber F_x of $\pi_3 : W_3 \rightarrow W_2$ is equal to the following:

$$F_x = \left\{ [b_{i3}] \in (\mathbb{C}^*)^m \mid \sum_{1 \leq i \leq m} b_{i3} = 1, [b_{i3}] \text{ is contained in the linear span of } [b_{i1}] \text{ and } [b_{i2}] \right\}.$$

Since each of the columns $[b_{i1}], [b_{i2}], [b_{i3}]$ sums to one, for any $[b_{i3}] \in F_x$ there exists $T \in \mathbb{C}$ such that $[b_{i3}] = T \cdot [b_{i1}] + (1 - T) \cdot [b_{i2}]$. For each i , this means $b_{i3} \neq 0$ is equivalent to $b_{i2} + T \cdot (b_{i1} - b_{i2}) \neq 0$. Therefore, for $x = [b_{i1} \ b_{i2}]$,

$$(20) \quad \begin{aligned} F_x &\cong \left\{ T \in \mathbb{C} \mid T \neq -\frac{b_{i2}}{b_{i1} - b_{i2}} \text{ if } b_{i1} - b_{i2} \neq 0 \text{ for } i = 1, 2, \dots, m \right\} \\ &\cong \mathbb{C} \setminus \left\{ -\frac{b_{i2}}{b_{i1} - b_{i2}} \mid 1 \leq i \leq m \text{ such that } b_{i1} - b_{i2} \neq 0 \right\}. \end{aligned}$$

Notice that $[b_{i1}] \neq [b_{i2}]$. Therefore, there has to be some i such that $b_{i1} \neq b_{i2}$. Thus F_x is isomorphic to \mathbb{C} minus some points of cardinality between 1 and m , and hence the first part of the lemma follows.

The condition that $\chi(F_x) \geq r$ is equivalent to the condition of some number of equalities $b_{i1} = b_{i2}$ and some number of overlaps among $\frac{b_{i2}}{b_{i1}-b_{i2}}$. Those conditions can be expressed by algebraic equations. Thus, the locus of x such that $\chi(F_x) \geq r$ is a closed algebraic subset in W_2 . ■

Example 3.10. When $m = 3$, the Euler characteristic $\chi(F_x)$ is in $\{0, -1, -2\}$. For example, if

$$x_0 = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ -2 & -3 \end{bmatrix}, \quad x_1 = \begin{bmatrix} 2 & 2 \\ 1 & -2 \\ -2 & 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 4 & 5 \\ -2 & 7 \\ -1 & -11 \end{bmatrix},$$

then $\chi(F_{x_0}) = 0, \chi(F_{x_1}) = -1, \chi(F_{x_2}) = -2$. This is because $F_{x_0} \cong \mathbb{C} \setminus \{3\}, F_{x_1} \cong \mathbb{C} \setminus \{-\frac{2}{3}, \frac{1}{3}\}$, and $F_{x_2} \cong \mathbb{C} \setminus \{5, -\frac{7}{9}, -\frac{11}{10}\}$. More generally, every matrix $x \in W_2$ has $\chi(F_x) \geq -2$. To have $x \in W_2$ such that $\chi(F_x) \geq -1$, at least one of the following six polynomials must vanish:

$$\begin{aligned} & (b_{31} - b_{32}), \quad (b_{21} - b_{22}), \quad (b_{11} - b_{12}), \\ & \det \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad \det \begin{bmatrix} b_{11} & b_{12} \\ b_{31} & b_{32} \end{bmatrix}, \quad \det \begin{bmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}. \end{aligned}$$

These conditions induce six irreducible components in W_2 whose points satisfy $\chi(F_x) \geq -1$. To have $x \in W_2$ such that $\chi(F_x) \geq 0$, two of the six polynomials above must vanish. However, this does not induce $15 = \binom{6}{2}$ components in W_2 . This is because in 12 of these cases there are no points in W_2 that satisfy the conditions. Instead, there are only three components.

