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 a 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 one matrices over the probability simplex) is well known to have an ML degree of one, meaning their is a unique local maxima of the likelihood function. However, for mixtures of independence models (i.e. rank two 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.

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 [3, 13] and gives a measure of complexity to the global optimization problem, as it bounds the number of local maxima.

The maximum likelihood degree has been studied in many contexts. Some of these contexts include Gaussian graphical models [22], variance component models [10], and in missing data [14]. In this manuscript, we work in the context of discrete random variables (for a recent survey in this context, see [17]). In our main results, we provide closed form expressions for ML degrees mixtures of independence models, which are sets of joint probability distributions for two random variables. This answers an outstanding conjecture of [12].

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

$$
P = \begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1n} \\
p_{21} & p_{22} & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{m1} & p_{m2} & \cdots & p_{mn}
\end{bmatrix}.
$$
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 \(\triangle_{mn-1}\) of all such matrices \(P\).

If 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 non-negative integers where \(u_{ij}\) is the number of samples drawn with state \((i, j)\):

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

The likelihood function corresponding to the data matrix \(u\) is given by

\[
\ell_u(p) := p_{11}^{u_{11}} p_{12}^{u_{12}} \cdots p_{mn}^{u_{mn}}.
\]

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 maximum likelihood 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 \(\triangle_{mn-1}\) is said to be the mixture of independence models and is defined to be the image of the following map:

\[
(\triangle_{m-1} \times \triangle_{n-1}) \times (\triangle_{m-1} \times \triangle_{n-1}) \times \triangle_1 \rightarrow \mathcal{M}_{mn} (R_i, B_j, C) \mapsto c_1R_1B_1^T + c_2R_2B_2^T,
\]

where \(R_i, B_j, C := [c_1, c_2]^T\) are \(m \times 1\) matrices, \(n \times 1\), and \(2 \times 1\) 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 [12], the following table of ML degrees of \(\mathcal{M}_{mn}\) were computed:

\[
\begin{array}{cccccccc}
n = & 3 & 4 & 5 & 6 & 7 \\
m = 3 : & 10 & 26 & 58 & 122 & 250 \\
m = 4 : & 26 & 191 & 843 & 3119 \\
m = 5 : & 58 & 843 & 6776 & & \\
\end{array}
\]

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 [20], 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 has four nucleobases. 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:

$$\text{ML degree } M_{3n} = 2^{n+1} - 6 \text{ for } n \geq 3.$$ 

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

Our techniques relate ML degrees to Euler characteristics. Prior, Huh has shown that the ML degree of smooth algebraic statistical models $M$ with Zariski closure $X$ equals the signed topological characteristic of an open subvariety $X^o$ where $X^o$ is the set of points of $X$ with nonzero coordinates and coordinate sums [16]. More recent work of Budur and the second author show that the ML degree of a singular model is a stratified topological invariant. In [1], they show that the Euler characteristic of $X^o$ 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 illustrating examples to set notation and definitions.

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

To $\mathbb{P}^{n+1}$ we associate the coordinates $p_0, p_1, \ldots, p_n$, and $p_s$ (were $s$ stands for sum). Consider the distinguished hyperplane in $\mathbb{P}^{n+1}$ defined by $p_0 + \cdots + p_n - p_s$ ($p_s$ 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 $u_0, \ldots, u_n \in \mathbb{C}$. The likelihood function has the nice property that, up to scaling, its gradient is a rational function $\nabla \ell_u(p) := \left[ \frac{u_0}{p_0} : \frac{u_1}{p_1} : \cdots : \frac{u_n}{p_n} : \frac{u_s}{p_s} \right]$.

**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 $[p_0 : p_1 : p_2 : p_3 : p_s] = [70 : 180 : 105 : 270 : -625]$. Whenever each $u_i$ is not equal to 0, 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 maximum likelihood degree of $X$ is defined to be the number of critical points of the likelihood function on $X$ for general $u_0, \ldots, u_n$. 

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 explicitly is often quite difficult and commonly, probability one algorithms are used to compute ML degrees instead. However, with our results, we compute ML degrees using Euler characteristics and topological arguments.

1.3. Using Euler characteristics. In the definition of maximum likelihood 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 of $X$. Let

$$\chi([16])$$

Example 1.4. Suppose $X$ is a smooth projective variety of $\mathbb{P}^{n+1}$. Then,

$$\chi(X^o) = (-1)^{\dim X^o} \operatorname{MLdeg} X. \tag{5}$$

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 $X$. The product property says $\chi(M \times N) = \chi(M) \cdot \chi(N)$. More generally, the fibration property says that if $E \to B$ is a fibration with fiber $F$ then $\chi(E) = \chi(F) \cdot \chi(B)$.

Example 1.5. Consider $X$ from Example 1.2. The variety $X$ has the parameterization shown below

$$\mathbb{P}^1 \times \mathbb{P}^1 \to X$$

$$[x_0 : x_1] \times [y_0 : y_1] \mapsto [x_0y_0 : x_0y_1 : x_1y_0 : x_1y_1 : x_0y_0 + x_0y_1 + x_1y_0 + x_1y_1].$$

Let $X^o$ be the underlying very affine variety of $X$ and consider $\mathcal{O} := \mathbb{P}^1 \setminus \{(0 : 1), (1 : 0), (1 : -1)\}$ the projective space with 3 points removed. The very affine variety $X^o$ has a parameterization:

$$\mathcal{O} \times \mathcal{O} \to X^o$$

$$[x_0 : x_1] \times [y_0 : y_1] \mapsto [x_0y_0 : x_0y_1 : x_1y_0 : x_1y_1, (x_0 + x_1)(y_0 + y_1)].$$

Since $\chi(\mathbb{P}^1) = 2$, after removing 3 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 the ML degree of $X^o$ is 1 as well.

