Maximum likelihood geometry in the presence of data zeros

Elizabeth Gross∗ and Jose Israel Rodriguez†

January 21, 2014

Abstract

Given a statistical model, the maximum likelihood degree is the number of complex solutions to the likelihood equations for generic data. We consider discrete algebraic statistical models and study the solutions to the likelihood equations when the data contain zeros and are no longer generic. Focusing on sampling and model zeros, we show that, in these cases, the solutions to the likelihood equations are contained in a previously studied variety, the likelihood correspondence. The number of these solutions give a lower bound on the ML degree, and the problem of finding critical points to the likelihood function can be partitioned into smaller and computationally easier problems involving sampling and model zeros. In fact, using numerical algebraic geometry and by considering data with sampling and model zeros, we gain traction on computing the ML degree for multi-way tables and provide an algorithm for finding critical points of the likelihood function using numerical algebraic geometry.

1 Introduction

The method of maximum likelihood estimation for a statistical model \( \mathcal{M} \) and an observed data vector \( u \in \mathbb{R}^{n+1} \) involves maximizing the likelihood function \( l_u \) over all distributions in \( \mathcal{M} \). This involves understanding the zero-set of a system of equations, and, thus, when the models of interest are algebraic, the process lends itself to investigation using algebraic geometry. In fact, likelihood geometry has been studied in a series of papers in the field of algebraic statistics beginning with [CHKS06] and [HKS05]. Subsequent papers include [BHR07], [HS10], [GDPT12], [HRSP12], [Uh12], [RPF13], and [H13] and cover both discrete and continuous models. In this paper, we look at discrete models and the case where the observed data vector contains zero entries.

In [HKS05], Hoşten, Khetan, and Sturmfels introduce the likelihood locus and its associated incidence variety for discrete statistical models. In [HS13], Hub and Sturmfels study this incidence variety further under the name of the likelihood correspondence. Given a discrete algebraic statistical model with sample space of size

∗Department of Mathematics, North Carolina State University, Raleigh, NC; eagross@ncsu.edu. This work was partially supported by NSF award DMS–1304167.

†Department of Mathematics, University of California at Berkeley, Berkeley, CA 94720; jo.ro@berkeley.edu. The second author is supported by the US National Science Foundation DMS-0943745.
Given positive integer data $u$ of $p$ points in $pr$ under the degree of a generic fiber of $M$ restricted to $pr$ under $pr$ (maximum likelihood degree) of $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Forels, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$. Els, which are defined by the vanishing of polynomial equations restricted to the $u$ coordinates $X$.

A statistical model $M$ is a subset of the probability simplex

$$\Delta_n = \{ (p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1} \mid \sum_{i=0}^{n} p_i = 1 \text{ and } p_i \geq 0 \text{ for } i = 0, 1, \ldots n \}. \quad \text{(1)}$$

Given positive integer data $u \in \mathbb{Z}_{\geq 0}^{n+1}$, the maximum likelihood estimation problem is to determine $\hat{p} \in M$ that maximizes the likelihood function

$$l_u = p_0^{u_0} p_1^{u_1} \cdots p_n^{u_n}$$

restricted to $M$. The point $\hat{p} \in M$ is called the maximum likelihood estimator, or MLE. The family of models we are interested in are algebraic statistical models, which are defined by the vanishing of polynomial equations restricted to the probability simplex.

To use algebraic methods, we consider points of $M \subset \mathbb{R}^{n+1}$ as representatives of points in $\mathbb{P}^n$ and study the Zariski closure $\overline{M} = X \subset \mathbb{P}^n$. This makes the problem easier by relaxing the nonnegative and real constraints, which allows us to obtain an understanding about the number of possible modes of the likelihood surface. There are subtleties when performing this relaxation as mentioned for example in Section 5 of [HRST2].

Let $p_+ := p_0 + p_1 + \cdots + p_n$ and $\mathcal{H}_u$ be the set of points where $p_+ p_0 p_1 \cdots p_n$ equals zero. With algebraic methods, our goal is to determine all complex critical points of $L_u := l_u/p_+^{u_+}$ when restricted to $X_{reg} \setminus \mathcal{H}_u \subset \mathbb{P}^n$, where $X_{reg}$ is the set of regular points of $X$. We work with $L_u$ since it is a rational function of degree zero and thus a function on $\mathbb{P}^n$ (see [DSS09], §2.2).

A point $p \in X_{reg}$ is said to be a critical point if the gradient of $L_u (p)$ is orthogonal to the tangent space of $X$ at $p$, that is

$$\nabla L_u (p) \perp T_p X. \quad \text{(2)}$$

If the maximum likelihood estimator $\hat{p}$ for the data vector $u$ is in the interior of $M$, then $\hat{p}$ will be a critical point of $L_u$ over $X$. By determining the critical points of $L_u$ on $X$, we find all local maxima of $l_u$ on $M$.

When the data vector $u$ contains zero entries, each zero entry is called either a sampling zero or a structural zero in the statistics literature. Considering $u$ as a flattened contingency table, a sampling zero at $u_i$ occurs when no observations fall into cell $i$ even though $p_i$ is nonzero. A structural zero occurs at $u_i$ when the probability of an observation falling into cell $i$ is zero. Structural and sampling zeros occur commonly in practice, for example, in large sparse data sets (for more on sampling and structural zeros see [BFH75], §5.1.1).

The terms “sampling zero” and “structural zero” are denotationally about contingency tables, but they also carry implications about $X$ as well. For example, the term “structural zero” connotes that maximum likelihood estimation should proceed
over a projection of $X$ (see [Rap06]). Due to this secondary definition imparted to the term “structural zero,” and in view of the fact that this study is concerned with the intersection of $X$ with the hyperplane $p_i = 0$ as opposed to the projection of $X$, we introduce the definition of a model zero.

**Definition 1 (Model zeros).** Given a model $\mathcal{M}$ with $\overline{\mathcal{M}} = X \subset \mathbb{P}^n$ and data vector $u$ with $u_i = 0$, a model zero at cell $i$ is a zero such that the maximum likelihood estimator $\hat{p}$ for $u$ is a critical point of $L_u$ over $X \cap \{p_i = 0\}$.

**Remark 2.** From the remainder of the paper, we will use “structural zero” to mean a zero at cell $i$ such that 1) $p_i = 0$, and 2) maximum likelihood estimation proceeds over the projection of $X$ onto all coordinates except the $i$th coordinate.

In this paper we explore in depth the algebraic considerations of maximum likelihood estimation when the data contains sampling and model zeros. In the main theorem of Section 3, Theorem 11, we show how solutions to the maximum likelihood estimation problems for data with zeros on $X$ are contained in the likelihood correspondence of $X$. This result gives statistical meaning to the likelihood correspondence when $u_i$ is equal to zero and has interesting theoretical and computational implications. On the theoretical side, we can use Theorem 11 to compute a lower bound on the ML degree of a variety $X$. On the computational side, Theorem 11 can be applied using coefficient-parameter homotopies to quickly find critical points of $L_u$ over $X$ (Algorithm 22) and can make the problem of computing the ML degree for multi-way tables tractable (Section 4.3).