Proof of Theorem 3.4. Thus far, we have shown with (16)–(18) the following relations:

$$(21) \quad \chi(\mathcal{Y}_n) = \chi(\mathcal{Y}_n^{(0)}) = (-1)^{n-1} \chi(V_n) = (-1)^{n-1} \sum_{2 \leq l \leq n} \chi(W_l).$$

Recall that $W_2^{(k)} = \{x \in W_2 \mid \chi(F_x) = -k\}$, and by Lemma 3.9 we have the stratification

$$(22) \quad W_2 = W_2^{(0)} \sqcup W_2^{(1)} \sqcup \dots \sqcup W_2^{(m-1)}.$$

Moreover, the $W_2^{(k)}$ are locally closed algebraic subsets of W_2 . The projection $\pi_3 : W_3 \rightarrow W_2$ induces a fiber bundle over each $W_2^{(k)}$, and by definition of $W_2^{(k)}$, the fiber has Euler characteristic $-k$. With Lemma 3.8, we showed W_l is isomorphic to an $(l-2)$ fiber product of W_3 over $W_2^{(k)}$. Thus, restricting $\pi_l : W_l \rightarrow W_2$ to $W_2^{(k)}$, the induced bundle’s fiber has Euler characteristic $(-k)^{l-2}$. Then

$$(23) \quad \chi(\pi_l^{-1}(W_2^{(k)})) = \lambda_k \cdot (-k)^{l-2}.$$

Since the $W_2^{(k)}$ are locally closed algebraic subsets of W_2 , $\pi_l^{-1}(W_2^{(k)})$ are locally closed algebraic subsets of W_l . Therefore, the additivity of Euler characteristic implies the following:

$$(24) \quad \chi(W_l) = \sum_{0 \leq k \leq m-1} \chi(\pi_l^{-1}(W_2^{(k)})) = \sum_{0 \leq k \leq m-1} \lambda_k \cdot (-k)^{l-2}.$$

Here our convention is $0^0 = 1$.

By equations (21) and (24) and Proposition 3.11, we have the following equalities:

$$\chi(Y_{mn}^o) = (-1)^{n-1} \sum_{2 \leq l \leq n} \left(\sum_{0 \leq k \leq m-1} \lambda_k \cdot (-k)^{l-2} \right) = (-1)^{n-1} \sum_{0 \leq k \leq m-1} \lambda_k \cdot \frac{1 - (-k)^{n-1}}{1 - (-k)}.$$

The last line becomes the same as (15) by replacing k by i and showing $\lambda_0 = 0$ in Proposition 3.11. ■

Proposition 3.11. *Let λ_0 of Λ_m be defined as above; then $\lambda_0 = 0$.*

We divert the proof of Proposition 3.11 to the end of the section. Now we specify to the case $m = 3$ for our first main result.

Theorem 3.12 (main result). *The ML degree of X_{3n} is given by the following formula.*

$$(25) \quad \text{MLdeg}(X_{3n}) = 2^{n+1} - 6.$$

Proof. In [15], the ML degree of X_{32} and ML degree of X_{33} are determined to be 1 and 10, respectively. With this information, it follows that $\lambda_1 + \lambda_2 = 0$ and $\lambda_2 = 12$ by Theorem 3.4. ■

The takeaway is that finitely many computations can determine infinitely many ML degrees. Using these techniques we may be able to determine ML degrees of other varieties, such as symmetric matrices and Grassmannians, with a combination of applied algebraic geometry and topological arguments.

Proof of Proposition 3.11. Recall that $\lambda_0 = \chi(W_2^{(0)})$. By definition, $W_2^{(0)}$ consists of those $[b_{i1} \ b_{i2}]$ in W_2 such that the cardinality of the set $\{b_{i1}/b_{i2} | 1 \leq i \leq m, b_{i1} \neq b_{i2}\}$ is equal to 1.