We call $X$ the variety of $2 \times 2$ matrices with rank 1 and restricting to the probability simplex yields the independence model. This example generalizes by considering the map $\mathbb{P}^{m-1} \times \mathbb{P}^{n-1} \to X$ given by $([x_0 : \cdots : x_{m-1}], [y_0 : \cdots : y_n]) \mapsto [x_0y_0 : \cdots : x_{m-1}y_{n-1} : \Sigma_{i,j} x_i y_j]$. In this case, $X$ is the variety of $m \times n$ rank 1 matrices, and a similar computations shows that the ML degree is 1 in these cases as well (see Example 1 of [12]).
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. For singular varieties, the ML degree is still related to the topology of the variety, but in a more subtle way from Whitney stratifications (Corollary 14). 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 taking the regular locus $X_{\text{reg}}$ as the first stratum, and then take the regular part of the singular locus of $X$, i.e., $(X_{\text{sing}})_{\text{reg}}$, and repeat 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 quite technical condition of Whitney regular, see [9, Page 37], [15, E3.7] for details.

Suppose $X_r$ is the subvariety of $\mathbb{P}^{mn}$ contained in the distinguished hyperplane $p_1 + \cdots + p_{mn} - p_s = 0$ parametrizing rank $\leq r$ matrices of size $m \times n$. Then, the naive stratification is indeed a Whitney stratification (see Proposition 3.1). Moreover, the stratification is given by the rank of the corresponding matrix.

We use Whitney stratifications in Corollary 14. We will see, 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 maximum likelihood degree of a projective variety. Sometimes, it is more convenient to restrict the projective variety to some affine torus and consider the notion of maximum likelihood degree of subvarieties of affine torus.

Let $Y$ be a closed irreducible subvariety of $(\mathbb{C}^*)^n$. Such a variety is called a very affine variety. Denote the coordinates of $(\mathbb{C}^*)^n$ by $z_1, z_2, \ldots, z_n$. The likelihood functions in the affine torus $(\mathbb{C}^*)^n$ are of the forms

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

**Definition 2.1.** Let $Y \subset (\mathbb{C}^*)^n$ be a very affine variety. Define the maximum likelihood degree of $Y$, denoted by $\text{MLdeg}_0(Y)$, to be the number of critical points of a likelihood function $l_u$ for general $u_1, u_2, \ldots, u_n$.

Fix an embedding of $\mathbb{P}^n \to \mathbb{P}^{n+1}$ by $[p_0 : p_1 : \ldots : p_n] \mapsto [p_0 : p_1 : \ldots : p_n : p_0 + p_1 + \cdots + p_n]$. 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 $p_0 + p_1 + \cdots + p_n - p_s = 0$.

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

$$(z_0, z_1, \ldots, z_n) \mapsto [z_0 : z_1 : \ldots : z_n : 1].$$
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 donate by $X^o$.

**Lemma 2.2.** 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.

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

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

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

As observed in [2], the maximum likelihood degree is equal to Gaussian degree defined by Franecki and Kapranov [7]. The main theorem of [7] relates the Gaussian degree with Euler characteristics. In this section, we will review their main result together with the explicit formula from [6] to compute characteristic cycles.

First, we follow the notation in [2]. Fix a positive integer $n$. 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 $G$. $T^*G$ has 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 $\gamma$ is contained in a nonempty Zariski open subset of $\mathfrak{g}^*$. This number is called the Gaussian degree of $\Delta$, and denoted by $\text{gdeg}(\Delta)$.

Let $Y \subset G$ be an irreducible closed subvariety of dimension $d$. Denote the conormal bundle of $Y_{\text{reg}}$ in $G$ by $T^*_Y G$, and denote its closure in $T^*G$ by $T^*_Y G$. Then $T^*_Y G$ is an irreducible conic Lagrangian subvariety of $T^*G$. Given any $\gamma \in \mathfrak{g}^*$, the left invariant 1-form corresponding to $\gamma$ degenerates at a point $P \in Y$ if and only if $T^*_Y G \cap \Omega_\gamma$ contains a point in $T^*_P G$. Thus, we have the following Lemma.

**Lemma 2.3 ([2]).**

$$\text{MLdeg}^o(Y) = \text{gdeg}(T^*_Y G).$$

Let $\mathcal{F}$ be a bounded constructible complex on $G$ and let $\text{CC}(\mathcal{F})$ be its characteristic cycle. Then $\text{CC}(\mathcal{F}) = \sum_j n_j(\Delta_j)$ 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 2.4 ([7]).**

$$\chi(G, \mathcal{F}) = \sum_j n_j \cdot \text{gdeg}(\Delta_j)$$
2.3. Euler obstructions. The Euler obstructions are defined to be the coefficients of some characteristic cycle decomposition (see equation (9)), and it is a theorem of Kashiwara that they can be computed as the Euler characteristic of some complex link (see Theorem 2.6).

Let \((S_1, S_2, \ldots, S_k)\) be a Whitney stratification of \(Y\) such that \(S_1 = Y_{\text{reg}}\). Let \(e_{j1}\) be the Euler obstruction of the pair \(S_j, S_1\), which measures the singular behavior of \(Y\) along \(S_j\). More precisely, \(e_{j1}\) are defined such that the following equality holds (See [6, 1.1] for more details).

\[
\text{CC}(C_{S_1}) = \sum_{0 \leq j \leq k} e_{j1}[T^*_j G].
\]

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

Since \(\chi(S_1) = \chi(G, C_{S_1})\), combining eqs. (7) to (9), we have the following corollary that expresses \(\chi(S_1)\) as a weighted sum of ML degrees.

