The maximum likelihood degree of Fermat hypersurfaces

Daniele Agostini[1], Davide Alberelli[2], Francesco Grande[3], Paolo Lella[4,*]

[1] *Institut für Mathematik, Humboldt Universität zu Berlin, Berlin, Germany*
[2] *Institut für Mathematik, University of Osnabrück, Osnabrück, Germany*
[3] *Institut für Mathematik, Freie Universität Berlin, Berlin, Germany*
[4] *Dipartimento di Matematica, Università degli Studi di Trento, Trento, Italy*

Abstract. We study the critical points of the likelihood function over the Fermat hypersurface. This problem is related to one of the main problems in statistical optimization: maximum likelihood estimation. The number of critical points over a projective variety is a topological invariant of the variety and is called maximum likelihood degree. We provide closed formulas for the maximum likelihood degree of any Fermat curve in the projective plane and of Fermat hypersurfaces of degree 2 in any projective space. Algorithmic methods to compute the ML degree of a generic Fermat hypersurface are developed throughout the paper. Such algorithms heavily exploit the symmetries of the varieties we are considering. A computational comparison of the different methods and a list of the maximum likelihood degrees of several Fermat hypersurfaces are available in the last section.

2000 Mathematics Subject Classifications: 14Q10, 14N10, 13P25, 62F10.

Key Words and Phrases: Maximum likelihood, Fermat hypersurface, Likelihood correspondence.

1. Introduction

One of the main problems in statistics is the maximum likelihood estimation (MLE) of a statistical model. It has been widely explored and provided of very efficient methods. Recently a new branch of mathematics, called algebraic statistics, opened new horizons to the study of some statistical models by means of polynomial algebra. Among others, the result of uniqueness of the MLE for linear and log-linear (toric) models can be easily deduced using the algebraic approach [19, Proposition 1.4, Theorem 1.10].

However, most algebraic statistical models have more than one local maximum for the maximum likelihood estimation problem (see for instance [19, Example 3.26]). Thus, it is natural to ask how many critical points of the likelihood function lie on the model; this question is known as the maximum likelihood (ML) degree problem. Some machinery from

algebraic geometry has been applied to study particular cases of the ML degree problem; such synergy has produced quite a few results (see for instance [12, 2, 13, 4, 8, 20, 11]). In particular, many hidden geometric features of algebraic statistical models have been brought to light. Most results in this area have been gathered and extended in [16].

The task is to compute the ML degree of a Fermat hypersurface of degree d in the complex projective space $\mathbb{P}^n$. No closed formula $\phi(n, d)$ for computing the ML degree of Fermat hypersurfaces seems to exist for general n and d. Hence, we look for the answer by a direct approach with the aid of algebraic geometry software (*Macaulay2* [7]).

The Fermat hypersurface $F_{n,d}$ of degree d in the projective space $\mathbb{P}^n$ is the zero locus of the polynomial

$$f_{n,d} := x_0^d + x_1^d + \ldots + x_n^d. \tag{1}$$

The ML degree of $F_{n,d}$ is the number of critical points of the likelihood function

$$\ell_{\boldsymbol{u}} := \frac{x_0^{u_0} \cdots x_n^{u_n}}{(x_0 + \cdots + x_n)^{u_0 + \cdots + u_n}}, \qquad \boldsymbol{u} = (u_0, \ldots, u_n) \in \mathbb{Z}_{>0}^{n+1}$$

on the hypersurface defined by (1) for a general $\boldsymbol{u}$. The standard approach to the problem is to consider the logarithmic derivatives of $\ell_{\boldsymbol{u}}$ and apply the theorem of Lagrange multipliers (see for instance [19, Chapter 1] and [16, Section 1]). Hence, $\boldsymbol{p} = (p_0, \ldots, p_n)$ is a critical point for $\ell_{\boldsymbol{u}}$ restricted to $F_{n,d}$ if, and only if, $f_{n,d}(\boldsymbol{p}) = 0$ and the rank of the matrix

$$\begin{pmatrix} \dfrac{u_0}{p_0} & \dfrac{u_1}{p_1} & \cdots & \dfrac{u_n}{p_n} \\ 1 & 1 & \cdots & 1 \\ dp_0^{d-1} & dp_1^{d-1} & \cdots & dp_n^{d-1} \end{pmatrix} \tag{2}$$

is not maximal. It is clear after few computations that even for small values of n and d the time needed to get an answer via the previous definition of critical point becomes huge. Refining this approach, in this paper we succeed in computing the ML degree of $F_{n,d}$ for a larger set of n and d (see Table 3). This requires an algorithmic method that uses tools of algebraic geometry and exploits the symmetries of the Fermat hypersurface. We do not explore the possible statistical applications. Nevertheless, it may happen that similar ideas can be applied to other classes of highly symmetric models derived from statistical observations.

Section 2 of the paper presents a formulation of the maximum likelihood degree problem in the language of algebraic geometry and introduces the standard concepts and tools required in the rest of the paper. Moreover, after pointing out the key difficulties of the algorithmic procedure, we illustrate a first try to improve the computations.

Section 3 contains the main result of the paper for the Fermat hypersurfaces. We show that the ML degree of $F_{n,d}$ can be computed by setting to 1 all the entries of the data vector $\boldsymbol{u}$. This is the point where algebraic geometry plays a key role. Indeed, to prove this fact, we construct a family of schemes $\mathcal{X} \to \mathbb{A}^1$ such that the fiber over $t \in \mathbb{A}^1$ is the ideal which encodes the solution of the ML degree problem for a family of data vectors

$\boldsymbol{u}_t$ (the vectors depend on t). We are particularly interested in the ideals defining two special fibers. The first one solves correctly the ML degree problem $F_{n,d}$ and the second one describes the critical points in the case $\boldsymbol{u} = (1, \ldots, 1)$. We prove that both ideals have the same number of solutions by showing that the family is flat.

In Section 4, we investigate the symmetries of the problem. In fact, using the data vector $(1, \ldots, 1)$, the action of the symmetric group $\mathcal{S}_{n+1}$ on $F_{n,d}$ extends to the matrix (2) and, in particular, to the critical points of the likelihood function. By looking at the orbits of the critical points and studying the number of distinct coordinates a critical point might have, we are able to subdivide the ML degree computation into parallel subtasks whose computations involve less than $n+1$ variables and, for this reason, are easier. (See [9] for another example of ML degree computation based on subdividing the main problem into several simpler problems.)

Section 5 is dedicated to closed formulas for two special families of Fermat hypersurfaces, namely $F_{n,2}$ and $F_{2,d}$. The first formula is an application of the results achieved in Section 4, while the second one is obtained using topological arguments and is amazingly simple, based exclusively on the congruence modulo 6 of the degree d.

The last section reports the computational results by mean of comparison tables between the running times of the naive algorithms and the improved algorithms. It is immediate to see that the advantages of using the second one are remarkable when $d \ll n$. As a conclusion we include a table with the maximum likelihood degrees for the Fermat hypersurfaces that have been computed so far.