This paper is organized as follows. In Section 2, we give preliminary definitions and introduce a square parameterized system called the Lagrange likelihood equations. Proposition 4 describes the properties of the Lagrange likelihood equations that will be referenced in later sections. We also describe how the variety of the Lagrange likelihood equations, homogenized appropriately, is related to the likelihood correspondence of Huh and Sturmfels [HS13].

In Section 3, we discuss how sampling and model zeros change the maximum likelihood problem. Theorem 11 describes the special fiber $pr_2^{-1}(u)$ when $u$ contains zero entries. We use this theorem to give a lower bound on the ML degree of $X$. The section continues with exploring how solutions to the Lagrange likelihood equations partition into solutions for different maximum likelihood estimation problems for sampling and model zeros; these partitions are captured in the ML tables introduced in this section. We end this section by fully characterizing the ML degree for different sampling and model zero configurations of a generic hypersurface of degree $d$ in $\mathbb{P}^n$.

We conclude with Section 4, which illustrates several computational advantages that can be achieved in ML degree computation by first considering data vectors with zeros. Algorithm 22 gives a method to find critical points of $L_u$ over $X$ by computing the critical points of $L_u$ when $u$ contains model zeros; these solutions are significantly easier to compute. We continue the section by looking at Grassmanian and tensor examples. We conclude by extending maximum likelihood duality to $u$ with zero entries and showing how ML duality offers further computational benefits.

## 2 Likelihood Equations and ML degree

The maximum likelihood degree (ML degree) of a variety $X \subset \mathbb{P}^n$ is defined as the number of critical points of the likelihood function $L_u$ on $X_{\text{reg}} \setminus \mathcal{H}$ for generic data $u$ [CHKS06]. The ML degree of $X$ quantifies the algebraic complexity of the maximum likelihood estimation problem over the model $\mathcal{M}$, indicating how feasible
symbolic algebraic methods are for finding the MLE. The ML degree has an explicit interpretation in numerical algebraic geometry as well. Assuming that the \textit{ab initio} stage of a coefficient-parameter homotopy has been run \cite{sw05} the ML degree is the number of paths that need to be followed for every subsequent run.

For each \( u \), all critical points of \( L_u \) over \( X \) form a variety. Thus, by varying \( u \) over \( \mathbb{P}^n_u \) we obtain a family of projective varieties with base \( \mathbb{P}_u \). In algebraic geometry, the natural way to view this family of parameterized varieties is as a subvariety \( L_X \) of the product variety \( \mathbb{P}^n_p \times \mathbb{P}^n_u \) where the elements of the family are the fibers of the canonical projection \( pr_2 : \mathbb{P}^n_p \times \mathbb{P}^n_u \to \mathbb{P}^n_u \) over the points \( u \) in \( \mathbb{P}^n_u \). The subvariety \( L_X \) is called the likelihood correspondence \cite{hst13}, which is the closure in \( \mathbb{P}^n_p \times \mathbb{P}^n_u \) of

\[
\{ (p, u) : p \in X_{\text{reg}} \setminus \mathcal{H}_n \text{ and } \text{dlog}(L_u) \text{ vanishes at } p \}.
\]

Just as we can talk about a parameterized family of varieties, we can also talk about a parameterized system of polynomial equations. For us, a \textit{parameterized polynomial system} is a family \( \mathcal{F} \) of polynomial equations in the variables \( p_0, \ldots, p_n \) and the parameters \( u_0, \ldots, u_n \). A member of the family is chosen by assigning a complex number to each parameter \( u_i \). If \( u \) is a generic vector in \( \mathbb{P}^n \), we call the resulting system \textit{generic}. A system of equations is said to be \textit{square} if the number of unknowns (variables) equals the number of equations of the system. Algebraic homotopies are an effective way to solve many members of a family \( \mathcal{F} \). By solving a generic member of the family, we determine the solutions to another system of the family using a \textit{coefficient-parameter homotopy} (see \cite{ms89}), thus, this viewpoint can be computationally advantageous.

In this section, we define a parameterized square system of polynomial equations called the Lagrange likelihood equations. The Lagrange likelihood equations for a variety \( X \subset \mathbb{P}^n \) of codimension \( c \) consists of \( n + 1 + c \) equations. The \( n + 1 + c \) unknowns are \( p_0, p_1, \ldots, p_n, \lambda_1, \ldots, \lambda_c \) and the parameters are \( u_0, \ldots, u_n \). The advantage of the Lagrange likelihood equations, in addition to being a parameterized square system, is that properties of a point \((p, u)\) in the likelihood correspondence become apparent. These properties are summarized in Proposition \[4\].

**Definition 3 (Lagrange likelihood equations).** Suppose \( X \) is a codimension \( c \) irreducible variety that is an irreducible component of the the projective variety defined by the homogeneous polynomials \( h_1, \ldots, h_c \). The Lagrange likelihood equations of \( X \) denoted by \( \text{LL}(X, u) \) are

\[
h_1 = h_2 = \cdots = h_c = 0
\]

(1)

\[
(u_i + p_i - u_i) = p_i \left( \lambda_1 \partial_i h_1 + \lambda_2 \partial_i h_2 + \cdots + \lambda_c \partial_i h_c \right) \text{ for } i = 0, \ldots, n
\]

(2)

If \( X \) is a complete intersection, then \( h_1, \ldots, h_c \) are the minimal generators of \( I(X) \). Otherwise, in order to satisfy the conditions imposed on \( X \), one can choose \( h_1, \ldots, h_c \) to be \( c \) randomized combinations of the minimal generators of \( I(X) \).

**Proposition 4.** The Lagrange likelihood equations have the following properties.

1. If \((p, \lambda)\) is a solution of \( \text{LL}(X, u) \) and \( u_+ \neq 0 \), then \( \sum p_i = 1 \).
2. If \( p_i = 0 \), then \( u_i = 0 \).
3. If the point \( p \) is a critical point of \( L_u \) restricted to \( X_{\text{reg}} \setminus \mathcal{H}_n \), then there exists an unique \( \lambda \) such that \((p, \lambda)\) is a solution to \( \text{LL}(X, u) \).
4. If \( p \in X_{\text{reg}} \) and \((p, \lambda)\) is a regular isolated solution to \( \text{LL}(X, u) \), then \( p \) is a critical point of \( L_u \) on \( X_{\text{reg}} \setminus \mathcal{H}_n \).
5. For generic choices of \( u \), the number of solutions of \( LL(X, u) \) with \( p \in X_{\text{reg}} \setminus \mathcal{H}_n \) equals the MLdegree of \( X \).

Proof. To arrive at property (1), we sum the equations of \([2]\) to get
\[
\sum_{i=0}^{n} (u_+ p_i - u_i - p_i (\lambda_1 \partial h_1 + \cdots + \lambda_c \partial h_c)) = \sum_{i=0}^{n} p_i u_+ - u_+ = u_+ (\sum_{i=0}^{n} p_i - 1).
\]
The first equality above follows by Euler’s relation of homogeneous polynomials.