**Corollary 2.5.** Let \((S_1, S_2, \ldots, S_k)\) be a Whitney stratification of \(Y\) such that \(S_1 = Y_{\text{reg}}\). Then,

\[
\chi(S_1) = \sum_{1 \leq j \leq k} e_{j1} \text{MLdeg}^0(\bar{S}_j)
\]

where \(\bar{S}_j\) denotes the closure of \(S_j\) in \(Y\) and \(e_{j1}\) be 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.

**Theorem 2.6 (Kashiwara\(^1\)).** Fix a point \(z \in Z\). Then

\[
e_{(mn)} = (-1)^{\dim Z + 1} \chi\left(B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)\right)
\]

where \(B\) is a ball of radius \(\delta\) in \(\mathbb{P}^{mn-1}\) centered at \(z\), \(\phi\) is a general linear function defined on a normal slice \(N\) of \(Z\) at \(z\) and \(0 < |\epsilon| \ll \delta \ll 1\).

By the formula above, every Euler obstruction we consider is computable.

**Example 2.7.** Consider the variety 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,

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

The Whitney stratification of \(X^o\) consists of \(S_1\), the regular points of \(X^o\), and \(S_2\), the singular point of \(X^o\), which are the rank 1 matrices. By Corollary 2.5, we have

\[
\chi(S_1) = e_{11} \text{MLdeg}^0(\bar{S}_1) + e_{21} \text{MLdeg}^0(\bar{S}_2).
\]

With this equation, we determine the ML degree of \(X\). The rank 1 matrices are known to have ML Degree one, so \(\text{MLdeg}^0(\bar{S}_2) = 1\). The Euler obstruction \(e_{21}\) is equal to the Euler characteristic of some complex link, up to a sign (see Theorem 2.6 for the precise formula). In this case, the complex link turns out to be homeomorphic to a vector bundle over \(\mathbb{P}^1\) (see [6, Theorem 1.1], see also [4, Page 100] and [8, 8.1]).

\(^1\)Here the formula is in the form of [6, Theorem 1.1], see also [4, Page 100] and [8, 8.1].
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 $\chi(S_1)$. In fact, $\chi(S_1) = -12$. Therefore, (11) implies that $\text{MLdeg}(X) = 10$ concluding the example.

In Example 2.7, we used Corollary 2.5 and topological computations to determine the ML degree of a singular variety. In the next section we will again use Corollary 2.5 and topological computations to determine ML degrees.

3. The ML degree for rank 2 matrices

To $\mathbb{P}^{mn}$ we associate the coordinates $p_{11}, \ldots, p_{1n}, \ldots, p_{m1}, \ldots, p_{mn}, p_s$. Let $X_{mn}$ denote the variety defined by $p_{11} + \cdots + p_{mn} - p_s = 0$ and the vanishing of the $3 \times 3$ minors of the matrix

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

(12)

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 vanishing of the $2 \times 2$ minors of the matrix (12). Then $Z_{mn}$ is the singular locus of $X_{mn}$ for $m, n \geq 3$.

We will make topological computations to determine the ML degree of $X_{mn}$. Summarizing, Proposition 3.1 gives a Whitney decomposition of $X_{mn}$ and determines the Euler obstructions for this stratification. Using these computations, we reduce our problem to determining a single Euler characteristic $\chi(X_{mn} \setminus Z_{mn})$ by Corollary 3.3. Next, Theorem 3.4 provides a closed form expression of $\chi(X_{mn} \setminus Z_{mn})$, for fixed $m$, in terms of the elements of the finite sequence $\Lambda_m$ of integers. We conclude this section by computing $\Lambda_m$ 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 $\lambda_1, \lambda_2$ of $\Lambda_3$, 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 $e_{21}$ for $X_{mn}$ and we have

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

Proposition 3.1. The stratification $X_{mn} = Y_{mn} \cup Z_{mn}$ is a Whitney stratification of $X_{mn}$. Moreover, if the Euler obstruction of the pair of strata $(Z_{mn}, Y_{mn})$ is denoted by $e_{(mn)}$, then

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

(13)

Proof. When $m = 2$ or $n = 2$, $X_{mn} = \mathbb{P}^{mn-1}$ and $Z_{mn} \subset X_{mn}$ is a smooth subvariety. Thus, the first part of the proposition follows. Moreover, it follows from definition that $e_{(mn)} = (-1)^{\dim Z_{mn}+1}$. Since $\dim Z_{mn} = m + n - 2$, the second part of the proposition follows. Therefore, we can assume $m, n \geq 3$ and $m \leq n$ without loss of generality.
Notice that there is a left $Gl(m, \mathbb{C})$ action and a right $Gl(n, \mathbb{C})$ action on $X_{mn}$ that both preserve $Z_{mn}$. The total action by $Gl(m, \mathbb{C}) \times Gl(n, \mathbb{C})$ is transitive on $Z_{mn}$. Since $Z_{mn}$ is the singular locus of $X_{mn}$, near a general point of $Z_{mn}$, $Z_{mn} \subset X_{mn}$ has to be Whitney regular. Now by the presence of the transitive action, $Z_{mn} \subset X_{mn}$ is Whitney regular at every point of $Z_{mn}$. Since $Y_{mn} = X_{mn} \setminus Z_{mn}$, it follows $(Y_{mn}, Z_{mn})$ is a Whitney stratification of $X_{mn}$.