Notice that for $[b_{i1} \ b_{i2}] \in W_2^{(0)}$,

$$(26) \quad \sum_{\substack{1 \leq i \leq m \\ b_{i1} \neq b_{i2}}} b_{i1} = \sum_{\substack{1 \leq i \leq m \\ b_{i2} \neq b_{i2}}} b_{i1} = 0.$$

Therefore, we can define a \mathbb{C}^* action on the set of m by two matrices $\{[b_{ij}]_{1 \leq i \leq m, j=1,2}\}$ by setting $t \cdot (b_{ij}) = (b'_{ij})$, where

$$(27) \quad b'_{ij} = \begin{cases} b_{ij} & \text{if } b_{i1} = b_{i2}, \\ t \times b_{ij} & \text{otherwise.} \end{cases}$$

Now, it is straightforward to check that this \mathbb{C}^* action preserves $W_2^{(0)}$ and that the action is transitive on $W_2^{(0)}$. Therefore, $\chi(W_2^{(0)}) = \chi(\emptyset) = 0$. ■

4. Recursions and closed form expressions. In this section, we use Theorem 3.4 to give recursions for the Euler characteristic $\chi(Y_{mn}^o)$ and thus the ML degree of X_{mn} by (14). We break the recursions and give closed form expressions in Corollary 4.2.

4.1. The recurrence. Recall that we have $Y_{mn}^o := X_{mn}^o \setminus Z_{mn}^o$. By Theorem 3.4, giving a recursion for $\chi(Y_{mn}^o)$ is equivalent to giving a recursion for $-\text{MLdeg}(X_{mn}) + (-1)^{m+n-1}(m-1)$, assuming $m \leq n$. The next theorem gives the recursion for $\chi(Y_{mn}^o)$.

Theorem 4.1. Fix m , and let $-c_i$ be the coefficient of t^{m-i} in $(t+1) \prod_{r=1}^{m-1} (t-r)$. For $n > m$, we have

$$\chi(Y_{mn}^o) = c_1 \chi(Y_{m(n-1)}^o) + c_2 \chi(Y_{m(n-2)}^o) + \dots + c_m \chi(Y_{m(n-m)}^o).$$

Proof. By Theorem 3.4, we have

$$\chi(Y_{mn}^o) = (-1)^{n-1} \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} - \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} \cdot i^{n-1}$$

for $n \geq 2$. Therefore $\chi(Y_{mn}^o)$ is an order m linear homogeneous recurrence relation with constant coefficients. The coefficients of such a recurrence are described by the characteristic polynomial with roots $-1, 1, \dots, m-1$, i.e., $t^m - c_1 t^{m-1} - \dots - c_m = (t+1) \prod_{r=1}^{m-1} (t-r)$. ■

With these recurrences we determine the ML degrees listed in Table 1.

Table 1
ML degrees of X_{mn} .

$n =$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
$m :$	1	10	191	6776	378477	30305766
$m + 1 :$	1	26	843	40924	2865245	274740990
$m + 2 :$	1	58	3119	212936	19177197	2244706374
$m + 3 :$	1	122	10587	1015564	118430045	17048729886
$m + 4 :$	1	250	34271	4586456	692277357	122818757286
$m + 5 :$	1	506	107883	19984444	3892815965	850742384190
$m + 6 :$	1	1018	333839	84986216	21284701677	5720543812614
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

Note that our methods are not limited to Table 1. We give closed form formulas in the next section for infinite families of mixture models.

4.2. Closed form expressions. In this subsection we provide additional closed form expressions. Using an inductive procedure (described in the proof of Corollary 4.2), we determine Λ_m for small m that can be extended to arbitrary m .²