2. The ML degree problem

The maximum likelihood degree problem we are facing has been described in the introduction. The formulation in the language of algebraic geometry is the following: we want to study the variety of $\mathbb{P}^n$ defined by the 3×3 minors of

$$\begin{pmatrix} u_0 & u_1 & \ldots & u_n \\ x_0 & x_1 & \ldots & x_n \\ x_0^d & x_1^d & \ldots & x_n^d \end{pmatrix} \tag{3}$$

and the equation (1) (notice that the matrix (3) is obtained from the matrix (2) by multiplying the i-th column by p_i for all $i = 0, \ldots, n$). In addition, by similarity to some constraints needed in the classical statistical modelling, we only want to consider the points not lying on the hyperplanes defined by the equations $x_0 = \ldots = x_n = x_0 + \cdots + x_n = 0$. We call this hyperplane arrangement the *distinguished arrangement* $\mathcal{H}$. Throughout the paper, the notation $\mathcal{M}_{n,d}^{\boldsymbol{u}}$ will be used to refer to the matrix (3). We denote by $\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{u}})$ the minor of $\mathcal{M}_{n,d}^{\boldsymbol{u}}$ corresponding to the columns i, j, k, i.e.

$$\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{u}}) = u_i\, x_j x_k (x_k^{d-1} - x_j^{d-1}) + u_j\, x_k x_i (x_i^{d-1} - x_k^{d-1}) + u_k\, x_i x_j (x_j^{d-1} - x_i^{d-1}),$$

and by $I_{n,d}^{\boldsymbol{u}}$ the ideal generated by the 3×3 minors and the Fermat equation:

$$I_{n,d}^{\boldsymbol{u}} := \big(f_{n,d},\ \mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{u}}) \mid \forall\, 0 \leqslant i < j < k \leqslant n\big) \subset \mathbb{C}[x_0, \ldots, x_n].$$

The critical points we are looking for are the solutions of the ideal $I_{n,d}^{\boldsymbol{u}}$ not lying on the distinguished arrangement, i.e. the ones defined by the saturated ideal

$$I_{n,d}^{\boldsymbol{u}} \setminus \mathcal{H} := \left(I_{n,d}^{\boldsymbol{u}} : \left(x_0 \cdots x_n (x_0 + \cdots + x_n) \right)^{\infty} \right). \tag{4}$$

If we consider also the entries of the data vector $\boldsymbol{u}$ as coordinates of a point in another projective space $\mathbb{P}^n = \operatorname{Proj} \mathbb{C}[y_0, \ldots, y_n]$, it is natural to consider the *likelihood correspondence*, which is the universal family of these critical points, i.e. the subscheme $\mathcal{L}_{n,d} \subset \mathbb{P}^n \times \mathbb{P}^n$ defined as the closure of

$$\left\{ (\boldsymbol{p}, \boldsymbol{u}) \in \mathbb{P}^n \times \mathbb{P}^n \mid \boldsymbol{p} \text{ is solution of } I_{n,d}^{\boldsymbol{u}} \setminus \mathcal{H} \right\}$$

whose equations are described by the ideal $I_{n,d}^{\boldsymbol{y}} \setminus \mathcal{H} \subset \mathbb{C}[x_0, \ldots, x_n, y_0, \ldots, y_n]$.

Let us denote by $\mathbb{P}_{\boldsymbol{x}}^n$ the projective space containing the Fermat hypersurface, by $\mathbb{P}_{\boldsymbol{y}}^n$ the projective space of data and by $\pi_{\boldsymbol{x}}$ and $\pi_{\boldsymbol{y}}$ the standard projections onto the two factors. Now we recall two theorems that motivate the definition of ML degree we will consider.

Theorem 1 ([16, Theorem 1.6]). *The likelihood correspondence $\mathcal{L}_X$ of any irreducible subvariety $X \subset \mathbb{P}_{\boldsymbol{x}}^n$ is an irreducible variety of dimension n in the product $\mathbb{P}_{\boldsymbol{x}}^n \times \mathbb{P}_{\boldsymbol{y}}^n$. The map $\pi_{\boldsymbol{y}} : \mathcal{L}_X \to \mathbb{P}_{\boldsymbol{y}}^n$ is generically finite-to-one.*

Theorem 2 ([16, Theorem 1.15]). *Let $\boldsymbol{u} \in \mathbb{R}_{>0}^{n+1}$, and let $X \subset \mathbb{P}^n$ be an irreducible variety such that no singular point of any intersection $X \cap \{x_i = 0\}$ lies in the hyperplane $\{x_0 + \cdots + x_n = 0\}$. Then*

1. the likelihood function $\ell_{\boldsymbol{u}}$ on X has only finitely many critical points in $X_{\mathrm{reg}} \setminus \mathcal{H}$;

2. if the fiber $\pi_{\boldsymbol{y}}^{-1}(\boldsymbol{u})$ is contained in X_{reg}, then its length equals the ML degree of X.

Notice that the hypotheses of Theorem 2 are automatically verified in the case we are interested in. Indeed, the Fermat hypersurface $F_{n,d}$ is smooth, i.e. $(F_{n,d})_{\mathrm{reg}} = F_{n,d}$, and every intersection $F_{n,d} \cap \{x_i = 0\}$ is again a Fermat hypersurface in a space of lower dimension. Thus, we adopt the following definition.

Definition 1. *The ML degree $\operatorname{MLdeg} X$ of the irreducible subvariety $X \subset \mathbb{P}_{\boldsymbol{x}}^n$ is the degree of the projection of the likelihood correspondence $\mathcal{L}_X$ to the second factor $\pi_{\boldsymbol{y}} : \mathcal{L}_X \to \mathbb{P}_{\boldsymbol{y}}^n$.*

In order to determine the ML degree of the Fermat hypersurface, we can begin considering the following two standard approaches.

Multidegree. We can consider the multidegree of the likelihood correspondence $\mathcal{L}_{n,d}$ in the sense of [18, Chapter 8] with respect to the natural $\mathbb{Z}^2$-grading on the polynomial ring $\mathbb{C}[x_0, \ldots, x_n, y_0, \ldots, y_n]$. The multidegree of $\mathcal{L}_{n,d}$ is a polynomial $B_{\mathcal{L}_{n,d}}$ in the ring $\mathbb{Z}[T_{\boldsymbol{x}}, T_{\boldsymbol{y}}]$ of degree $n = \dim \mathcal{L}_{n,d}$. It can be computed by means of the prime ideal defining $\mathcal{L}_{n,d}$ [18, Proposition 8.49] and turns out to have the following shape

$$B_{\mathcal{L}_{n,d}}(T_{\boldsymbol{x}}, T_{\boldsymbol{y}}) = (\operatorname{MLdeg} F_{n,d}) T_{\boldsymbol{x}}^n + \cdots + (\deg F_{n,d}) T_{\boldsymbol{x}} T_{\boldsymbol{y}}^{n-1}. \tag{5}$$

Hence, we can compute the ML degree of the Fermat hypersurface as the leading coefficient of the multidegree $B_{\mathcal{L}_{n,d}}(T_{\boldsymbol{x}}, T_{\boldsymbol{y}})$ of the likelihood correspondence.

Random data. The degree of the map $\pi_{\boldsymbol{y}} : \mathcal{L}_{n,d} \to \mathbb{P}^n_{\boldsymbol{y}}$ is the degree of the generic fiber of $\pi_{\boldsymbol{y}}$. Thus, there is an open dense subset $\mathcal{U} \subset \mathbb{P}^n_{\boldsymbol{y}}$ whose points have fiber of constant degree. A computational strategy to determine $\mathrm{MLdeg}\, F_{n,d}$ is to randomly pick a point of $\mathbb{P}^n_{\boldsymbol{y}}$ and to calculate the degree of its fiber. Indeed, the probability to randomly choose a point in the Zariski closed subset $\mathbb{P}^n_{\boldsymbol{y}} \setminus \mathcal{U}$ is negligible, so that we get almost surely the degree of the projection $\pi_{\boldsymbol{y}}$, i.e. the ML degree.