The implication stated in property (2) is clearly seen by setting \( p_i \) equal to zero in the \( i \)th equation of Equations \([2]\).

For properties (3) and (4), we note that, as discussed in [DR12], \( p \in X_{\text{reg}} \setminus \mathcal{H}_n \) is a critical point of \( L_u \) on \( X \) if and only if the linear subspace \( T^*_p \) contains the point
\[
\left( \frac{u_0}{p_0}, \frac{u_1}{p_1}, \ldots, \frac{u_n}{p_n}, \frac{u_+}{p_+} \right).
\]
When \( X \) is of codimension \( c \), this is equivalent to saying that \( p \in X_{\text{reg}} \setminus \mathcal{H}_n \) is a critical point for \( L_u \) on \( X \) if and only if there exist \( \lambda_1, \ldots, \lambda_c \in \mathbb{C} \) such that for all \( 0 \leq i \leq n \),
\[
\frac{u_i}{p_i} - \frac{u_+}{p_+} = \lambda_1 \cdot \partial h_1 + \cdots + \lambda_c \cdot \partial h_c.
\]
The Langrange likelihood equations are a restatement of this condition with the denominators cleared. Property (5) follows from (3) and (4).

If we homogenize the Lagrange likelihood equations using \( p_+ \) and \( u_+ \) so that each equation is homogeneous in both the coordinates \( p_0, \ldots, p_n \) and the coordinates \( u_0, \ldots, u_n, \lambda_1, \ldots, \lambda_c \), the Lagrange likelihood equations define a variety \( \hat{\mathcal{L}}_X \) in the product space \( \mathbb{P}^n \times \mathbb{P}^n_{u, \lambda} \). The variety \( \hat{\mathcal{L}}_X \) is related to the likelihood correspondence as follows. Let
\[
\pi : \mathbb{P}^n \times \mathbb{P}^{n+c} \to \mathbb{P}^n \times \mathbb{P}^n
\]
\[
((p_0 : \ldots : p_n), (u_0 : \ldots : u_n : \lambda_1 : \ldots : \lambda_n)) \mapsto ((p_0 : \ldots : p_n), (u_0 : \ldots : u_n)).
\]
Then by Proposition \( [4] \) the morphism \( \pi \) maps a dense open set of \( \hat{\mathcal{L}}_X \) to a dense open set of \( \mathcal{L}(X) \), thus,
\[
\mathcal{L}(X) = \pi(\hat{\mathcal{L}}_X).
\]

The implication of this equality is that by studying the Lagrange likelihood equations, we are in fact studying fibers of the projection \( pr_2 : \mathbb{P}^n_p \times \mathbb{P}^n_u \to \mathbb{P}^n_u \).

We conclude this section with an example of using the Lagrange likelihood equations to find critical points of \( L_u \).

**Example 5.** Let \( X = \text{Gr}_{2,6} \subset \mathbb{P}^{14} \) be the variety defined by
\[
p_{ij} p_{kl} - p_{ik} p_{jl} + p_{ij} p_{jk}, \quad 1 \leq i < j < k < l \leq 6.
\]
The Grassmanian \( \text{Gr}_{2,6} \) parameterizes lines in the projective space \( \mathbb{P}^5 \). It has codimension 6 and is not a complete intersection. However, the 6 equations
\[
h_1 = p_{36} p_{45} - p_{35} p_{46} + p_{34} p_{56}, \quad h_4 = p_{26} p_{45} - p_{25} p_{46} + p_{24} p_{56},
\]
\[
h_2 = p_{25} p_{34} - p_{24} p_{35} + p_{23} p_{45}, \quad h_5 = p_{16} p_{45} - p_{15} p_{46} + p_{14} p_{56},
\]
\[
h_3 = p_{15} p_{34} - p_{14} p_{35} + p_{13} p_{45}, \quad h_6 = p_{14} p_{23} - p_{13} p_{24} + p_{12} p_{34}
\]
define a reducible variety that has \( \text{Gr}_{2,6} \) as an irreducible component (the other components live in the coordinate hyperplanes). The system of equations \( LL(X, u) \) consists of 21 equations: the equations \( h_1 = \cdots = h_6 = 0 \) and the 15 below
In this section, we determine what happens when the data vector $X$ thereby, by Proposition 4 the ML degree of $X$ is 156.

3 Sampling and model zeros

In this section, we determine what happens when the data vector $u$ contains zero entries. By understanding the maximum likelihood estimation problems for sampling and model zeros we gain insight into the ML degree of a variety $X$.

For a subset $S \subseteq \{0, 1, \ldots, n\}$, we define

$$U_S := \{ u \in \mathbb{P}^n | u_i = 0 \text{ if } i \in S \text{ and nonzero otherwise} \}.$$ 

The set $U_S$ specifies which entries of the data vector are zero. A partial order on the set of all $\{ U_S : S \subseteq \{0, 1, \ldots, n\} \}$ is induced by inclusion and we notice $U_S \subseteq U_{S'}$ if and only if $S' \subseteq S$. For ease of notation, we define $U := U_0$. When $u \in U_S$, every $u_i$ with $i \in S$ is considered a sampling zero or a model zero.

A sampling zero at cell $i$ changes the likelihood function since the monomial $p_i^{u_i}$ no longer appears in $l_u$. In the case of a model zero at cell $i$, the model zero is not considered as part of the data, and thus, the likelihood function is changed as well: $p_i^{u_i}$ no longer appears in the function and $p_i$ is set to zero in $p_+$. Below, we make precise how the maximum likelihood estimation problem changes in the presence of model zeros and sampling zeros and describe the maximum likelihood estimation problem on $X$ for data $u \in U_S$ with model zeros $R$.

Let $S \subseteq \{0, 1, \ldots, n\}$ and $R \subseteq S$ and consider the following modified likelihood function

$$L_{u,S} := \prod_{i \notin S} p_i^{u_i} / p_+^{u_+}.$$ 

The set $X_R := X \cap \{ p \in \mathbb{P}^n | p_i = 0 \text{ for all } i \in R \}$ will be called the model zero variety for $X$ and $R$. We consider $X_R$ as a projective variety in $\mathbb{P}^{n-|R|}$ and define $\mathcal{H}_R$ as the set of points in $\mathbb{P}^{n-|R|}$ where $(\prod_{i \notin R} p_i)^{-1} \cdot p_+$ vanishes. The model zero variety $X_R$ is called proper if the codimension of $X_R \subset \mathbb{P}^{n-|R|}$ equals the codimension of $X \subset \mathbb{P}^n$.

**Definition 6.** The maximum likelihood estimation problem on $X$ for data $u \in U_S$ with model zeros $R$, denoted $ML_{R,S}$, is to determine the critical points of $L_{u,S}$ on $X_R \setminus \mathcal{H}_R$. The ML degree $(X_R, S)$ is defined to be the number of critical points of $L_{u,S}$ on $X_R \setminus \mathcal{H}_R$ for generic $u \in U_S$ when $X_R$ is proper and zero otherwise.