Corollary 4.2. For fixed $m = 2, 3, \dots, 7$, the closed form expressions for $\text{MLdeg}(X_{mn})$ with $m \leq n$ are

$$\text{MLdeg}(X_{2n}) = 1,$$

$$\text{MLdeg}(X_{3n}) = \left(\frac{-12}{2} \cdot 1^{n-1} + \frac{12}{3} \cdot 2^{n-1} \right),$$

$$\text{MLdeg}(X_{4n}) = \left(\frac{50}{2} \cdot 1^{n-1} + \frac{-120}{3} \cdot 2^{n-1} + \frac{72}{4} \cdot 3^{n-1} \right),$$

$$\text{MLdeg}(X_{5n}) = \left(\frac{-180}{2} \cdot 1^{n-1} + \frac{780}{3} \cdot 2^{n-1} + \frac{-1080}{4} \cdot 3^{n-1} + \frac{480}{5} \cdot 4^{n-1} \right),$$

$$\text{MLdeg}(X_{6n}) = \left(\frac{602}{2} \cdot 1^{n-1} + \frac{-4200}{3} \cdot 2^{n-1} + \frac{10080}{4} \cdot 3^{n-1} + \frac{-10080}{5} \cdot 4^{n-1} + \frac{3600}{6} \cdot 5^{n-1} \right),$$

$$\text{MLdeg}(X_{7n}) = \left(\frac{-1932}{2} + \frac{-20412}{3} \cdot 2^{n-1} + \frac{-75600}{4} \cdot 3^{n-1} + \frac{127680}{5} \cdot 4^{n-1} + \frac{-100800}{6} \cdot 5^{n-1} + \frac{30240}{7} \cdot 6^{n-1} \right).$$

²See the appendix for an implementation.

Proof. We find these formulas using an inductive procedure to determine Λ_m from Λ_{m-1} . With (14) and (15), we have the following $m - 1$ relations with $n = 2, 3, \dots, m$:

$$\begin{bmatrix} \text{MLdeg}(X_{m2}) \\ \text{MLdeg}(X_{m3}) \\ \vdots \\ \text{MLdeg}(X_{mm}) \end{bmatrix} + (-1)^m \begin{bmatrix} 1 \\ -2 \\ \vdots \\ (-1)^m(m-1) \end{bmatrix} = \mathbf{M} \begin{bmatrix} \lambda_1/2 \\ \lambda_2/3 \\ \vdots \\ \lambda_{m-1}/m \end{bmatrix}, \text{ where}$$

$$\mathbf{M} := \left(\begin{bmatrix} 1^1 & 2^1 & \cdots & (m-1)^1 \\ 1^2 & 2^2 & \cdots & (m-1)^2 \\ \vdots & \vdots & \ddots & \vdots \\ 1^{m-1} & 2^{m-1} & \cdots & (m-1)^{m-1} \end{bmatrix} - \begin{bmatrix} (-1)^1 & (-1)^1 & \cdots & (-1)^1 \\ (-1)^2 & (-1)^2 & \cdots & (-1)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{m-1} & (-1)^{m-1} & \cdots & (-1)^{m-1} \end{bmatrix} \right).$$

For fixed m , this system of linear equations has $2m - 2$ unknowns: $\text{MLdeg}(X_{mj})$ for $j = 2, \dots, m$ and $\lambda_1, \dots, \lambda_{m-1}$ of Λ_m . By induction, we may assume we know Λ_{m-1} . The Λ_{m-1} gives us a closed form expression for the ML degrees of $X_{(m-1)j}$ with $j \geq 2$. Since $\text{MLdeg}(X_{(m-1)j}) = \text{MLdeg}(X_{j(m-1)})$, we have reduced our system of linear equations to m unknowns by substitution. By Proposition 4.4, we have λ_{m-1} of Λ_m equals $(m - 1) \cdot m!$. Substituting this value as well, we have a linear system of $m - 1$ equations in $m - 1$ unknowns: $\text{MLdeg}(X_{mm}), \lambda_1, \lambda_2, \dots, \lambda_{m-2}$.