Computational experiments (see Table 2(a) and Table 2(c) for details) show that these two methods for determining the ML degree of $F_{n,d}$ have a long execution time even for small values of n and d. This is mainly due to the computation of a Gröbner basis of the ideal, which is needed to determine the degree (resp. multidegree) of the ideal of the generic fiber (resp. likelihood correspondence). A little more accurate analysis of the computational experiments reveals that the time required by the elimination of the critical points lying on the distinguished arrangement $\mathcal{H}$ is the hardest part of the process. Indeed, to accomplish this task, we need to saturate the ideal $I^{\boldsymbol{u}}_{n,d}$ (resp. $I^{\boldsymbol{y}}_{n,d}$) by $n+2$ linear forms $(x_0, \ldots, x_n$ and $x_0 + \cdots + x_n)$ and this operation may be computationally expensive. In the case of the Fermat hypersurface, we see that we can reduce to saturation by a single linear form.

Lemma 1. *Let $\boldsymbol{u}$ be a data vector such that $u_i \neq 0$, $\forall\, i$, and let $\boldsymbol{p} = [p_0 : \ldots : p_n] \in \mathbb{P}^n_{\boldsymbol{x}}$ be a solution of the ideal $I^{\boldsymbol{u}}_{n,d}$. If for some i the coordinate p_i vanishes, then $\sum_{j=0}^n p_j = 0$.*

Proof. First, note that there are at least two non-zero coordinates, as the point is a solution of the equation $f_{n,d} = 0$. Let p_k and p_h be any two non-zero coordinates. Evaluating the minor $\mathrm{e}_{ikh}(\mathcal{M}^{\boldsymbol{u}}_{n,d})$ on the 3-tuple $(p_i = 0, p_k, p_h)$, we obtain

$$\mathrm{e}_{ikh}\big(\mathcal{M}^{\boldsymbol{u}}_{n,d}\big)(0, p_k, p_h) = u_i\, p_k p_h (p_h^{d-1} - p_k^{d-1}),$$

so that for any solution of the ideal we have $p_h^{d-1} = p_k^{d-1}$. Finally,

$$0 = \sum_{j=0}^n p_j^d = \sum_{\substack{j=0 \\ p_j \neq 0}}^n p_j^d = \sum_{\substack{j=0 \\ p_j \neq 0}}^n p_j\, p_j^{d-1} = p_k^{d-1} \sum_{\substack{j=0 \\ p_j \neq 0}}^n p_j = p_k^{d-1} \sum_{j=0}^n p_j,$$

where p_k is just one of the non-zero coordinates.

The previous lemma implies that saturating the ideal by the linear form $x_0 + \cdots + x_n$ suffices to guarantee that no solution lies on the hyperplane arrangement $\mathcal{H}$ for a generic data vector $\boldsymbol{u}$. Saturating with respect to a unique linear form instead of $n+2$ is certainly an improvement. Unfortunately, the saturation with respect to $x_0 + \cdots + x_n$ is much more expensive than the saturation with respect to a single variable. Next proposition shows that we can avoid the saturation by $x_0 + \cdots + x_n$ and compute $\mathrm{MLdeg}\, F_{n,d}$ as the difference between the number of points defined by $I^{\boldsymbol{u}}_{n,d}$ and the number of points defined by the ideal $I^{\boldsymbol{u}}_{n,d} + (x_0 + \cdots + x_n)$.

Proposition 1. *For a generic $\boldsymbol{u} \in \mathbb{P}^n_{\boldsymbol{y}}$ the points defined by $I^{\boldsymbol{u}}_{n,d}$ lying on the hyperplane $x_0 + \ldots + x_n = 0$ are simple. Hence,*

$$\deg\left(I^{\boldsymbol{u}}_{n,d} \setminus \mathcal{H}\right) = \deg I^{\boldsymbol{u}}_{n,d} - \deg\left(I^{\boldsymbol{u}}_{n,d} + (x_0 + \cdots + x_n)\right). \tag{6}$$

Proof. By symmetry, we can restrict our attention to the affine chart $\mathcal{U}_0 = \{x_0 \neq 0\} \subset \mathbb{P}^n_{\boldsymbol{x}}$ with affine coordinates $t_i = \frac{x_i}{x_0}$ for $i = 1, \ldots, n$. The ideal $I^{\boldsymbol{u}}_{n,d}$ restricted to $\mathcal{U}_0$ is defined by the polynomial $1 + t_1^d \cdots + t_n^d$ and by the 3×3 minors of the matrix

$$\begin{pmatrix} u_0 & u_1 & \ldots & u_n \\ 1 & t_1 & \ldots & t_n \\ 1 & t_1^d & \ldots & t_n^d. \end{pmatrix}.$$

First, we observe that

$$\mathrm{rk}\begin{pmatrix} u_0 & u_1 & \ldots & u_n \\ 1 & t_1 & \ldots & t_n \\ 1 & t_1^d & \ldots & t_n^d. \end{pmatrix} = \mathrm{rk}\begin{pmatrix} u_0 + \cdots + u_n & u_1 & \ldots & u_n \\ 1 + t_1 + \cdots + t_n & t_1 & \ldots & t_n \\ 1 + t_1^d + \cdots + t_n^d & t_1^d & \ldots & t_n^d. \end{pmatrix}.$$

Since $\boldsymbol{u}$ is generic we can assume that $\sum_{i=0}^n u_i \neq 0$ and then normalize to $\sum_{i=0}^n u_i = 1$. Moreover, $1 + t_1^d + \cdots + t_n^d$ vanishes since the points we are looking at lie on the Fermat hypersurface $F_{n,d}$. Let $h = 1 + t_1 + \cdots + t_n$. Applying a sequence of column operations to the matrix yields the following equalities

$$\mathrm{rk}\begin{pmatrix} 1 & u_1 & \ldots & u_n \\ h & t_1 & \ldots & t_n \\ 0 & t_1^d & \ldots & t_n^d \end{pmatrix} = \mathrm{rk}\begin{pmatrix} 1 & 0 & \ldots & 0 \\ h & t_1 - u_1 h & \ldots & t_n - u_n h \\ 0 & t_1^d & \ldots & t_n^d \end{pmatrix} =$$

$$= \mathrm{rk}\begin{pmatrix} t_1 - u_1 h & \ldots & t_n - u_n h \\ t_1^d & \ldots & t_n^d \end{pmatrix} + 1 =$$

$$= \mathrm{rk}\begin{pmatrix} \ldots & t_i - u_i h & \ldots & (h-1) - h(1 - u_0) \\ \ldots & t_i^d & \ldots & t_1^d + \cdots + t_n^d \end{pmatrix} + 1 =$$

$$= \mathrm{rk}\begin{pmatrix} \ldots & t_i - u_i h & \ldots & u_0 h - 1 \\ \ldots & t_i^d & \ldots & -1 \end{pmatrix} + 1 =$$

$$= \mathrm{rk}\begin{pmatrix} \ldots & (u_0 h - 1)t_i^d + t_i - u_i h & \ldots & u_0 h - 1 \\ \ldots & 0 & \ldots & -1 \end{pmatrix} + 1$$

so that

$$I^{\boldsymbol{u}}_{n,d}\big|_{\mathcal{U}_0} = \left(1 + t_1^d + \cdots + t_n^d, (u_0 h - 1)t_i^d + t_i - u_i h \mid \forall\, 1 \leqslant i \leqslant n - 1\right).$$