Next, we will compute the Euler obstructions using Theorem 2.6. Consider $X_{mn}$ as a subvariety of $\mathbb{P}^{mn-1}$, the projective space of all $m \times n$ matrices. In the next theorem and its proof, we simply write $X$ and $Z$ instead of $X_{mn}$ and $Z_{mn}$ and use the notion 'normal slice'. A normal slice of a variety $Z$ at the point $z$ is a general linear space with complementary dimension to $Z$ containing the point $z$.

One can easily compute that $\dim Z = m + n - 2$. Since $B \cap (X \setminus Z) \cap \phi^{-1}(\epsilon)$ is an even-dimensional manifold, its Euler characteristics with or without compact support are equal by Poincare 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 the link is homotopy equivalent to $\mathbb{P}^{m-2}$.

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

**Proof.** First, we give a concrete description of the normal slice $N$. Notice that $X \subset \mathbb{P}^{mn}$ is contained in the distinguished hyperplane $p_{11} + \cdots + p_{mn} - p_s = 0$. In this proof, we will consider $X$ as a subvariety of $\mathbb{P}^{nm-1}$ with homogeneous coordinates $p_{11}, \ldots, p_{mn}$. 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 $O$. Note, we define a projection $\pi : U_{11} \to Z \cap U_{11}$ by $(a_{ij}) \mapsto (b_{ij})$, where $b_{ij} = a_{11} \cdot a_{ij}$ and $a_{11} = 1$. Then $U_{11}$ becomes a vector bundle over $Z \cap U_{11}$ via $i\pi : U_{11} \to Z \cap U_{11}$. The preimage of $O$ is the vector space parametrized by $a_{ij}$ with $2 \leq i \leq m, 2 \leq j \leq n$.

In terms of matrices, we can think of $\pi$ as the following map

$$
\begin{pmatrix}
1 & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
\mapsto
\begin{pmatrix}
1 & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
$$

and we think of the preimage of $O$ as

$$
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
0 & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
= \pi^{-1}
\begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}.
$$

By the above construction, we can take the normal slice $N$ at $O$ to be the fiber $\pi^{-1}(O)$. The intersection $N \cap X$ is clearly isomorphic to the affine variety $\{(a_{ij})_{2 \leq i \leq m, 2 \leq j \leq n} | \text{rank } a_{ij} \leq 1\}$. Thus, we can define a map $\rho : N \cap (X \setminus Z) \to \mathbb{P}^{m-2}$ 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
Theorem 3.4. Fix \( m \) to be an integer greater than two. Then, there exists a sequence, denoted \( \lambda \), the Euler characteristic \( \chi \) (15) calculating Euler characteristics and ML degrees.

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

Recall that \( X^o_{mn} \) 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 \( \Lambda_m \), of integers \( \lambda_1, \lambda_2, \ldots, \lambda_{m-1} \) such that

\[
\chi(Y^o_{mn}) = (-1)^{n-1} \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} - \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} \cdot i^{n-1}, \quad \text{for } n \geq 2.
\]

Before proving Theorem 15, we quote a hyperplane arrangement result, which follows immediately from the theorem of Orlik-Solomon (see e.g. [19] Theorem 5.90).

Lemma 3.5. Let \( L_1, \cdots, L_r \) be distinct hyperplanes in \( \mathbb{C}^s \). Suppose they are in general position, that is the intersection of any \( t \) hyperplanes from \( \{L_1, \cdots, L_r\} \) has codimension \( t \), for any \( 1 \leq t \leq s \). Denote the complement of \( L_1 \cup \cdots \cup L_r \) 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 proceeding lemma to compute Euler characteristics of stratum of the following stratifications. Throughout, we assume that \( m \) is fixed. Let \( U_n \) denote the rank 2 matrices with nonzero coordinates whose entries sum to 1, i.e.,

\[
U_n := \left\{ (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \mid a_{ij} \in \mathbb{C}^s, \sum_{1 \leq i \leq m, 1 \leq j \leq n} a_{ij} = 1, \text{ rank}(a_{ij}) = 2 \right\}.
\]

A stratification of \( U_n \) is given by the number of columns summing to zero a matrix has. Defining \( U_n^{(l)} \) below,

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

yields the stratification:

\[
U_n = U_n^{(0)} \sqcup \cdots \sqcup U_n^{(n)},
\]
where each $U_n^{(l)}$ is a locally closed subvariety of $U_n$.

Let $U_n'$ denote the set of rank 2 matrices with nonzero column sums, i.e.,

$$U_n' := \left\{ (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \in U_n \mid \sum_{1 \leq i \leq m} a_{ij} \neq 0 \text{ for each } 1 \leq j \leq n \right\}.$$

Note by definition, we have $U_n' = U_n^{(0)}$ and

$$U_n \cong Y_{mn}^o.$$

The following lemmas will show $\chi(U_n') = \chi(U_n)$ by proving $\chi(U_n^{(l)}) = 0$ for $l \geq 1$.

**Lemma 3.6.**

$$\chi(U_n) = \chi(U_n').$$

**Proof of Lemma.** We define a $C^*$ action on $U_n$ by putting $t \cdot (a_{ij}) = a'_{ij}$, where

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

It is straightforward to check the action preserves each $U_n^{(l)}$ and the action is transitive and continuous on $U_n^{(l)}$ for any $l \geq 1$. Therefore, $\chi(U_n^{(l)}) = 0$ for any $l \geq 1$, and hence $\chi(U_n) = \chi(U_n^{(0)}) = \chi(U_n')$. \hfill $\square$

A column sum of $U_n'$ can be any element of $C^*$. Now we consider a subset of $U_n'$ 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.,

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

In Lemma 3.7, we express the Euler characteristic of $U_n'$ in terms of $\chi(V_n)$.