A linear algebra argument shows that there exists a unique solution of the system yielding each λ_j of Λ_m as well as $\text{MLdeg}(X_{mm})$. It proceeds as follows. Since the unknown $\text{MLdeg}(X_{mm})$ appears only in the last equation, the linear system above has a unique solution if and only if the determinant of the $(m - 2) \times (m - 2)$ upper left submatrix of \mathbf{M} , denoted \mathbf{N} , is nonsingular. Consider the row vector $\mathbf{c} := [c_1, c_2, \dots, c_{m-2}]$, and let $f_{\mathbf{c}}(x) = c_1x^1 + c_2x^2 + \dots + c_{m-2}x^{m-2}$. We will show that \mathbf{c} is a null vector for \mathbf{N} if and only if \mathbf{c} is the zero vector. If \mathbf{c} is a null vector, then by multiplying \mathbf{c} with \mathbf{N} , we see that

$$[f_{\mathbf{c}}(1) - f_{\mathbf{c}}(-1), f_{\mathbf{c}}(2) - f_{\mathbf{c}}(-1), \dots, f_{\mathbf{c}}(m-2) - f_{\mathbf{c}}(-1)]^T = 0.$$

In other words, $(f_{\mathbf{c}}(x) - f_{\mathbf{c}}(-1))$ has $m - 1$ distinct roots and must be the zero polynomial. This means $f_{\mathbf{c}}(x)$ is a constant, and $\mathbf{c} = \mathbf{0}$.

Using the inductive procedure described above, we determined the Λ_m listed in Table 2 to yield the closed form expressions we desired. ■

Table 2
The λ_i of Λ_m .

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
$\Lambda_2 :$	2					
$\Lambda_3 :$	-12	12				
$\Lambda_4 :$	50	-120	72			
$\Lambda_5 :$	-180	780	-1080	480		
$\Lambda_6 :$	602	-4200	10080	-10080	3600	
$\Lambda_7 :$	-1932	20412	-75600	127680	-100800	30240

Remark 4.3. With our recursive methods we determined the ML degree of $X_{20,20}$ to be

$$19\ 674\ 198\ 689\ 452\ 133\ 729\ 973\ 092\ 792\ 823\ 813\ 947\ 695 \approx 1.967 \times 10^{40}.$$

While it is currently infeasible to track this large number of paths via numerical homotopy continuation as in [14], it may be possible to track much fewer paths using the topology of the variety. Preliminary results by the 2016 Mathematics Research Community in Algebraic Statistics Likelihood Geometry Group [1] for weighted independence models deform a variety with ML degree one to a variety with large ML degree. These results deform the ML estimate of one model to the estimate of another. The stratifications presented here may lead to similar results in the case of mixtures.

Proposition 4.4. *Let λ_k of Λ_m be defined as in (19). Then λ_{m-1} of Λ_m equals $(m-1) \cdot m!$.*

Proof. Recall that λ_{m-1} of Λ_m equals $\chi(W_2^{(m-1)})$. By definition, $W_2^{(m-1)}$ consists of all $[b_{i1}\ b_{i2}] \in W_2$ such that $b_{i1} \neq b_{i2}$ for all $1 \leq i \leq m$ and b_{i1}/b_{i2} are distinct for $1 \leq i \leq m$.

Denote by B_m the subset of $(\mathbb{C}^* \setminus \{1\})^m$ corresponding to m distinct numbers. Then there is a natural map $\pi : W_2^{(m-1)} \rightarrow B_m$, defined by $[b_{ij}] \mapsto (\frac{b_{11}}{b_{12}}, \dots, \frac{b_{m1}}{b_{m2}})$. The map is surjective. Moreover, one can easily check that under the map π , $W_2^{(m-1)}$ is a fiber bundle over B_m . In addition, the fiber is isomorphic to the complement of m hyperplanes in \mathbb{C}^{m-2} in a general position. By Lemma 3.5, the fiber has Euler characteristic $(-1)^{m-2}(m-1)$.