In order to prove that the solutions of $I^{\boldsymbol{u}}_{n,d}\big|_{\mathcal{U}_0}$ that satisfy $h = 0$ are simple, we compute the Jacobian matrix of $I^{\boldsymbol{u}}_{n,d}\big|_{\mathcal{U}_0}$ and we check that the locus where it is not of maximal rank

does not intersect the set defined by $I_{n,d}^{\boldsymbol{u}}\big|_{\mathcal{U}_0} + (h)$. The Jacobian matrix restricted to the set $\{h = 0\}$ is

$$
\begin{pmatrix}
dt_1^{d-1} & \dots & u_0 t_i^d - u_i & & \dots \\
\vdots & & \vdots & & \\
dt_i^{d-1} & \dots & u_0 t_i^d - u_i + (1 - dt_i^{d-1}) & \dots \\
\vdots & & \vdots & & \\
dt_n^{d-1} & \dots & u_0 t_i^d - u_i & & \dots
\end{pmatrix}
$$

and the rank of this matrix equals the rank of the matrix

$$
W := \begin{pmatrix}
t_1^{d-1} - t_n^{d-1} & 1 - dt_1^{d-1} & \dots & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots & & \vdots \\
t_i^{d-1} - t_n^{d-1} & 0 & \dots & 1 - dt_i^{d-1} & \dots & 0 \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
t_{n-1}^{d-1} - t_n^{d-1} & 0 & \dots & 0 & \dots & 1 - dt_{n-1}^{d-1} \\
t_n^{d-1} & u_0 t_1^d - u_1 & \dots & u_0 t_i^d - u_i & \dots & u_0 t_{n-1}^d - u_{n-1}
\end{pmatrix}.
$$

Now, we have to show that the ideal $J := I_{n,d}^{\boldsymbol{u}}\big|_{\mathcal{U}_0} + (\det W, h)$ has no solution. Reducing by h the generators of $I_{n,d}^{\boldsymbol{u}}\big|_{\mathcal{U}_0}$ we obtain

$$
J = (\det W, h, 1 + t_1^d + \dots + t_n^d, t_i^d - t_i \mid \forall\, 1 \leqslant i \leqslant n-1) = (\det W, h, t_i^d - t_i \mid \forall\, 1 \leqslant i \leqslant n),
$$

which implies that every solution of J satisfy the following condition: either $t_i = 0$ or $t_i^{d-1} = 1$ for every $i = 1, \dots, n$. Since the matrix W is symmetric in the variables $t_1, \dots, t_{n-1}$, we can assume $t_i \neq 0$ for $1 \leqslant i \leqslant r \leqslant n-1$ and $t_i = 0$ for $r+1 \leqslant i \leqslant n-1$. Thus, the matrix W has the following form

$$
\begin{pmatrix}
1 - t_n^{d-1} & 1 - d & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & 0 & \vdots \\
1 - t_n^{d-1} & 0 & \dots & 1 - d & 0 & \dots & 0 \\
-t_n^{d-1} & 0 & \dots & 0 & 1 & \dots & 0 \\
\vdots & \vdots & 0 & \vdots & \vdots & \ddots & \vdots \\
-t_n^{d-1} & 0 & \dots & 0 & 0 & \dots & 1 \\
t_n^{d-1} & u_0 t_1 - u_1 & \dots & u_0 t_r - u_r & -u_{r+1} & \dots & -u_{n-1}
\end{pmatrix}
$$

from which we can easily compute its determinant and express it (up to a constant) as follows

$$
\det W = t_n^{d-1} - \frac{1 - t_n^{d-1}}{d-1} \sum_{i=1}^{r} (u_0 t_i - u_i) - \sum_{i=r+1}^{n-1} u_i t_n^{d-1} =
$$

$$
= \begin{cases}
\dfrac{1}{d-1}\left(\displaystyle\sum_{i=1}^{r} u_i + u_0 \right), & \text{if } t_n = 0, \\[3ex]
1 - \displaystyle\sum_{i=r+1}^{n-1} u_i, & \text{if } t_n^{d-1} = 1.
\end{cases}
$$

For a generic point $\boldsymbol{u} \in \mathbb{P}_{\boldsymbol{y}}^n$, the determinant does not vanish and since this reasoning works for every $1 \leqslant r \leqslant n-1$, we proved that $J = (1)$.

Finally, in order to prove the last statement, we notice that, as the points of $I_{n,d}^{\boldsymbol{u}}$ lying on $\mathcal{H}$ in fact lies on $x_0 + \cdots + x_n = 0$ (Lemma 1) and are simple, the following equality holds:

$$I_{n,d}^{\boldsymbol{u}} \setminus \mathcal{H} = \left(I_{n,d}^{\boldsymbol{u}} : (x_0 + \cdots + x_n)^\infty \right) = \left(I_{n,d}^{\boldsymbol{u}} : (x_0 + \cdots + x_n) \right).$$

Remark 1. *Notice that the proof of Proposition 1 applies also for $\boldsymbol{u} = [1 : \ldots : 1]$.*

We conclude this section by describing the shape of the ideal $\left(I_{n,d}^{\boldsymbol{u}} + (x_0 + \cdots + x_n) \right)$.

Lemma 2. *Let $\boldsymbol{u} = [u_0 : \ldots : u_n]$ be a point in $\mathbb{P}_{\boldsymbol{y}}^n$ such that $u_0 + \cdots + u_n \neq 0$. The solutions of $I_{n,d}^{\boldsymbol{u}}$ lying on the hyperplane $x_0 + \cdots + x_n = 0$ do not depend on the point $\boldsymbol{u}$. More precisely,*

$$I_{n,d}^{\boldsymbol{u}} + (x_0 + \cdots + x_n) = \left(\sum_{j=0}^n x_j, \ \sum_{j=0}^n x_j^d, \ x_k x_h (x_h^{d-1} - x_k^{d-1}), \ \forall \, 0 \leqslant k < h \leqslant n \right). \quad (7)$$

Proof. The ideal of 3×3 minors of $\mathcal{M}_{n,d}^{\boldsymbol{u}}$ contains also the minors of any matrix obtained from $\mathcal{M}_{n,d}^{\boldsymbol{u}}$ by column operations. For instance, we can fix i and replace the i-th column with the sum of all columns. We get the matrix

$$\begin{pmatrix} u_0 & \cdots & u_{i-1} & \sum_j u_j & u_{i+1} & \cdots & u_n \\ x_0 & \cdots & x_{i-1} & \sum_j x_j & x_{i+1} & \cdots & x_n \\ x_0^d & \cdots & x_{i-1}^d & \sum_j x_j^d & x_{i+1}^d & \cdots & x_n^d \end{pmatrix} = \begin{pmatrix} u_0 & \cdots & u_{i-1} & \sum_j u_j & u_{i+1} & \cdots & u_n \\ x_0 & \cdots & x_{i-1} & 0 & x_{i+1} & \cdots & x_n \\ x_0^d & \cdots & x_{i-1}^d & 0 & x_{i+1}^d & \cdots & x_n^d \end{pmatrix}$$

as we are assuming both $x_0 + \cdots + x_n$ and $x_0^d + \cdots + x_n^d$ equal to 0. Among the minors of the second matrix, we have, up to sign, $(\sum_j u_j) x_k x_h (x_h^{d-1} - x_k^{d-1})$, $k, h \neq i$, and varying i, we prove that

$$\left(\sum x_j, \ \sum x_j^d, \ x_k x_h (x_h^{d-1} - x_k^{d-1}), \ \forall \, 0 \leqslant k < h \leqslant n \right) \subseteq I_{n,d}^{\boldsymbol{u}} + (x_0 + \cdots + x_n).$$

The other inclusion is straightforward if we notice that the 3×3 minors of $\mathcal{M}_{n,d}^{\boldsymbol{u}}$ are linear combinations of the polynomials $x_k x_h (x_h^{d-1} - x_k^{d-1})$.