In terms of the likelihood correspondence, the ML degree $(X_R, S)$ is the cardinality of the subset of points $(p, u)$ of $pr_X^{-1}(u)$ such that $p_i = 0$ for all $i \in R$ for generic $u \in U_S$. Whenever $R = S$, then ML degree $(X_R, S)$ simply equals ML degree $(X_R \subset \mathbb{P}^{n-|R|})$. In terms of optimization, the ML degree $(X_R, S)$ gives an upper bound on the local maxima of $L_{u,S} := \prod_{i \notin S} p_i^{u_i}$ on $\mathcal{M} \cap \{ p_i = 0 \text{ for all } i \in R \}$.

Next, we take the time to explain the subtleties of sampling zeros, model zeros, and structural zeros. When given a model $\mathcal{M}$ with closure $X$ and structural zeros $R,$
common practice is to optimize \( l_{u,R} \) restricted to \( \pi_R(X) \), the closure of the projection of \( X \) onto all coordinates not indexed by \( R \). In contrast, given a model \( M \) with closure \( X \) and model zeros \( R \), the goal is to optimize \( l_{u,R} \) restricted to \( X_R \). In general, \( \pi_R(X) \neq X_R \), and so, the number of critical points will differ. We illustrate the differences in the next three examples.

**Notation 7.** We use \( S \) to denote the indices of the data zeros in \( u \) and \( R \subset S \) to denote the indices of the model zeros. While we defined \( S \subset \{0,1,\ldots,n\} \), in some examples, it is more natural to index the entries of \( u \) by ordered pairs. In this case, \( S \) will be a set of ordered pairs indicating the positions of the data zeros and \( R \) will be a set of ordered pairs indicating the positions of the model zeros.

**Example 8 (Model, sampling, and structural zeros).** Let \( X \) denote the set of \( 3 \times 3 \) matrices of rank 2 in \( \mathbb{P}^8 \). The variety \( X \) is a hypersurface defined by the polynomial \( f = p_{11}p_{22}p_{33} - p_{11}p_{23}p_{32} - p_{12}p_{21}p_{33} + p_{13}p_{21}p_{32} - p_{13}p_{22}p_{31} \). The ML degree of \( X \) is 10.

When we have data \( u \) as a \( 3 \times 3 \) table and the upper left entry \( u_{11} \) is a model zero, then optimization proceeds over \( X_R = X_{\{(1,1)\}} \). The model zero variety \( X_R \) is defined by the polynomial \(-p_{12}p_{21}p_{33} + p_{13}p_{23}p_{31} + p_{13}p_{21}p_{32} - p_{13}p_{22}p_{31} \), obtained by setting \( p_{11} = 0 \) in \( f \). In this case, there are 5 complex critical points, that is, \( \text{MLdegree}(X_R) = 5 \).

When \( u_{11} \) is a sampling zero, optimization proceeds over \( X \) and critical points on the coordinate hyperplanes are ignored. In this case, there are 5 complex critical points whose coordinates are all non-zero, i.e., \( \text{MLdegree}(X, \{(1,1)\}) = 5 \).

When \( u_{11} \) is a structural zero, optimization proceeds over \( \pi_R(X) = \mathbb{P}^7 \). The projection is onto since \( X \) is a hypersurface. In this case, there is one complex critical point.

**Example 9.** Let \( X \) denote the set of \( 3 \times 4 \) matrices of rank 2 in \( \mathbb{P}^{11} \). The defining ideal of \( X \) is generated by the four \( 3 \times 3 \) minors of \( p \),

\[
\begin{align*}
I(X) = \langle & p_{11}p_{22}p_{33} - p_{11}p_{23}p_{32} - p_{12}p_{21}p_{33} + p_{13}p_{21}p_{32} - p_{13}p_{22}p_{31}, \\
p_{11}p_{22}p_{34} - p_{11}p_{24}p_{32} - p_{12}p_{21}p_{34} + p_{13}p_{21}p_{32} + p_{14}p_{21}p_{32} - p_{14}p_{22}p_{31}, \\
p_{11}p_{23}p_{34} - p_{12}p_{24}p_{33} - p_{13}p_{21}p_{34} + p_{13}p_{24}p_{31} + p_{14}p_{21}p_{33} - p_{14}p_{23}p_{31}, \\
p_{12}p_{23}p_{34} - p_{12}p_{24}p_{33} - p_{13}p_{22}p_{34} + p_{13}p_{24}p_{32} + p_{14}p_{22}p_{33} - p_{14}p_{23}p_{32} \rangle.
\end{align*}
\]

The ML degree of \( X \) is 26.

Now let \( u_{11} \) be a model zero in the contingency table \( u \). In this case, \( R = \{(1,1)\} \) and the defining ideal of \( X_R \) is

\[
\begin{align*}
I(X_R) = \langle & p_{12}p_{21}p_{33} + p_{12}p_{23}p_{31} + p_{13}p_{21}p_{32} - p_{13}p_{22}p_{31}, \\
p_{12}p_{21}p_{34} + p_{12}p_{24}p_{31} + p_{14}p_{21}p_{32} - p_{14}p_{22}p_{31}, \\
p_{12}p_{23}p_{34} + p_{13}p_{24}p_{31} + p_{14}p_{23}p_{31} - p_{14}p_{23}p_{31}, \\
p_{12}p_{23}p_{34} + p_{12}p_{24}p_{33} - p_{13}p_{22}p_{34} + p_{13}p_{24}p_{32} + p_{14}p_{22}p_{33} - p_{14}p_{23}p_{32} \rangle.
\end{align*}
\]

The \( \text{MLdegree}(X_R) = 13 \).

When \( u_{11} \) is a structural zero, we follow [Rap06] and eliminate \( p_{11} \) from the ideal \( I(X) \) to obtain the defining ideal of \( \pi_R(X) \),

\[
I(\pi_R(X)) = \langle p_{12}p_{23}p_{34} - p_{12}p_{24}p_{33} - p_{13}p_{22}p_{34} + p_{13}p_{24}p_{32} + p_{14}p_{22}p_{33} - p_{14}p_{23}p_{32} \rangle.
\]

Optimizing over \( \pi_R(X) \), yields 10 complex critical points, which is also the ML degree for \( 3 \times 3 \) rank 2 matrices.
Example 10. Let $X$ be the set of $3 \times 3$ matrices of rank 1 in $\mathbb{P}^8$. It is well known that the ML degree of $X$ equals 1 and that the corresponding critical point of $L_u$ is $\frac{u_{i+}^T}{u_{i+}^T} [u_{i+} u_{i+}]$ for generic choices of data. Now consider the case when

$$u = \begin{bmatrix} 0 & u_{12} & u_{13} \\ u_{21} & 0 & u_{23} \\ u_{31} & u_{32} & 0 \end{bmatrix}.$$ 

The zeros of $u$ are indexed by $S = \{(1,1), (2,2), (3,3)\}$.