The Euler characteristic of B_m is equal to $(-1)^m \cdot m!$ by induction. In fact, B_m is a fiber bundle over B_{m-1} with fiber homeomorphic to $\mathbb{C}^* \setminus \{m \text{ distinct points}\}$. Therefore,

$$\chi(W_2^{(m-1)}) = (-1)^{m-2}(m-1) \cdot (-1)^m m! = (m-1) \cdot m!. \quad \blacksquare$$

5. Conclusion and additional questions. We have developed the topological tools to determine the maximum likelihood (ML) degree of singular models. We proved a closed form expression for $3 \times n$ matrices with rank 2 conjectured by [14]. In addition, our results provide a recursion to determine the ML degree of a mixture of independence models, where the first random variable has m states and the second random variable has n states. Furthermore, we have shown how a combination of computational algebraic calculations and topological arguments can determine an infinite family of ML degrees.

The next natural question is to determine the ML degree for higher order mixtures (rank r matrices for $r > 2$). Our results give closed form expressions in the corank 1 case by *maximum likelihood duality* [6]. Maximum likelihood duality is quite surprising here because our methods would initially suggest that the corank 1 matrices have a much more complicated ML degree. It would be very nice to give a topological proof in terms of Euler characteristics of maximum likelihood duality. One possible approach is by applying these techniques to the *dual maximum likelihood estimation problem* described in [23].

Another question concerns the *boundary components* of statistical models as described in [20] for higher order mixtures. Can we also use these topological methods to give closed form expressions of the ML degrees of the boundary components of the statistical model?

Finally, one should notice that the formulas in Corollary 4.2 for the ML degrees involve alternating signs. It would be beneficial to give a canonical transformation of our alternating

sum formula into a positive sum formula. One reason why this might be possible is motivated by the work of [13]. Here, entries of the data u are degenerated to zero, and some of the critical points go to the boundary of the algebraic torus (i.e., coordinates go to zero). Based on how the critical points go to the boundary, one partitions the ML degree into a sum of positive integers.

Appendix. The following code creates a function called `recursiveMLDegrees` that will compute a table of ML degrees and other Euler characteristics. The input consists of two integers m, n with $m \leq n$. The output is a sequence of four elements:

- Element 0 of the output is the ML degree of X_{mn} .
- Element 1 of the output is the list $\Lambda_m = \lambda_1, \lambda_2, \dots, \lambda_{m-1}$.
- Element 2 of the output is a table of ML degrees given by a list of lists.
- Element 3 of the output lists Λ_i for $i = 2, 3, \dots, m$.

For example, `recursiveMLDegrees(3,4)` returns

(26, {-12, 12}, {{1, 1, 1}, {10, 26}}, {{2}, {-12, 12}})

and `(recursiveMLDegrees(7,n))_3` returns Table 2 for $n \geq 7$. Here is the code.

```

----MACAULAY2 CODE-----
recursiveMLDegrees=(M,N)->(m1R=QQ[a_0..a_N]**QQ[c_0..c_N]**QQ[ML_0..ML_N];
----m1Equations is a function that gives the relations in the proof of Corollary 4.2.
m1Equations=(m,n,m1degX)->( (-m1degX+(-1)^(m+n-1)*(min(m,n)-1))-
  sum for i from 1 to m-1 list (a_i*( (-1)^(n-1-i^(n-1))/(i+1)));
----M2N lists MLDegree(X_2,j) for j=2,3,...,N; These ML degrees are 1.
M2N:=for i from 2 to N list 1;
----tableML gives a list of lists of ML degrees.
----tableLam gives, for i=2,3,...,m,
-----the lists \Lambda_i={\lambda_1,\lambda_2,...,\lambda_{i-1}}.
tableML:={M2N};tableLam:={{2}};
----The loop constructs \Lambda_i and ML degrees using the recursion.
for fixM from 3 to M do (
  m1dList:=(for i from 2 to fixM-1 list tableML_(i-2)_(fixM-i))|{ML_fixM};
  solveI:=radical ideal({a_(fixM-1)-fixM!*(fixM-1)}|
    for fixN from 2 to fixM list m1Equations(fixM,fixN,m1dList_(fixN-2)));
  newLam:=for j from 1 to fixM-1 list a_j %solveI;
  newLamSub:=for j from 1 to fixM-1 list a_j=>newLam_(j-1);
  tableLam=append(tableLam,(toList newLamSub)/last);
  newMLdegrees:=for i from fixM to N list (sum for j from 1 to N list
    if j>= fixM then 0 else sub(a_j/(j+1)*j^(i-1), a_j=>newLam_(j-1)));
  tableML=append(tableML,newMLdegrees);
----The output of the function has four elements.
return (last last tableML,last tableLam, tableML,tableLam))
----EXAMPLE-----
INPUT: recursiveMLDegrees(3,4)
OUTPUT: (26, {-12, 12}, {{1, 1, 1}, {10, 26}}, {{2}, {-12, 12}})
----EXAMPLE-----
INPUT: (recursiveMLDegrees(3,4))_0
OUTPUT: 26
-----