**Lemma 3.7.**

$$\chi(U_n') = (-1)^{n-1}\chi(V_n).$$

**Proof of Lemma.** Let $T_n = \{(t_{ij})_{1 \leq j \leq n} \in (C^*)^{n} \mid \sum_{j} t_{ij} = 1\}$. Define a map $F : T_n \times V_n \rightarrow U_n'$ by putting $a_{ij} = t_{ij}b_{ij}$; 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, $\chi(U_n') = \chi(T_n) \cdot \chi(V_n)$. $T_n$ can be considered as $C^{n-1}$ removing $n$ hyperplanes in general position. By Lemma 3.5, $\chi(T_n) = (-1)^{n-1}$, and hence $\chi(U_n') = (-1)^{n-1}\chi(V_n)$. \hfill $\square$

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 $V_n^{(l)}$ below,

$$V_n^{(l)} := \left\{ (b_{ij}) \in V_n \mid \text{the column vectors satisfy } (b_{i1}) = (b_{i2}) = \cdots = (b_{il}) \neq (b_{i(l+1)}) \right\},$$
yields the stratification of $V_n$ by locally closed subvarieties:

$$V_n = V^{(1)}_n \sqcup V^{(2)}_n \sqcup \cdots \sqcup V^{(n-1)}_n.$$  

Hence

$$\chi(V_n) = \chi(V^{(1)}_n) + \chi(V^{(2)}_n) + \cdots + \chi(V^{(n-1)}_n).$$

We use $W_i$ defined below to stratify $V_n$.

$$W_n := \{(b_{ij}) \in V_n| \text{ the first two column vectors } (b_{i1}) \text{ and } (b_{i2}) \text{ are linearly independent}\}.$$

Then, we have the isomorphism $V^{(i)}_n \approx W_{n-l+1}$ by the following map,

$$\begin{bmatrix} b_{i1} & \ldots & b_{i(n+1)} \\ \vdots & \ddots & \vdots \\ b_{i1} & \ldots & b_{i(n+1)} \end{bmatrix} \mapsto [b_{i1} \ b_{i(l+1)} \ldots \ b_{in}].$$

Therefore,

$$\chi(V_n) = \chi(W_2) + \chi(W_3) + \cdots + \chi(W_n).$$

For any $l \geq 2$, we can define a map $\pi_l : W_l \to W_2$ by taking the first two column. Thus, we can consider all $W_l$ as varieties over $W_2$.

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

$$W_l \cong W_3 \times_{w_2} W_3 \times_{w_2} \cdots \times_{w_2} W_3$$

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, take any point $x \in W_2$ the fiber of $\pi_l : W_l \to W_2$ over $x$ is equal to the $(l-2)$-th power of the fiber of $\pi_3 : W_3 \to W_2$ over $x$.

**Proof of Lemma.** Given $l-2$ elements in $W_3$. Suppose they all belong to the same fiber of $\pi_3 : W_3 \to W_2$. 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 column of each matrix, and put them after the same first two columns. Thus we obtain a $m \times l$ matrix, whose rank is still 2. In this way, we obtain a map $W_3 \times_{w_2} W_3 \times_{w_2} \cdots \times_{w_2} W_3 \to W_l$,

$$\begin{bmatrix} b_{i1} & b_{i2} & b_{i3} \\ b_{i1} & b_{i2} & b_{i4} \\ \vdots & \vdots & \vdots \\ b_{i1} & b_{i2} & b_{il} \end{bmatrix}, \ldots, \begin{bmatrix} b_{i1} & b_{i2} & b_{i3} \\ b_{i1} & b_{i2} & b_{i4} \end{bmatrix} \mapsto \begin{bmatrix} b_{i1} & b_{i2} & b_{i3} \ldots & b_{il} \end{bmatrix},$$

which is clearly an isomorphism.\[\square\]

Given any point $x \in W_2$, we denote the fiber of $\pi_3 : W_3 \to W_2$ by $F_x$. Since 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

$$W_2^k := \{x \in W_2|\chi(F_x) = -k\}.$$  

Then to compute $\chi(W_l)$, it suffices to compute $\chi(W_2^k)$ for all $k$. Let $\Lambda_m$ denote the sequence $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$, where $\lambda_k$ are defined by

$$\lambda_k := \chi(W_2^k) \text{ for } 0 \leq k \leq m-1.$$
As we will soon see in the proofs, we can consider $F_x$ as the complement of a point arrangement in $C$. The arrangement is parametrized by the point $x \in W_2$. In other words, $W_2$ is naturally a parameter space of point arrangement in $C$. 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 to generalize 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 $W_2 \to \mathbb{Z}$ defined by $x \mapsto \chi(F_x)$ is a semi-continuous function. In other words, for any integer $k$, $\{x \in W_2 | \chi(F_x) \geq k\}$ is a closed algebraic subset of $W_2$. In particular, the subsets $\{x \in W_2 | \chi(F_x) = k\}, 0 \geq k \geq -(m - 1)$, give a stratification of $W_2$ into locally closed subsets.