If all zeros of $u$ are sampling zeros, then we ask how many critical points of $L_{u,0} = p_{11}^{u_{11}} p_{12}^{u_{12}} \cdots p_{33}^{u_{33}} / p_{+++}^u$ restricted to $X \subset \mathbb{P}^8$ there are. We find the unique critical point is again $\frac{1}{u_{i+}} [u_{i+} u_{i+}]$.

If the zeros of $u$ are model zeros, then we let $R = S$ and we ask how many critical points of $L_{u,R} = p^{u_{12}}_{12} p^{u_{13}}_{13} p^{u_{21}}_{21} p^{u_{23}}_{23} p^{u_{31}}_{31} p^{u_{32}}_{32} / p_{+++}^u$ restricted to $X \cap \{p_{11} = p_{22} = p_{33} = 0\} \setminus H_R \subset \mathbb{P}^5$ there are. We find there are no such critical points. This is because $X \cap \{p_{11} = p_{22} = p_{33} = 0\} \subseteq H_R$.

If the zeros of $u$ are structural zeros, then the model under consideration is a quasi-independence model; such models have been well-studied. The projection $\pi_R(X)$ is defined by one equation $p_{12} p_{23} p_{31} - p_{13} p_{21} p_{32}$, and we find the ML degree of $\pi_R(X)$ is 3.

We now come to the description of the special fiber $pr_2^{-1}(u)$ when $u$ is a generic data vector in $U_S$, which connects this work with previous work on the likelihood correspondence [HS13].

Theorem 11. Let $u$ be a generic data vector in $U_S$ for some $S \subseteq \{0, \ldots, n\}$. Let $X \subseteq \mathbb{P}^n$ be a codimension $c$ irreducible component of a projective variety defined by homogeneous polynomials $h_1, \ldots, h_c$. Let $X_R$ be a proper model zero variety for all $R \subseteq S$. Then, the special fiber $pr_2^{-1}(u)$ contains the critical points of the problem $ML_{R,S}$ for all $R \subseteq S$.

Moreover, if $(p, u) \in pr_2^{-1}(u)$ with $p \in (X_R \setminus H_R)_{reg}$ then $p$ is a critical point of the problem $ML_{R,S}$ for some $R \subseteq S$.

Proof. Most of the work of this proof comes from the formulation of the Lagrange likelihood equations. First, note that for a variety $Y \subseteq \mathbb{P}^n$ and $u \in U_S'$ for $S' \subseteq \{0, 1, \ldots, n\}$, the point $p \in Y_{reg} \setminus H$ is a critical point on $Y$ for $L_{u,S'}$ if and only if the linear subspace $T_p^\perp$ contains the point $v \in [\mathbb{P}^n-\{p\}]$ where

$$v_i = \begin{cases} \frac{u_i}{p_i} - \frac{u_{i+}}{p_+} & \text{if } i \notin S, \\ -\frac{u_{i+}}{p_+} & \text{if } i \in S. \end{cases}$$

This condition results in the same equations as in $LL(Y, u)$ when $u_i = 0$ for all $i \in S$ and $p_i$ is assumed not to be zero when $i \notin S$.

Second, note that when we substitute $p_i = 0$ in to $LL(X, u)$, we get the equations for $LL(X_R, u)$. Thus, by substituting $p_i = 0$ for $i \in R$ and $u_i = 0$ for $i \in S$ into $LL(X, u)$, we get a system of equations whose solutions are the critical points of $L_{u,S}$ on $X_R$.

This implies that if $X_R$ is a proper model zero variety then $p$ is a critical point on $X_R$ for $L_{u,S}$ if and only if there exists $\lambda$ such that $(p, \lambda)$ is an isolated solution to $LL(X, u)$, or equivalently, the point $(p, u) \in L_X$.

From Proposition [32] we know $u_i \neq 0$ implies $p_i \neq 0$, thus, we can account for all solutions to $LL(X, u)$ since we consider every subset $R \subseteq S$. 


In the proof of Theorem 11, we also proved the following statement (Proposition 12). We state Proposition 12 separately in order to highlight the equations for $ML_{R,S}$.

**Proposition 12.** Fix $u \in U_S$ and $X \subset \mathbb{P}^n$ with codimension $c$ that is an irreducible component of the projective variety defined by homogeneous polynomials $h_1, \ldots, h_c$. Whenever $X_R$ is proper, the critical points of $L_{u,S}$ restricted to $X_R$ are regular isolated solutions of the equations:

$$h_1 = h_2 = \cdots = h_c = 0$$
$$p_i = 0 \text{ for } i \in R, \text{ and}$$

$$u_+ = \left(\lambda_1 \partial_i h_1 + \lambda_2 \partial_i h_2 + \cdots + \lambda_c \partial_i h_c\right) \text{ for } i \in S \setminus R$$

$$\left(u_+ p_i - p_+ u_i\right) = p_i \left(\lambda_1 \partial_i h_1 + \lambda_2 \partial_i h_2 + \cdots + \lambda_c \partial_i h_c\right) \text{ for } i \notin S$$

Moreover, the solutions to (3) and (4) for all $R \subseteq S$ account for all the solutions to $LL(X,u)$.

An important consequence of Theorem 11 is that we can use a parameter homotopy to take the solutions of $LL(X,u)$ for $u \in U$ to the solutions of $LL(X,v)$ for $v \in U$. Such methods are discussed in [SW05] and can be implemented in Bertini [BHSW06] or PHCpack [Ver99]. Doing so, we solve $2^{|S|}$ different optimization problems corresponding to the $2^{|S|}$ subsets of $S$. In the case $|S| = 1$, we get the following corollary.

**Corollary 13 (ML degree bound).** Suppose $S = \{n\}$ and $X \subset \mathbb{P}^n$ is an irreducible projective variety. Then for generic $u \in U_S$, we have

$$ML\text{degree}(X) \geq ML\text{degree}(X_S) + ML\text{degree}(X,S)$$

Moreover, when $X$ is a generic complete intersection, the inequality becomes an equality.

**Proof.** This follows from Theorem 11 and the fact that the number of solutions to a parameterized family of polynomial systems for a generic choice of parameters can only decrease on nested parameter spaces [MS89]. Equality holds when $u$ remains off an exceptional subset $E \subset U$ which is defined by an algebraic relation among the $p$ coordinates and $u$ coordinates [SW05]. Since $X$ is a generic intersection, we have $U_S$ is not strictly contained in $E$, and the equality holds.

As we can see from Corollary 13, solutions to $LL(X,u)$ with $u \in U_S$ get partitioned into sampling zero and model zero solutions, in fact, we see this same behavior even as we increase the size of $S$. We encode $ML\text{degree}(X,R,S)$ for all possible choices of $(R,S)$ in a table called the ML table of $X$ whose rows are indexed by $R \subset \{0,1,\ldots,n\}$ and whose columns are indexed by $S \subset \{0,1,\ldots,n\}$. Due to space considerations, in our examples, we only print partial ML tables, i.e. that is subtables of the complete ML table.