```

Acknowledgments. We would like to thank Nero Budur and Jonathan Hauenstein for introducing the authors to one another. We would also like to especially thank Jan Draisma for his comments to improve an earlier version of this manuscript. We thank Sam Evens for the explanation of Theorem 2.10.

REFERENCES

- [1] C. AMENDOLA, N. BLISS, C. R. GIBBONS, M. HELMER, S. HOSTEN, E. D. NASH, J. I. RODRIGUEZ, AND D. SMOLKIN, *The Maximum Likelihood Degree of Toric Varieties*, preprint, <https://arxiv.org/abs/1703.02251>, 2017.
- [2] N. BUDUR AND B. WANG, *The signed Euler characteristic of very affine varieties*, Int. Math. Res. Not. IMRN, 2015 (2015), pp. 5710–5714, <https://doi.org/10.1093/imrn/rnu107>.
- [3] N. BUDUR AND B. WANG, *Bounding the maximum likelihood degree*, Math. Res. Lett., 22 (2015), pp. 1613–1620, <https://doi.org/10.4310/MRL.2015.v22.n6.a4>.
- [4] F. CATANESE, S. HOŞTEN, A. KHETAN, AND B. STURMFELS, *The maximum likelihood degree*, Amer. J. Math., 128 (2006), pp. 671–697, http://muse.jhu.edu/journals/american_journal_of_mathematics/v128/128.3catanese.pdf.
- [5] A. DIMCA, *Sheaves in Topology*, Universitext, Springer-Verlag, Berlin, 2004, <https://doi.org/10.1007/978-3-642-18868-8>.
- [6] J. DRAISMA AND J. RODRIGUEZ, *Maximum likelihood duality for determinantal varieties*, Int. Math. Res. Not. IMRN, 2014 (2014), pp. 5648–5666.
- [7] S. EVENS AND I. MIRKOVIĆ, *Characteristic cycles for the loop Grassmannian and nilpotent orbits*, Duke Math. J., 97 (1999), pp. 109–126, <https://doi.org/10.1215/S0012-7094-99-09705-3>.
- [8] J. FRANECKI AND M. KAPRANOV, *The Gauss map and a noncompact Riemann-Roch formula for constructible sheaves on semiabelian varieties*, Duke Math. J., 104 (2000), pp. 171–180, <https://doi.org/10.1215/S0012-7094-00-10417-6>.
- [9] W. FULTON, *Introduction to Toric Varieties*, The William H. Roever Lectures in Geometry, Ann. of Math. Stud. 131, Princeton University Press, Princeton, NJ, 1993, <https://doi.org/10.1515/9781400882526>.
- [10] V. GINSBURG, *Characteristic varieties and vanishing cycles*, Invent. Math., 84 (1986), pp. 327–402, <https://doi.org/10.1007/BF01388811>.
- [11] M. GORESKY AND R. MACPHERSON, *Stratified Morse theory*, in Singularities, Part 1 (Arcata, CA, 1981), Proc. Sympos. Pure Math. 40, Amer. Math. Soc., Providence, RI, 1983, pp. 517–533.
- [12] E. GROSS, M. DRTON, AND S. PETROVIĆ, *Maximum likelihood degree of variance component models*, Electron. J. Stat., 6 (2012), pp. 993–1016, <https://doi.org/10.1214/12-EJS702>.
- [13] E. GROSS AND J. I. RODRIGUEZ, *Maximum likelihood geometry in the presence of data zeros*, in Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation, ISSAC '14, ACM, New York, 2014, pp. 232–239, <https://doi.org/10.1145/2608628.2608659>.
- [14] J. D. HAUSTEIN, J. I. RODRIGUEZ, AND B. STURMFELS, *Maximum likelihood for matrices with rank constraints*, J. Algebr. Stat., 5 (2014), pp. 18–38.
- [15] S. HOŞTEN, A. KHETAN, AND B. STURMFELS, *Solving the likelihood equations*, Found. Comput. Math., 5 (2005), pp. 389–407, <https://doi.org/10.1007/s10208-004-0156-8>.
- [16] S. HOSTEN AND S. SULLIVANT, *The algebraic complexity of maximum likelihood estimation for bivariate missing data*, in Algebraic and Geometric Methods in Statistics, P. Gibilisco, E. Riccomagno, M. P. Rogantin, and H. P. Wynn, eds., Cambridge University Press, Cambridge, UK, 2009, pp. 123–134, <https://doi.org/10.1017/CBO9780511642401.008>.
- [17] R. Hotta, K. TAKEUCHI, AND T. TANISAKI, *D-Modules, Perverse Sheaves, and Representation Theory*, translated from the 1995 Japanese edition by K. Takeuchi, Progr. Math. 236, Birkhäuser Boston, Inc., Boston, MA, 2008, <https://doi.org/10.1007/978-0-8176-4523-6>.
- [18] J. HUH, *The maximum likelihood degree of a very affine variety*, Compos. Math., 149 (2013), pp. 1245–1266, <https://doi.org/10.1112/S0010437X13007057>.
- [19] J. HUH AND B. STURMFELS, *Likelihood geometry*, in Combinatorial Algebraic Geometry, Lecture Notes in Math. 2108, Springer, Cham, 2014, pp. 63–117, https://doi.org/10.1007/978-3-319-04870-3_3.