**Proof of Lemma.** 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 $x = (b_{ij}) \in W_2$. By definition, the fiber $F_x$ of $\pi_3 : W_3 \to W_2$ is equal to the following:

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

Since $\sum_{1 \leq i \leq m} b_{ij} = 1$, $j = 1, 2, 3$, for any $(b_{i3}) \in F_x$ there exists $b \in \mathbb{C}$ such that $(b_{i3}) = b \cdot (b_{i1}) + (1 - b) \cdot (b_{i2})$. The condition that $b_{i3} \neq 0$ is equivalent to $b_{i2} + b \cdot (b_{i1} - b_{i2}) \neq 0$. Therefore,

$$F_x \cong \mathbb{C} \setminus \left\{ -\frac{b_{i2}}{b_{i1} - b_{i2}} \mid 1 \leq i \leq m \text{ such that } b_{i1} - b_{i2} \neq 0 \right\}.$$

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 $W_2$. \qed

**Example 3.10.** When $m = 3$, the Euler characteristic $\chi(F_x)$ is in $\{0, -1, -2\}$. More precisely, we can think of each $F_x$ to be the complement of 3 points in $C$, where we allow the points to overlap or go to infinity, but we do not allow all three points collapsing to one point. The closed algebraic subset $\{x \in W_2 | \chi(F_x) = 0\}$ consists of 3 irreducible components of codimension 2.
One of them is equal to
\[
\left\{ \begin{pmatrix} 1 & 1 \\ \alpha & \beta \\ -\alpha & -\beta \end{pmatrix} \right| \alpha, \beta \in \mathbb{C}^*, \alpha \neq \beta \right\}
\]
and the other two are obtained from the first one by column permutations. When \( x \in W_2 \) is in contained in the subset (21), the first point of the three-point configuration goes to infinity, and the other two points overlap. Thus, the corresponding fiber \( F_x \) is isomorphic to \( \mathbb{C} \) minus a point.

The closed algebraic subset \( \{ x \in W_2|\chi(F_x) \geq -1 \} \) consists of 6 irreducible components of codimension one. Two of them are equal to
\[
\left\{ \begin{pmatrix} t & t \\ \alpha & \beta \\ 1-t-\alpha & 1-t-\beta \end{pmatrix} \right| \alpha, \beta, t \in \mathbb{C}^* \right\}
\]
and
\[
\left\{ \begin{pmatrix} 1-\alpha-s\alpha & 1-\beta-s\beta \\ \alpha & \beta \\ sa & s\beta \end{pmatrix} \right| \alpha, \beta, s \in \mathbb{C}^* \right\}.
\]
The other 4 components are obtained from the above two components by column permutations. Notice that the listed two components are invariant under the permutation of the second and third columns. When \( x \in W_2 \) belongs to the set (22), the first point of the three-point configuration goes to infinity. When \( x \in W_2 \) belongs to the set (23), the second and the third point overlap. Therefore, in any of these cases, \( F_x \) is isomorphic to \( \mathbb{C} \) minus two points.

Proof of Theorem 3.4 . Thus far, we have shown with eqs. (16) to (19), the following relations:
\[
\chi(Y_{mn}^0) = \chi(U_n) = \chi(U_n^i) = (-1)^{n-1} \chi(V_n) = (-1)^{n-1} \sum_{2 \leq l \leq n} \chi(W_l).
\]
Recall that \( W_k^2 = \{ x \in W_2|\chi(F_x) = -k \} \) and by Lemma 3.9 we have the following stratification,
\[
W_2 = W_2^0 \cup W_2^1 \cup \ldots \cup W_2^{m-1}.
\]
Moreover, \( W_k^2 \) are locally closed algebraic subsets of \( W_2 \). The projection \( \pi_3 : W_3 \to W_2 \) induces a fiber bundle over each \( W_k^2 \), and by definition of \( W_k^2 \), 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 \( W_k^2 \). Thus, restricting \( \pi_l : W_l \to W_2 \) to \( W_k^2 \), the induced bundle’s fiber has Euler characteristic \((-k)^{l-2} \).

Then,
\[
\chi(\pi_l^{-1}(W_k^2)) = \lambda_k \cdot (-k)^{l-2}.
\]
Since $W^2_k$ are locally closed algebraic subsets of $W^2$, $\pi^{-1}(W^2_k)$ are locally closed algebraic subsets of $W_l$. Therefore, the additivity of Euler characteristic implies the following,

$$\chi(W_l) = \sum_{0 \leq k \leq m-1} \chi(\pi^{-1}(W^2_k)) = \sum_{0 \leq k \leq m-1} \lambda_k \cdot (-k)^{l-2}. \quad (27)$$

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

By equations (24), (27) and Proposition 3.11, we have the following equalities.

$$\chi(Y_{mn}^o) = (-1)^{n-1} \sum_{2 \leq l \leq n} \left( \sum_{0 \leq k \leq m-1} \lambda_k \cdot (-k)^{l-2} \right) = (-1)^{n-1} \sum_{0 \leq k \leq m-1} \lambda_k \cdot \frac{1 - (-k)^{n-1}}{1 - (-k)}. \quad (29)$$

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 $\lambda_0$ of $\Lambda_m$ 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 maximum likelihood degree of $X_{3n}$ is given by the following formula.

$$\text{MLdeg}(X_{3n}) = 2^{n+1} - 6. \quad (28)$$

**Proof.** In [13], 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 $\lambda_1 + \lambda_2 = 0$ and $\lambda_2 = 12$ by Theorem 3.4. □

The take away 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 Grassmanians, with a combination of applied algebraic geometry and topological arguments.

**Proof of Proposition 3.11.** Recall that $\lambda_0 = \chi(W^0_2)$. By definition, $W^0_2$ consists of those $(b_{ij})_{1 \leq i \leq m, j=1,2}$ in $W^2$ such that the cardinality of the set $\{b_{ij} / b_{ij} \neq b_{i2} \}$ is equal to 1.