See Table 2(a),(b) and Table 2(c),(d) for a comparison between the running time of the naive strategy (based on the saturation) and the strategy based on Proposition 1 applied both to the multidegree and random data approach.

3. Symmetrizing the problem

To improve further the computations, we would like to extend some symmetries of the Fermat hypersurfaces to symmetries of the ML degree problem. More precisely, we will consider the action of the symmetric group $\mathcal{S}_{n+1}$ on the variables of $\mathbb{C}[x_0, \ldots, x_n]$. By

looking at the matrix (3), we notice that the ideal (4) is symmetric with respect to $\mathcal{S}_{n+1}$ if the polynomials $\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{u}})$ are invariant under the action of $\mathcal{S}_{n+1}$. This is equivalent to the requirement that all the entries of the data vector $\boldsymbol{u}$ are equal. From a statistical point of view, we are restricting to the very specific case where we observe the same number of occurrences for each random variable. From the algebraic geometry point of view, we are claiming that the point $\mathbf{1} := [1 : \ldots : 1] \in \mathbb{P}_{\boldsymbol{y}}^n$ belongs to the open subset $\mathcal{U}$ of points whose fiber has the correct degree.

For any multi-index $\boldsymbol{u} = (u_0, \ldots, u_n)$, let us denote by $|\boldsymbol{u}|$ the sum $u_0 + \cdots + u_n$, by $\hat{u}_i$ the difference $|\boldsymbol{u}| - u_i$ and by $\hat{\boldsymbol{u}}$ the multi-index $(\hat{u}_0, \ldots, \hat{u}_n)$.

Consider a generic point $\boldsymbol{u} \in \mathcal{U} \subset \mathbb{P}_{\boldsymbol{y}}^n$ for the morphism $\mathcal{L}_{n,d} \to \mathbb{P}_{\boldsymbol{y}}^n$, i.e. such that the degree of its fiber equals the ML degree of $F_{n,d}$. We assume that $|\boldsymbol{u}| \neq 0$ and $u_i \neq 0$, for all i. We prove that the fiber of the point $[|\boldsymbol{u}| : \ldots : |\boldsymbol{u}|] = \mathbf{1} \in \mathbb{P}_{\boldsymbol{y}}^n$ has the same degree. Consider the affine line in $\mathbb{P}_{\boldsymbol{y}}^n$ passing through $\boldsymbol{u}$ and $\mathbf{1}$, which is the image of the map $\phi : \mathbb{A}^1 \to \mathbb{P}_{\boldsymbol{y}}^n$ induced by the ring homomorphism

$$
\begin{aligned}
\mathbb{C}[y_0, \ldots, y_n] &\longrightarrow \quad \mathbb{C}[t] \\
y_i &\longmapsto u_i t + |\boldsymbol{u}|(1 - t).
\end{aligned}
\tag{8}
$$

We will show that the induced subfamily of the family $\mathcal{X} := \mathrm{Proj}\left(\mathbb{C}[y_0, \ldots, y_n]/I_{n,d}^{\boldsymbol{y}}\right) \to \mathbb{P}_{\boldsymbol{y}}^n$

$$
\begin{array}{ccc}
\mathbb{A}^1 \times_{\mathbb{P}_{\boldsymbol{y}}^n} \mathcal{X} & \longrightarrow & \mathcal{X} \\
\downarrow & & \downarrow \\
\mathbb{A}^1 & \xrightarrow{\ \phi\ } & \mathbb{P}_{\boldsymbol{y}}^n
\end{array}
$$

is flat, i.e. all the fibers have the same degree (see [10, III, Theorem 9.9]).

To prove such a property, we will need a flatness criterion for filtered modules. Thus, we briefly recall few features of filtered modules (see [6, Chapter 5]) for the particular case we are dealing with. Let us consider a polynomial ring R and its irrelevant ideal $\mathfrak{m}$. The $\mathfrak{m}$-adic filtration of R is the descending multiplicative filtration of ideals

$$
R \supset \mathfrak{m} \supset \mathfrak{m}^2 \supset \cdots \supset \mathfrak{m}^\ell \supset \cdots
$$

that induces the standard graded structure of R by considering the direct sum

$$
\mathrm{gr}\, R := \bigoplus_{\ell \geqslant 0} \mathfrak{m}^\ell / \mathfrak{m}^{\ell+1} = \bigoplus_{\ell \geqslant 0} R_\ell.
$$

The same construction extends to any R-module M. The $\mathfrak{m}$-adic filtration of M is

$$
M \supset \mathfrak{m}M \supset \mathfrak{m}^2 M \supset \cdots \supset \mathfrak{m}^\ell M \supset \cdots
$$

and the associated graded $\mathrm{gr}\, R$-module is

$$
\mathrm{gr}\, M := \bigoplus_{\ell \geqslant 0} \mathfrak{m}^\ell M / \mathfrak{m}^{\ell+1} M = M/\mathfrak{m}M \oplus \mathfrak{m}M/\mathfrak{m}^2 M \oplus \cdots
$$

For any $f \in M$, we define the *initial form of f* to be the element

$$\text{in}(f) := f \bmod \mathfrak{m}^{\ell+1}M \subset \mathfrak{m}^\ell M/\mathfrak{m}^{\ell+1}M$$

where ℓ is the greatest index such that $f \in \mathfrak{m}^\ell M$. If M' is a submodule of M, we can consider the $(\mathfrak{gr}\,R)$-submodule of $\mathfrak{gr}\,M$ generated by the elements $\text{in}(f)$, $\forall\, f \in M'$. In particular, for any ideal $J \subset R$, we have $\mathfrak{gr}(R/J) = (\mathfrak{gr}\,R)/\text{in}(J) = R/\text{in}(J)$ (see [6, Exercise 5.3]).

Theorem 3 ([17, Theorem 2.20], [1, Proposition 3.12]). *Let S be a quotient of a polynomial ring and let M be a S-module. If $\mathfrak{gr}\,M$ is flat over $\mathfrak{gr}\,S$, then M is flat over S.*

With the help of these new tools, we can prove a crucial result for this paper.