- [20] K. KUBJAS, E. ROBEVA, AND B. STURMFELS, *Fixed points of the EM algorithm and nonnegative rank boundaries*, Ann. Statist., 43 (2015), pp. 422–461, <https://doi.org/10.1214/14-AOS1282>.
- [21] P. ORLIK AND H. TERAOKA, *Arrangements of Hyperplanes*, Grundlehren Math. Wiss. [Fundamental Principles of Mathematical Sciences] 300, Springer-Verlag, Berlin, 1992, <https://doi.org/10.1007/978-3-662-02772-1>.
- [22] L. PACTHER AND B. STURMFELS, *Statistics*, in Algebraic Statistics for Computational Biology, Cambridge University Press, New York, 2005, pp. 3–42, <https://doi.org/10.1017/CBO9780511610684.004>.
- [23] J. I. RODRIGUEZ, *Maximum likelihood for dual varieties*, in Proceedings of the 2014 Symposium on Symbolic-Numeric Computation (SNC '14), ACM, New York, 2014, pp. 43–49, <https://doi.org/10.1145/2631948.2631959>.
- [24] S. SMALE, *A Vietoris mapping theorem for homotopy*, Proc. Amer. Math. Soc., 8 (1957), pp. 604–610, <https://doi.org/10.2307/2033527>.
- [25] E. H. SPANIER, *Algebraic Topology*, McGraw-Hill, New York, 1966.
- [26] C. UHLER, *Geometry of maximum likelihood estimation in Gaussian graphical models*, Ann. Statist., 40 (2012), pp. 238–261, <https://doi.org/10.1214/11-AOS957>.
- [27] H. WHITNEY, *Tangents to an analytic variety*, Ann. of Math. (2), 81 (1965), pp. 496–549, <https://doi.org/10.2307/1970400>.