**Example 14.** The ML table of a generic curve of degree $d$ in $\mathbb{P}^2$ is below. The top left entry of the table is the ML degree of a generic curve of degree $d$ in $\mathbb{P}^2$.

<table>
<thead>
<tr>
<th>$R \setminus S$</th>
<th>${}$</th>
<th>${0}$</th>
<th>${1}$</th>
<th>${2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${}$</td>
<td>$d + d^2$</td>
<td>$d^2$</td>
<td>$d^2$</td>
<td>$d^2$</td>
</tr>
<tr>
<td>${0}$</td>
<td>$d$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>${1}$</td>
<td>$d$</td>
<td>$0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>${2}$</td>
<td>$d$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 15. Let $X \subset \mathbb{P}^8$ be the projectivization of all $3 \times 3$ matrices of rank 2. A partial ML table of $X$ is below.

<table>
<thead>
<tr>
<th>$R \setminus S$</th>
<th>${}$</th>
<th>${1}$</th>
<th>${12}$</th>
<th>${11,12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${}$</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>${11}$</td>
<td>5</td>
<td>$-$</td>
<td>4</td>
<td>.</td>
</tr>
<tr>
<td>${12}$</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>${11,12}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Example 14 and Example 15 above, each of the columns of the ML table of $X$ sum to ML degree($X$). This does not happen for all varieties, but, in general, the column sums are lower bounds of the ML degree of $X$.

Corollary 16. The column sums of the ML table of $X$ are less than or equal to ML degree($X$), meaning ML degree($X$) $\geq \sum_{R \subseteq S} \text{ML degree}(X_{R,S})$. Moreover, when $X$ is a generic complete intersection, the inequality becomes an equality.

The inequality in Corollary 16 above can be strict as the next example shows.

Example 17. Let $f = p_0^3 + p_1^3 + p_2^3 + p_3^3$ define a hypersurface $X \subset \mathbb{P}^3$. Some of the entries of the ML table of $X$ are below. We have ML degree($X$) = 30 but for $S = \{0,1\}$, we have $\sum_{R \subseteq S} \text{ML degree}(X_{R,S}) = 28$.

<table>
<thead>
<tr>
<th>$R \setminus S$</th>
<th>${}$</th>
<th>${0}$</th>
<th>${0,1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${}$</td>
<td>30</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>${0}$</td>
<td>9</td>
<td>7</td>
<td>$-$</td>
</tr>
<tr>
<td>${1}$</td>
<td>7</td>
<td></td>
<td>$-$</td>
</tr>
<tr>
<td>${0,1}$</td>
<td>2</td>
<td></td>
<td>$-$</td>
</tr>
</tbody>
</table>

Remark 18. Our definition for the entries of the ML table ignores multiplicities and singularities of the variety. We only take account regular isolated solutions. An interesting research direction would be to take into account multiplicities to obtain an equality in the statement of Corollary 13.

We conclude this section with a full description of the ML table for a generic hypersurface of degree $d$ in $\mathbb{P}^n$.

Theorem 19. Suppose $X$ is a generic hypersurface of degree $d$ in $\mathbb{P}^n$ and let $s = |S|$ and $r = |R|$. Then

$$\text{ML degree}(X_{R,S}) = \begin{cases} \frac{d}{d-s} (d^{n-s} - 1), & s = r \\ \frac{d}{d-s+1} (d-1)^{s-r-1}, & s > r \\ 0 & \text{otherwise}. \end{cases}$$

Proof. Since the entries of the ML table of generic degree $d$ hypersurfaces $X \subset \mathbb{P}^n$ depend only on $d$, $n$, and the size of $R$ and $S$, we ease notation and let ML degree($X_r \subset \mathbb{P}^n$, $s$) := ML degree($X_{R \subset \mathbb{P}^n}$, $S$). By Proposition 12, it follows

$$\text{ML degree}(X_r \subset \mathbb{P}^{n+1}, s) = \text{ML degree}(X_{r-1} \subset \mathbb{P}^n, s-1) \quad \text{for } r, s \geq 1 \quad (5)$$

because a section of a generic hypersurface projected into a smaller projective space is again generic degree $d$ hypersurface. We will use (5) to induct on $n$.

Recall by [HKS05] the ML degree of a generic degree $d$ hypersurface in $\mathbb{P}^n$ is $\frac{d}{d-1} (d^n - 1)$. So when $s = r$, we have ML degree($X_{R,S}$) $= \frac{d}{d-s} (d^{n-s} - 1)$ as desired. So for $n = 2$ we have

$$\text{ML degree}(X_0, \emptyset) = \text{ML degree}(X_{\emptyset}, \{0\}) + \text{ML degree}(X_{\emptyset}, \{0\})$$

10
Simple algebra reveals \( \text{MLdegree} \left( X_0 \subset \mathbb{P}^2, \{0\} \right) = d^2 \). With this we have shown the theorem holds when \( n = 2 \). To complete the proof by induction, we need only show \( \text{MLdegree} \left( X_r \subset \mathbb{P}^{n+1}, s \right) \) equals \( d^{n-s+1} (d-1)^{s-r-1} \), when \( r = 0 \) and \( r < s \). To show this we recall

\[
\text{MLdegree} (X \subset \mathbb{P}^{n+1}) = \sum_{R \subset S} \text{MLdegree} \left( X_R \subset \mathbb{P}^{n+1}, S \right).
\]  

(6)

The right hand side of (6) becomes

\[
\text{MLdegree} \left( X_0 \subset \mathbb{P}^{n+1}, s \right) + \sum_{r=1}^{s} \binom{s}{r} \text{MLdegree} \left( X_{r-1} \subset \mathbb{P}^n, s-1 \right).
\]

Letting \( D = \text{MLdegree} \left( X_0 \subset \mathbb{P}^{n+1}, s \right) \) we have that (6) simplifies to

\[
\frac{d}{d-1} (d^{n+1} - 1) = D + \sum_{r=1}^{s-1} \binom{s}{r} d^{n-s+2} (d-1)^{s-r-1} + \frac{d}{d-1} (d^{n-s+1} - 1).
\]

With the binomial formula it follows \( D = d^{n-s+2} (d-1)^{s-1} \) finishing the proof. \( \square \)

**Example 20.** By Theorem 19 we have the following MLtable of a generic degree \( d \) hypersurface \( X \subset \mathbb{P}^n \).

\[
\begin{array}{ccccccc}
R \setminus S & \{\} & \{0\} & \{0, 1\} & \{0, 1, 2\} & \cdots \\
\{\} & \frac{d}{d-1} (d^n - 1) & d^n (d-1)^0 & d^n (d-1)^0 & d^n (d-1)^0 & \cdots \\
\{0\} & d^n (d-1) & \frac{d}{d-1} (d^n-1) & d^n (d-1)^0 & d^n (d-1)^0 & \cdots \\
\{0, 1\} & d^n (d-1) & d^n (d-1) & \frac{d}{d-1} (d^n-2) & d^n (d-1)^0 & \cdots \\
\{0, 1, 2\} & d^n (d-1) & d^n (d-1) & d^n (d-1) & \frac{d}{d-1} (d^n-3) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\
\end{array}
\]

4 Applications and further directions

In this section we illustrate the computational gains acquired by working with model zero varieties. This section has four brief subsections focused on different applications: ML table homotopies, ML duality, tensors (multi-way tables), and Grassmannians.