Notice that for $(b_{ij})_{1 \leq i \leq m, j=1,2} \in W^0_2$,

$$\sum_{1 \leq i \leq m \atop b_{i1} \neq b_{i2}} b_{i1} = \sum_{1 \leq i \leq m \atop b_{i2} \neq b_{i1}} b_{i1} = 0. \quad (29)$$

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

$$b'_{ij} = \begin{cases} b_{ij} & \text{if } b_{i1} = b_{i2} \\ t \times b_{ij} & \text{otherwise}. \end{cases} \quad (30)$$

Now, it is straightforward to check this $\mathbb{C}^*$ action preserves $W^0_2$ and the action is transitive on $W^0_2$. Therefore, $\chi(W^0_2) = 0$. □
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 (15). 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 \(-\text{MLdeg}(X_{mn}) + (-1)^{m+n-1}(\min\{m, n\} - 1)\). 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
\[
\chi(Y_{mn}^o) = c_1 \chi(Y_{m(n-1)}^o) + c_2 \chi(Y_{m(n-2)}^o) + \cdots + c_m \chi(Y_{m(m-n)}^o).
\]

**Proof.** By Theorem 3.4, we have
\[
\chi(Y_{mn}^o) = (-1)^{n-1} \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} - \sum_{1 \leq i \leq m-1} \frac{\lambda_i}{i+1} \cdot t^{n-1}
\]
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 \(-1, 1, \ldots, m - 1\), i.e. \(t^m - c_1 t^{m-1} - \cdots - c_m = (t + 1) \prod_{r=1}^{m-1} (t - r)\). \(\square\)

With these recurrences we determine the following table of ML degrees:

<table>
<thead>
<tr>
<th>(n)</th>
<th>(m = 2)</th>
<th>(m = 3)</th>
<th>(m = 4)</th>
<th>(m = 5)</th>
<th>(m = 6)</th>
<th>(m = 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m)</td>
<td>1</td>
<td>10</td>
<td>191</td>
<td>6776</td>
<td>378477</td>
<td>3030576</td>
</tr>
<tr>
<td>(m + 1)</td>
<td>1</td>
<td>26</td>
<td>843</td>
<td>40924</td>
<td>2865245</td>
<td>27474099</td>
</tr>
<tr>
<td>(m + 2)</td>
<td>1</td>
<td>58</td>
<td>3119</td>
<td>212936</td>
<td>19177197</td>
<td>2244706374</td>
</tr>
<tr>
<td>(m + 3)</td>
<td>1</td>
<td>122</td>
<td>10587</td>
<td>1015564</td>
<td>118430045</td>
<td>17048729886</td>
</tr>
<tr>
<td>(m + 4)</td>
<td>1</td>
<td>250</td>
<td>34271</td>
<td>4586456</td>
<td>692277357</td>
<td>122818757286</td>
</tr>
<tr>
<td>(m + 5)</td>
<td>1</td>
<td>506</td>
<td>107883</td>
<td>19984444</td>
<td>3892815965</td>
<td>850742384190</td>
</tr>
<tr>
<td>(m + 6)</td>
<td>1</td>
<td>1018</td>
<td>333839</td>
<td>84986216</td>
<td>21284701677</td>
<td>5720543812614</td>
</tr>
</tbody>
</table>

**Table 1.** ML degrees of \(X_{mn}\).

Note that our methods are not limited to Table 4.1. 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 \(\Lambda_m\) for small \(m\) that can be extended to arbitrary \(m\).\(^2\)

\(^2\)Macaulay2 implementation at http://home.uchicago.edu/~joisro/quickLinks/ECML16/MLDegreeMixtures.m2.
Corollary 4.2. For fixed $m = 2, 3, \ldots, 7$, the closed form expressions for $\text{MLdeg}(X_{mm})$ with $m \leq n$ are below:
\[
\begin{align*}
\text{MLdeg}(X_{22}) & = 1 \\
\text{MLdeg}(X_{33}) & = \left(\frac{-12}{2} \cdot 1^{n-1} + \frac{12}{3} \cdot 2^{n-1}\right) \\
\text{MLdeg}(X_{44}) & = \left(\frac{30}{2} \cdot 1^{n-1} + \frac{-120}{3} \cdot 2^{n-1} + \frac{72}{4} \cdot 3^{n-1}\right) \\
\text{MLdeg}(X_{55}) & = \left(\frac{-180}{2} \cdot 1^{n-1} + \frac{780}{3} \cdot 2^{n-1} + \frac{-1080}{4} \cdot 3^{n-1} + \frac{480}{5} \cdot 4^{n-1}\right) \\
\text{MLdeg}(X_{66}) & = \left(\frac{600}{2} \cdot 1^{n-1} + \frac{-2000}{3} \cdot 2^{n-1} + \frac{10800}{4} \cdot 3^{n-1} + \frac{-600}{5} \cdot 4^{n-1} + \frac{30240}{6} \cdot 5^{n-1}\right) \\
\text{MLdeg}(X_{77}) & = \left(\frac{-1932}{2} \cdot 1^{n-1} + \frac{-20412}{3} \cdot 2^{n-1} + \frac{-75600}{4} \cdot 3^{n-1} + \frac{-127680}{5} \cdot 4^{n-1} + \frac{-100800}{6} \cdot 5^{n-1} + \frac{30240}{7} \cdot 6^{n-1}\right).
\end{align*}
\]

Proof. We find these formulas using an inductive procedure to determine $\Lambda_m$ from $\Lambda_{m-1}$.