Lemma 3. *The family $\mathbb{A}^1 \times_{\mathbb{P}^n_{\boldsymbol{y}}} \mathcal{X} \to \mathbb{A}^1$ induced by the morphism (8) is flat.*

Proof. The family induced by (8) is described by the ideal

$$I_{n,d}^{\boldsymbol{u},t} := (F_{n,d}, \mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{u},t}),\ \forall\, 0 \leqslant i < j < k \leqslant n)$$

(without saturation by $x_0 + \cdots + x_n$), where

$$\mathcal{M}_{n,d}^{\boldsymbol{u},t} := \begin{pmatrix} |\boldsymbol{u}| - \hat{u}_0 t & \dots & |\boldsymbol{u}| - \hat{u}_i t & \dots & |\boldsymbol{u}| - \hat{u}_n t \\ x_0 & \dots & x_i & \dots & x_n \\ x_0^d & \dots & x_i^d & \dots & x_n^d \end{pmatrix}.$$

We show that the module $\mathbb{C}[x_0, \dots, x_n, t]/I_{n,d}^{\boldsymbol{u},t}$ is flat over $\mathbb{C}[t]$ via Theorem 3, i.e. proving that the module $\mathfrak{gr}\left(\mathbb{C}[x_0, \dots, x_n, t]/I_{n,d}^{\boldsymbol{u},t}\right)$ is flat over $\mathfrak{gr}\,\mathbb{C}[t] = \mathbb{C}[t]$. Since we have

$$\mathfrak{gr}\left(\mathbb{C}[x_0, \dots, x_n, t]/I_{n,d}^{\boldsymbol{u},t}\right) = \mathfrak{gr}\,\mathbb{C}[x_0, \dots, x_n, t]/\text{in}\big(I_{n,d}^{\boldsymbol{u},t}\big) = \mathbb{C}[x_0, \dots, x_n, t]/\text{in}\big(I_{n,d}^{\boldsymbol{u},t}\big),$$

we focus on the ideal $I_{n,d}^{\boldsymbol{u},t}$ and the corresponding $\text{in}(I_{n,d}^{\boldsymbol{u},t})$. By linearity, we can split each minor $\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{u},t})$ in the following way

$$\begin{vmatrix} |\boldsymbol{u}| - \hat{u}_i t & |\boldsymbol{u}| - \hat{u}_j t & |\boldsymbol{u}| - \hat{u}_k t \\ x_i & x_j & x_k \\ x_i^d & x_j^d & x_k^d \end{vmatrix} = |\boldsymbol{u}| \begin{vmatrix} 1 & 1 & 1 \\ x_i & x_j & x_k \\ x_i^d & x_j^d & x_k^d \end{vmatrix} - t \begin{vmatrix} \hat{u}_i & \hat{u}_j & \hat{u}_k \\ x_i & x_j & x_k \\ x_i^d & x_j^d & x_k^d \end{vmatrix} =$$
$$= |\boldsymbol{u}|\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{1}}) - t\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\hat{\boldsymbol{u}}})$$

and we can also deduce that $\text{in}\big(|\boldsymbol{u}|\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{1}}) - t\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\hat{\boldsymbol{u}}})\big) = |\boldsymbol{u}|\mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{1}})$. Hence,

$$I_{n,d}^{\boldsymbol{1}} = \big(F_{n,d}, \mathrm{e}_{ijk}(\mathcal{M}_{n,d}^{\boldsymbol{1}}),\ \forall\, 0 \leqslant i < j < k \leqslant n\big) \subseteq \text{in}\big(I_{n,d}^{\boldsymbol{u},t}\big).$$

To prove that in fact equality holds, we can study the relations among these polynomials. The ideal $I_{n,d}^{\boldsymbol{1}}$ is the ideal of initial forms of $I_{n,d}^{\boldsymbol{u},t}$ if, and only if, each non-trivial syzygy

between a pair of generators of $I^1_{n,d}$ can be lifted to a syzygy of $I^{u,t}_{n,d}$. The polynomial defining the Fermat hypersurface is irreducible and it has no non-trivial syzygies with the other generators of $I^1_{n,d}$. The other generators have four irreducible factors

$$\mathrm{e}_{ijk}(\mathcal{M}^1_{n,d}) = (x_i - x_j)(x_j - x_k)(x_k - x_i) \sum_{\substack{(e_i,e_j,e_k)\in\mathbb{Z}^3_{\geqslant 0} \\ e_i+e_j+e_k=d-2}} x_i^{e_i} x_j^{e_j} x_k^{e_k}$$

so that two distinct generators $\mathrm{e}_{ijk}(\mathcal{M}^1_{n,d})$ and $\mathrm{e}_{i'j'k'}(\mathcal{M}^1_{n,d})$ have a non-trivial syzygy if, and only if, two of the three indices are equal. Let us consider $\mathrm{e}_{ijk}(\mathcal{M}^1_{n,d})$ and $\mathrm{e}_{ijh}(\mathcal{M}^1_{n,d})$. Without loss of generality, we may assume $i < j < k < h$. The corresponding syzygy is

$$\frac{\mathrm{e}_{ijh}(\mathcal{M}^1_{n,d})}{x_i - x_j}\mathrm{e}_{ijk}(\mathcal{M}^1_{n,d}) - \frac{\mathrm{e}_{ijk}(\mathcal{M}^1_{n,d})}{x_i - x_j}\mathrm{e}_{ijh}(\mathcal{M}^1_{n,d}) = 0.$$

By direct computation, it is possible to check that the syzygy between these two generators of $I^1_{n,d}$ can be lifted to the syzygy

$$\frac{\mathrm{e}_{ijh}(\mathcal{M}^1_{n,d})}{x_i - x_j}\mathrm{e}_{ijk}(\mathcal{M}^{u,t}_{n,d}) - \frac{\mathrm{e}_{ijk}(\mathcal{M}^1_{n,d})}{x_i - x_j}\mathrm{e}_{ijh}(\mathcal{M}^{u,t}_{n,d}) =$$

$$= |\boldsymbol{u}|x_i x_j \left(\sum_{\substack{(e_i,e_j)\in\mathbb{Z}^2_{\geqslant 0} \\ e_i+e_j=d-2}} x_i^{e_i} x_j^{e_j} \right) \left(-\mathrm{e}_{ijk}(\mathcal{M}^{u,t}_{n,d}) + \mathrm{e}_{ijh}(\mathcal{M}^{u,t}_{n,d}) - \mathrm{e}_{ikh}(\mathcal{M}^{u,t}_{n,d}) + \mathrm{e}_{jkh}(\mathcal{M}^{u,t}_{n,d}) \right)$$

among the generators of $I^{u,t}_{n,d}$. Finally, $\mathbb{C}[x_0,\dots,x_n,t]/I^{u,t}_{n,d}$ is flat, as the graded module $\mathfrak{gr}\big(\mathbb{C}[x_0,\dots,x_n,t]/I^{u,t}_{n,d}\big) = \mathbb{C}[x_0,\dots,x_n,t]/I^1_{n,d}$ is free over $\mathbb{C}[t]$ (the quotient does not depend on t).