4.1 ML table homotopy

Let \( X \subset \mathbb{P}^n \) be a generic complete intersection of codimension \( c \) defined by homogeneous polynomials \( h_1, \ldots, h_c \). Let \( u \) be generic data vector in \( U \), and let \( u_s \) be a generic data vector in \( U_S \) with \( S \subseteq \{0, 1, \ldots, n\} \). Our first application of Corollary 13 is the construction of a homotopy to determine critical points of \( L_u \) on \( X \). We determine the critical points of \( L_{u,S} \) on \( X \cap H_R \) for each subset \( R \) of \( S \). So rather than doing a single expensive computation to determine the critical points of \( L_u \) on \( X \), we perform several easier computations to determine critical points of \( L_{u,S} \).

Doing so allows us to use Proposition 12 to get the critical points of \( L_u \) using a coefficient-parameter homotopy. The homotopy requires two steps. Step 1 determines the start points by solving multiple systems of equations. Step 2 constructs the coefficient-parameter homotopy (see [SW05, §7]) that will do the path tracking.
Example 21. Let $X \subset \mathbb{P}^3$ be defined by $f = 2p_0^3 - 3p_1^3 + 5p_2^3 - 7p_3^3$. We note that $\text{MLdegree}(X) = 39$ and the ML table of $X$ is:

$$
\begin{array}{cccc}
R \setminus S & \{\} & \{0\} & \{1\} & \{0, 1\} \\
\{\} & 39 & 27 & 27 & 18 \\
\{0\} & 12 & - & 9 \\
\{1\} & 12 & 9 \\
\{0, 1\} & 3
\end{array}
$$

Let $S = \{0, 1\}$ and let $u_s$ be a generic vector in $U_S$. For Step 1 of the algorithm, we solve four systems of equations. Each system of equations corresponds to a choice of $R$ from $R := \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. For example, when $R = \{0, 1\}$, we solve the following system

$$
\begin{align*}
&f = 0 \\
&p_0 = 0 \\
&p_1 = 0 \\
&(u + p_2 - p_1) = p_2 \lambda_1 \cdot \partial_2 f \\
&(u + p_3 - p_1) = p_3 \lambda_1 \cdot \partial_3 f
\end{align*}
$$

and find 3 solutions. In general, we solve the equations in Proposition 12. So when $R = \emptyset, \{0\}, \{1\}, \{0, 1\}$ we determine there are 18, 12, 12, 3 solutions for the respective systems for a total of 39 solutions. For Step 2, by Proposition 12 the computed 39 solutions are solutions to the Lagrange likelihood equations $LL(X, u_s)$. So by using the coefficient-parameter homotopy $LL(X, u_s \rightarrow u)$, we can go from data with zeros $u_s$ to generic data $u$.

Algorithm 22.

- Input $u_s \in U_S$ and homogeneous polynomials $h_1, h_2, \ldots, h_c$ defining $X$ with codimension $c$.
- (Step 1) Solve $LL(X_R, u_s)$ for each $R \subset S$ to determine the start points of the homotopy.
- (Step 2) Construct and solve the coefficient-parameter homotopy $LL(X, u_s \rightarrow u)$.
- Output solutions to $LL(X, u)$ yielding the critical points of $L_u$ on $X$.

The immediate advantage of this homotopy is that we can get several critical points of $L_u$ quickly. Thus, we get some insight if the ML degree of $X$ is small. Moreover, one can use monodromy methods [SVW01] to attempt to recover additional solutions. One drawback is that by increasing the size of $S$ we also increase the number of subproblems we need to solve, a second drawback is that we may not know a priori that $\sum_{R \subseteq S} \text{MLdegree}(X_R, S)$ equals the ML degree. To address the first drawback, one can take advantage of the structure of the problem to lessen the number of subproblems. For example, in the case when $X$ is a generic hypersurface, we know that the ML degree of $X$ depends only on the size of $R$ and $S$. Taking advantage of this structure and pairing change of variables with parameter homotopies, we preprocess much fewer subproblems—namely $|S| \cdot |\text{subproblems versus } 2^{|S|}$. While we do not have equality in Corollary 14 in general, equality does occur in some examples (see Theorem 19).

4.2 Maximum Likelihood Duality

In this section, we extend ML duality for matrix models when $u$ contains zero entries. We let $X \subset \mathbb{P}^{mn-1}$ be the variety of $m \times n$ matrices of rank less than or equal to $r$
and we let \( Y \subset \mathbb{P}^{mn-1} \) be the variety of \( m \times n \) matrices of rank less than or equal to \( m - r + 1 \) where \( m \leq n \). In [DR12], it is shown that ML degree \( X = \text{ML degree } Y \) by considering critical points of \( L_u \) on subvarieties of the algebraic torus; a bijection between said critical points is also given. Translating these results into the language of determining critical points of \( L_u \) on subvarieties of projective space, we are able to talk about sampling zeros and model zeros. As a consequence, we can use maximum likelihood duality to gain computational advantages by exploiting that sampling zeros are dual to model zeros.

**Proposition 23.** Let \( X \) and \( Y \) be defined as above so that they are ML dual varieties. Let \( S \subset [n] \) and \( u \in U_S \). If \( P \in \mathbb{C}^{mn} \) is a solution to \( \text{LL}(X, u) \), then there exists a \( Q \in \mathbb{C}^{mn} \) such that \( Q \) is a solution to \( \text{LL}(Y, u) \) and

\[
P \ast Q = \Omega_U \tag{7}
\]

where \( \Omega_U = \frac{u}{u^{++}} \ast \frac{u_i+ u_{++}}{u_{i++}} \tag{8} \)

**Proof.** Let \( D \subset \mathbb{P}^{mn-1} \times \mathbb{P}^{mn-1} \times \mathbb{P}^{mn-1} \) be the set of all points \((p,q,u)\) such that \((p,u) \in L_X, (q,u) \in L_Y \) and

\[
u_{i++}^3 p_{ij} q_{ij} - p_{++} q_{++} u_{i+} u_{j+} = 0 \text{ for } 0 \leq i \leq m, 0 \leq i \leq n.
\]

The set \( D \) is a projective variety, thus, if we consider the projection

\[
\phi : \mathbb{P}^n \times \mathbb{P}^n \times \mathbb{P}^n \rightarrow \mathbb{P}^n \times \mathbb{P}^n \\
(p,q,u) \mapsto (p,u),
\]

the image of \( D \) under \( \phi \) is a variety. By [DR12], we know that a dense open subset of \( L_X \) is contained in \( \phi(D) \), therefore, \( L_X \subseteq \phi(D) \) and the statement of the theorem follows.