By Theorem 3.4, we have (15) gives us the following $m - 1$ relations with $n = 2, 3, \ldots, m$:
\[
\begin{bmatrix}
\text{MLdeg}(X_{m2}) \\
\text{MLdeg}(X_{m3}) \\
\vdots \\
\text{MLdeg}(X_{mn})
\end{bmatrix} + (-1)^n 
\begin{bmatrix}
1 \\
n-2 \\
\vdots \\
(-1)^n (m-1)
\end{bmatrix} = 
\begin{bmatrix}
1^1 & 2^1 & \ldots & (m-1)_1^1 \\
1^2 & 2^2 & \ldots & (m-1)_2^2 \\
\vdots & \vdots & \ddots & \vdots \\
1^{m-1} & 2^{m-1} & \ldots & (m-1)_{m-1}^{m-1}
\end{bmatrix} - 
\begin{bmatrix}
(-1)^1 & (-1)^1 & \ldots & (-1)^1 \\
(-1)^2 & (-1)^2 & \ldots & (-1)^2 \\
\vdots & \vdots & \ddots & \vdots \\
(-1)^{m-1} & (-1)^{m-1} & \ldots & (-1)^{m-1}
\end{bmatrix} 
\begin{bmatrix}
\lambda_1/2 \\
\lambda_3/3 \\
\vdots \\
\lambda_{m-1}/m
\end{bmatrix}
\]

For fixed $m$, this system of linear equations has $2m - 2$ unknowns: $\text{MLdeg}(X_{mj})$ for $j = 2, \ldots, m$ and $\lambda_1, \ldots, \lambda_{m-1}$ of $\Lambda_m$. By induction, we may assume we know $\Lambda_{m-1}$. The $\Lambda_m$ gives us a closed form expression for the ML degrees of $X_{(m-1)}$ 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 $\lambda_{m-1}$ of $\Lambda_m$ equals $(m-1) \cdot m!$. Substituting this value as well, we have a linear system of $m - 1$ equations in $m - 1$ unknowns: $\text{MLdeg}(X_{mm}), \lambda_1, \lambda_2, \ldots, \lambda_{m-2}$.

A simple linear algebra argument shows that there exists a unique solution of the system yielding each $\lambda_i$ of $\Lambda_m$ as well as $\text{MLdeg}(X_{mm})$.

Using the inductive procedure described above, we determined the following table of $\Lambda_m$ to yield the closed form expressions we desired.

<table>
<thead>
<tr>
<th>$\Lambda_i$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\lambda_3$</th>
<th>$\lambda_4$</th>
<th>$\lambda_5$</th>
<th>$\lambda_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda_2$</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda_3$</td>
<td>-12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda_4$</td>
<td>50</td>
<td>-120</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda_5$</td>
<td>-180</td>
<td>780</td>
<td>-1080</td>
<td>480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Lambda_6$</td>
<td>602</td>
<td>-4200</td>
<td>10080</td>
<td>-10080</td>
<td>3600</td>
<td></td>
</tr>
<tr>
<td>$\Lambda_7$</td>
<td>-1932</td>
<td>20412</td>
<td>-75600</td>
<td>127680</td>
<td>-100800</td>
<td>30240</td>
</tr>
</tbody>
</table>

Table 2. The $\lambda_i$ of $\Lambda_m$.

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 [12], 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 for weighted independence models deform a variety with ML degree
one to a variety with large ML degree. This deformation deforms the maximum likelihood 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 \( \lambda_1 \) of \( \Lambda_m \) be defined as above. Then \( \lambda_{m-1} \) of \( \Lambda_m \) equals \( (m - 1) \cdot m! \).

**Proof.** Recall that for \( \lambda_{m-1} \) of \( \Lambda_m \) equals \( \chi(W_2^{m-1}) \). By definition, \( W_2^{m-1} \) consists of all \( (b_{ij})_{1 \leq i \leq m, j=1,2} \in W_2 \) 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 \( (C^* \setminus \{1\})^m \) corresponding to \( m \) distinct numbers. Then there is a natural map \( \pi : W_2^{m-1} \rightarrow B_m \), defined by \( (b_{ij}) \mapsto (b_{11}/b_{12}, \ldots, b_{m1}/b_{m2}) \). The map is surjective. Moreover, one can easily check that under the map \( \pi \), \( W_2^{m-1} \) is a fiber bundle over \( B \), whose fiber is isomorphic to the complement of \( m \) hyperplanes in \( C^{m-2} \) in 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 \( C^* \setminus \{m \text{ distinct points}\} \). Therefore,

\[
\chi(W_2^{m-1}) = (-1)^{m-2}(m - 1) \cdot (-1)^m m! = (m - 1) \cdot m!.
\]

\( \square \)

5. Conclusion and additional questions

We have developed the topological tools to determine the ML degree of singular models. We proved a closed form expression for \( 3 \times n \) matrices with rank 2 conjectured by \[12\]. 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 algebra 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 \[5\]. 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 \[21\].

Another question concerns the boundary components of statistical models as described in \[18\] 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 great 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 \[11\]. Here, entries of the data \( u \) are degenerated to zero and some of the 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.
6. Acknowledgements

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 2.6.

References

JOSE ISRAEL RODRIGUEZ, THE UNIVERSITY OF CHICAGO, DEPT. OF STATISTICS, 5734 S. UNIVERSITY AVE. CHICAGO, IL 60637
E-mail address: joisro@uchicago.edu
URL: http://home.uchicago.edu/~joisro

BOTONG WANG, UNIVERSITY OF WISCONSIN-MADISON, DEPARTMENT OF MATHEMATICS, VAN VLECK HALL, 480 LINCOLN DRIVE, MADISON, WI
E-mail address: wang@math.wisc.edu
URL: http://www.math.wisc.edu/~wang/