Theorem 4.

$$\mathrm{MLdeg}\, F_{n,d} = \deg\left(I^1_{n,d} \setminus \mathcal{H}\right) = \deg I^1_{n,d} - \deg\left(I^1_{n,d} + (x_0 + \cdots + x_n)\right). \qquad (9)$$

Moreover, for every point $\boldsymbol{v} \in \mathbb{P}^n_{\boldsymbol{y}}$ such that $|\boldsymbol{v}| \neq 0$,

$$\mathrm{MLdeg}\, F_{n,d} = \deg I^v_{n,d} - \deg\left(I^v_{n,d} + (x_0 + \cdots + x_n)\right).$$

Proof. Let $\boldsymbol{u} \in \mathbb{P}^n_{\boldsymbol{y}}$ be a generic point for the projection $\pi_{\boldsymbol{y}} : \mathcal{L}_{n,d} \to \mathbb{P}^n_{\boldsymbol{y}}$, i.e. such that the length of the fiber $\pi^{-1}_{\boldsymbol{y}}(\boldsymbol{u})$ is equal to the ML degree of $F_{n,d}$. We may assume by Proposition 1 that $\mathrm{Proj}\,\mathbb{C}[x]/I^{\boldsymbol{u}}_{n,d}$ is smooth along the hyperplane $x_0 + \cdots + x_n = 0$, $|\boldsymbol{u}| \neq 0$ and $u_i \neq 0$, $\forall\, i = 0,\dots,n$. Proving (9) is equivalent to prove $\deg I^{\boldsymbol{u}}_{n,d} = \deg I^1_{n,d}$. Indeed, by Lemma 1 and Proposition 1 (Remark 1), we have $\deg(I^{\boldsymbol{u}}_{n,d} \setminus \mathcal{H}) = \deg I^{\boldsymbol{u}}_{n,d} - \deg\left(I^{\boldsymbol{u}}_{n,d} + (x_0 + \cdots + x_n)\right)$, $\deg(I^1_{n,d} \setminus \mathcal{H}) = \deg I^1_{n,d} - \deg\left(I^1_{n,d} + (x_0 + \cdots + x_n)\right)$ and $I^{\boldsymbol{u}}_{n,d} + (x_0 + \cdots + x_n) = I^1_{n,d} + (x_0 + \cdots + x_n)$ by Lemma 2. The equality $\deg I^{\boldsymbol{u}}_{n,d} = \deg I^1_{n,d}$

follows from Lemma 3. In fact, the ideals $I_{n,d}^{\mathbf{1}}$ and $I_{n,d}^{\boldsymbol{u}}$ correspond respectively to the fibers over $t = 0$ and $t = 1$ of the family $\mathbb{A}^1 \times_{\mathbb{P}_{\boldsymbol{y}}^n} \mathcal{X} \to \mathbb{A}^1$ and in the case of a family of zero-dimensional schemes with an irreducible reduced base, flatness is equivalent to the fact that all the fibers over closed points have the same degree [10, III, Theorem 9.9].

To prove the last part of the statement, it suffices to repeat the same reasoning starting with MLdeg $F_{n,d} = \deg I_{n,d}^{\mathbf{1}} - \deg \left(I_{n,d}^{\mathbf{1}} + (x_0 + \cdots + x_n) \right)$.

Example 1. *Consider the Fermat surface $F_{3,3} \subset \mathbb{P}^3$. Its ML degree is 30 and we can determine it by computing the multidegree of the likelihood correspondence $\mathcal{L}_{3,3} \subset \mathbb{P}_{\boldsymbol{x}}^3 \times \mathbb{P}_{\boldsymbol{y}}^3$. Now, we look at the behavior at the points $\mathbf{1} = [1 : 1 : 1 : 1]$, $\boldsymbol{u} = [2 : 3 : 5 : 6]$, $\boldsymbol{v}_1 = [2 : 3 : -7 : 7]$, $\boldsymbol{v}_2 = [2 : 3 : 5 : 0]$, $\boldsymbol{v}_3 = [1 : 3 : -6 : 2]$ and $\boldsymbol{v}_4 = [1 : -1 : 1 : -1]$. By Theorem 4, we know that*

$$30 = \text{MLdeg}\, F_{3,3} = \deg \left(I_{3,3}^{\mathbf{1}} \setminus \mathcal{H} \right) = \deg I_{3,3}^{\mathbf{1}} - \deg \left(I_{3,3}^{\mathbf{1}} + (x_0 + x_1 + x_2 + x_3) \right) =$$
$$= \deg I_{3,3}^{\boldsymbol{u}} - \deg \left(I_{3,3}^{\boldsymbol{u}} + (x_0 + x_1 + x_2 + x_3) \right) =$$
$$= \deg I_{3,3}^{\boldsymbol{v}_1} - \deg \left(I_{3,3}^{\boldsymbol{v}_1} + (x_0 + x_1 + x_2 + x_3) \right) =$$
$$= \deg I_{3,3}^{\boldsymbol{v}_2} - \deg \left(I_{3,3}^{\boldsymbol{v}_2} + (x_0 + x_1 + x_2 + x_3) \right),$$

as $|\boldsymbol{u}| \neq 0$, $|\boldsymbol{v}_1| \neq 0$ and $|\boldsymbol{v}_2| \neq 0$. However, the theorem says nothing about the critical points of the likelihood functions $\ell_{\boldsymbol{u}}$, $\ell_{\boldsymbol{v}_1}$, $\ell_{\boldsymbol{v}_2}$ not lying over $\mathcal{H}$. By direct computation, we can check that $\boldsymbol{u}$ is "generic" for the projection map $\pi_{\boldsymbol{y}} : \mathcal{L}_{3,3} \to \mathbb{P}_{\boldsymbol{y}}^3$, i.e. $\deg I_{3,3}^{\boldsymbol{u}} - \deg \left(I_{3,3}^{\boldsymbol{u}} + (x_0 + x_1 + x_2 + x_3) \right) = \deg \left(I_{3,3}^{\boldsymbol{u}} \setminus \mathcal{H} \right)$, whereas $\ell_{\boldsymbol{v}_1}$ and $\ell_{\boldsymbol{v}_2}$ have more critical points lying on $\mathcal{H}$. More precisely, $\deg \left(I_{3,3}^{\boldsymbol{v}_1} \setminus \mathcal{H} \right) = 28$ and $\deg \left(I_{3,3}^{\boldsymbol{v}_2} \setminus \mathcal{H} \right) = 21$. To understand this behavior, we notice that the data vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ do not satisfy the assumption of Theorem 2 and that $\boldsymbol{v}_2$ does not even satisfy the hypothesis of Lemma 1.

As $|\boldsymbol{v}_3| = |\boldsymbol{v}_4| = 0$, the last two points do not even satisfy the weaker assumption of Theorem 4 and can not be use to determine the ML degree of $F_{3,3}$. In the case of $\boldsymbol{v}_3$, we check that $\ell_{\boldsymbol{v}_3}$ has a finite number of critical points and that $\deg I_{3,3}^{\boldsymbol{v}_3} = \deg I_{3,3}^{\mathbf{1}}$ (even if we can not deduce it from Lemma 3), but Lemma 2 does not apply. Indeed, $\deg \left(I_{3,3}^{\boldsymbol{v}_3} + (x_0 + x_1 + x_2 + x_3) \right) = 12 \neq \deg \left(I_{3,3}^{\mathbf{1}} + (x_0 + x_1 + x_2 + x_3) \right)$. The case of $\boldsymbol{v}_4$ is even more special as the ideal $I_{3,3}^{\boldsymbol{v}_4}$ describing the entire set of critical points of $\ell_{\boldsymbol{v}_4}$ defines a pair of lines plus 22 points. These 22 points represent the set of critical points of $\ell_{\boldsymbol{v}_4}$ not lying on $\mathcal{H}$.