**Theorem 24.** Let \( X \) and \( Y \) be defined as in Lemma 23. Fix \( S \subset [m] \times [n] \) and generic \( u \in U_S \). Then a solution to the maximum likelihood estimation problem \( \text{ML}_{R,S}(u) \) is dual to a solution to the maximum likelihood estimation problem \( \text{ML}_{R',S}(u) \), with \((S \setminus R) \subset R'\).

When \( |S| = 1 \), the theorem says that a sampling zero critical point is dual to a model zero critical point. We also believe that the converse, model zero critical points are dual to sampling zero critical points is true, and that in general, \((S \setminus R) \subset R'\) is actually an equality in the theorem. Nonetheless, because computing model zeros is heuristically easier than computing sampling zeros, we make computational gains with Theorem 24.

In Example 15, we see that a column of the ML table is symmetric. This is because the variety of \( 3 \times 3 \) matrices of rank 2 is ML self dual. Other examples of varieties that are ML self dual include \( m \times n \) matrices of rank \( \frac{m+1}{2} \) with \( m \) being odd. We conclude this subsection with an ML table of \( 4 \times 4 \) matrices of rank 2 and 3 respectively. An ongoing project is to give a recursive formula for the ML table of \( m \times n \) matrices of rank \( r \) similar to what was done in Theorem 19.
Example 25. ML table of $4 \times 4$ rank 2 matrices:

$$
\begin{array}{c|cccc}
R \setminus S & \{\} & \{11\} & \{11, 44\} & \{11, 22, 44\} & \{11, 22, 33, 44\} \\
{\} & 191 & 118 & 76 & 51 & 35 \\
\{11\} & 73 & 42 & 25 & 16 & \\
\{22\} & - & 25 & 16 & \\
\{33\} & - & - & 16 & \\
\{44\} & 42 & 25 & 16 & \\
\{11, 22\} & - & 17 & 9 & \\
\{11, 33\} & - & - & 9 & \\
\{11, 44\} & 31 & 17 & 9 & \\
\{22, 33\} & - & - & 9 & \\
\{22, 44\} & - & - & 9 & \\
\{33, 44\} & - & - & 9 & \\
\{11, 22, 33\} & - & - & 8 & \\
\{11, 22, 44\} & - & - & 8 & \\
\{11, 33, 44\} & 8 & \\
\{22, 33, 44\} & 8 & \\
\{11, 22, 33, 44\} & 6 & \\
\end{array}
$$

ML table of $4 \times 4$ rank 3 matrices:

$$
\begin{array}{c|cccc}
R \setminus S & \{\} & \{11\} & \{11, 44\} & \{11, 22, 44\} & \{11, 22, 33, 44\} \\
{\} & 191 & 73 & 31 & 14 & 6 \\
\{11\} & 118 & 42 & 17 & 8 & \\
\{22\} & - & - & 8 & \\
\{33\} & - & - & 8 & \\
\{44\} & 42 & 17 & 8 & \\
\{11, 22\} & - & - & 8 & \\
\{11, 33\} & - & - & 8 & \\
\{11, 44\} & 76 & 25 & 9 & \\
\{22, 33\} & - & - & 9 & \\
\{22, 44\} & - & - & 9 & \\
\{33, 44\} & - & - & 9 & \\
\{11, 22, 33\} & - & - & 16 & \\
\{11, 22, 44\} & - & - & 16 & \\
\{11, 33, 44\} & - & - & 16 & \\
\{22, 33, 44\} & - & - & 16 & \\
\{11, 22, 33, 44\} & - & - & 16 & \\
\end{array}
$$

4.3 Tensors

Let $T$ be the set of $2 \times 2 \times 2 \times 2$ tensors with border rank $\leq 2$. The ML degree of this variety is unknown. The variety is defined by the $3 \times 3$ minors of all possible flattenings. This is an overdetermined system of equations with codimension 6. We choose 6 of the equations to be $h_1, \ldots, h_6$ for the Lagrange likelihood equations. For the model zero variety with $p_{1111} = p_{2222} = 0$ we find 3 solutions for a generic $u \in U_S$ with $S = \{1111, 2222\}$. When we solve the Lagrange likelihood equations for $R = \{1111\}$, we find 52 solutions with $p \in X$.

**Theorem 26.** Let $T$ be the set of $2 \times 2 \times 2 \times 2$ tensors with border rank $\leq 2$.

$$\text{MLdegree}(T) \geq 52$$

In this example, we also see that when we have data with zeros the number of critical points can drop significantly as we introduce more model zeros.
4.4 Grassmanians

Let the ideal $I_{2,n}$ be generated by the quadrics

\[ p_{ij}p_{kl} - p_{ik}p_{jl} + p_{il}p_{jk}, \quad 1 \leq i < j < k < l \leq n. \]

Then the variety of $I_{2,n}$ is the Grassmanian $Gr_{2,n} \subset \mathbb{P}^{n-1}$. The Grassmanian $Gr_{2,n}$ parameterizes lines in the projective space $\mathbb{P}^{n-1}$. Below we have a table of computations. The top line of numbers are ML degrees of Grassmanians while the next line are ML degrees of a model zero variety for Grassmanians. The bottom line has ML degrees of sampling zeros.

<table>
<thead>
<tr>
<th>$Gr_{2,4}$</th>
<th>$Gr_{2,5}$</th>
<th>$Gr_{2,6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLdegree $X$</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>MLdegree $(X_{{1}, {12}})$</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>MLdegree $(X_{\emptyset}, {12})$</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>

These computations were performed by choosing $c = \text{codim} X$ generators of $I_{2,n}$ to be $h_1 \ldots h_c$ for $\text{LL}(X, u)$. We used the numerical software bertini and symbolic packages available in M2. From this data we make the following conjecture to motivate the pursuit of a recursive formula for ML degrees of Grassmanians.

**Conjecture 27.** For $n \geq 4$ we conjecture

\[ \text{MLdegree } Gr_{2,n} = \text{MLdegree}(Gr_{2,n+1} \cap \{p_{12} = 0\}). \]

5 Conclusion

Understanding model and sampling zeros gives us insights into the maximum likelihood degree for a given model. When the data vector contains a zero entry, we see that critical points to the likelihood function partition into two groups: critical points for the sampling zero problem and critical points for the model zero problem. This split can help us obtain bounds for the ML degree and provides interesting directions for further research within the study of likelihood geometry, for example, determining which varieties yield an equality in Corollary 13. Furthermore, model zeros can help with the computational problem of finding all the solutions to a set of likelihood equations. This paper illustrates some of the advantages of working with model zeros, as seen by the lower bound obtained on the set of $2 \times 2 \times 2 \times 2$ tensors of border rank $\leq 2$. We hope that the problem mentioned in Section 4.2 of determining whether model zero critical points are dual to sampling zero critical points is furthered explored. A positive answer would yield significant gains in understanding the ML degree for determinantal varieties.

6 Acknowledgements

The authors would like to thank Kaie Kubjas, and Anton Leykin, Bernd Sturmfels, and Seth Sullivant for their helpful suggestions and comments.

References