4. Determine solutions

The result of the previous section allows us to restrict our attention to the ideal $I_{n,d}^{\mathbf{1}}$ generated by $f_{n,d}$ and by the 3×3 minors of the matrix

$$\mathcal{M}_{n,d}^{\mathbf{1}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_0 & x_1 & \dots & x_n \\ x_0^d & x_1^d & \dots & x_n^d \end{pmatrix}.$$

It is clear that if two coordinates coincide, then any 3×3-minor of $\mathcal{M}_{n,d}^{\mathbf{1}}$ involving the two corresponding columns vanishes. Furthermore, we know that there are at least two distinct coordinates, since the point $[1 : \ldots : 1]$ does not lie on the Fermat hypersurface $F_{n,d}$. The next natural question is about finding the maximal number of distinct coordinates for a critical point. Consider a critical point with at least three distinct coordinates p_i, p_j, p_k. Since the 3×3 minors of $\mathcal{M}_{n,d}^{\mathbf{1}}$ split into factors

$$
\det \begin{pmatrix} 1 & 1 & 1 \\ x_i & x_j & x_k \\ x_i^d & x_j^d & x_k^d \end{pmatrix} = (x_i - x_j)(x_j - x_k)(x_k - x_i) \sum_{\substack{(e_i,e_j,e_k)\in\mathbb{Z}_{\geqslant 0}^3 \\ e_i+e_j+e_k=d-2}} x_i^{e_i} x_j^{e_j} x_k^{e_k}, \tag{10}
$$

any 3-tuple of distinct coordinates of a critical point is a solution to the polynomial

$$
\sum_{\substack{(e_i,e_j,e_k)\in\mathbb{Z}_{\geqslant 0}^3 \\ e_i+e_j+e_k=d-2}} x_i^{e_i} x_j^{e_j} x_k^{e_k}. \tag{11}
$$

Lemma 4. *Consider the Fermat hypersurface of degree d. The coordinates of each solution of $I_{n,d}^{\mathbf{1}} \setminus \mathcal{H}$ can take at most $\min\{d, n+1\}$ distinct values.*

Proof. Consider k distinct values $p_1, \ldots, p_k$ and assume they appear as coordinates of a critical point. Clearly $k \leqslant n+1$. Furthermore, we see from polynomial (11) that once we fix p_1 and p_2, the other values $p_3, \ldots, p_k$ have to be roots of the univariate polynomial

$$
\sum_{e_k=0}^{d-2} \left(\sum_{e_i+e_j=d-2-e_k} p_1^{e_i} p_2^{e_j} \right) x_k^{e_k},
$$

which has degree $d-2$. Thus, k can be at most d. ∎

Definition 2. *Let $\mathsf{a} := (a_1, \ldots, a_s)$ be an integer partition of $n+1$ such that $a_1 \geqslant \cdots \geqslant a_s \geqslant 1$ and $2 \leqslant s \leqslant \min\{d, n+1\}$. We say that a critical point $\boldsymbol{p}$ of the ML degree problem of the Fermat hypersurface $F_{n,d}$ is of type $(a_1, \ldots, a_s)$, or a $(a_1, \ldots, a_s)$-critical point, if $\boldsymbol{p}$ has s distinct coordinates and each distinct value p_i appears a_i times as coordinate of the point, for $1 \leqslant i \leqslant s$.*

Let us recall some standard notation for partitions. If some integer b is repeated α times in a, we will write b^α instead of $b, \ldots, b$. Moreover, for any partition a, we denote by α_{a} the set of "exponents". For instance, if $\mathsf{a} = (4, 3, 3, 3, 2, 1, 1) = (4, 3^3, 2, 1^2)$, then $\alpha_{\mathsf{a}} = (1, 3, 1, 2)$. The length of a partition a is the sum of all exponents $|\alpha_{\mathsf{a}}|$. In the previous example the partition has length 7.

Let us define $\mathcal{P}_{n+1,d}$ as the set of partitions of $n+1$ of length at most $\min\{d, n+1\}$. We will present each element in $\mathcal{P}_{n+1,d}$ as a vector whose components form a sequence of weakly decreasing positive integers.

Using the notation we introduced above, the property stated in Lemma 4 translates in the following formula to compute the ML degree of the Fermat hypersurface:

$$\text{MLdeg}\, F_{n,d} = \sum_{\mathsf{a}\in\mathcal{P}_{n+1,d}} \#\{\mathsf{a}\text{-critical points}\}.$$

Let $\mathsf{a} = (a_1,\ldots,a_s)$ be a partition in $\mathcal{P}_{n+1,d}$ and let $\alpha_{\mathsf{a}} = (\alpha_1,\ldots,\alpha_\sigma)$ be the corresponding set of exponents. We consider the integers

$$c_{\mathsf{a}} := \binom{n+1}{a_1} \cdot \binom{n+1-a_1}{a_2} \cdot \ldots \cdot \binom{n+1-(a_1+\ldots+a_{s-2})}{a_{s-1}} \text{ and } \alpha_{\mathsf{a}}! := \alpha_1! \cdot \ldots \cdot \alpha_\sigma!.$$

In addition, we want to determine the ideal that defines the subset of a-critical points of $F_{n,d}$ by specifying the identifications among the coordinates. We can compute this ideal from $I^{\mathbf{1}}_{n,d}$ by adding the linear form $x_i - x_j$ if the i-th and j-th coordinates have to be equal and by saturating with respect to the linear form $x_k - x_h$ if the k-th and h-th coordinates have to be different. The identifications allow to rewrite the ideal in terms of s variables, say $z_1,\ldots,z_s$, and the ideal written in terms of these variables (up to relabeling of the variables) does not depend on the identifications. Thus, we define

$$I_{\mathsf{a}}^d := \left(a_1 z_1^d + \cdots + a_s z_s^d, \mathrm{e}_{ijk}(\mathcal{M}^{\mathbf{1}}_{s-1,d}), \forall\, 1 \leqslant i < j < k \leqslant s\right) : \left(\prod_{i<j}(z_i - z_j)\right),$$

$$I_{\mathsf{a}}^d \setminus \mathcal{H} := I_{\mathsf{a}}^d : (a_1 z_1 + \cdots + a_s z_s), \qquad I_{\mathsf{a}}^d \cap \mathcal{H} := I_{\mathsf{a}}^d + (a_1 z_1 + \cdots + a_s z_s),$$

$$(12)$$

where $\mathrm{e}_{ijk}(\mathcal{M}^{\mathbf{1}}_{s-1,d})$ is the polynomial (10) in the variables z_i, z_j and z_k.

Theorem 5.

$$\text{MLdeg}\, F_{n,d} = \sum_{\mathsf{a}\in\mathcal{P}_{n+1,d}} c_{\mathsf{a}} \frac{\deg\left(I_{\mathsf{a}}^d \setminus \mathcal{H}\right)}{\alpha_{\mathsf{a}}!} = \sum_{\mathsf{a}\in\mathcal{P}_{n+1,d}} c_{\mathsf{a}} \frac{\deg I_{\mathsf{a}}^d - \deg\left(I_{\mathsf{a}}^d \cap \mathcal{H}\right)}{\alpha_{\mathsf{a}}!}$$

Proof. In order to prove the statement, we show that the number of a-critical points equals $c_{\mathsf{a}} \frac{\deg(I_{\mathsf{a}}^d \setminus \mathcal{H})}{\alpha_{\mathsf{a}}!}$ for each $\mathsf{a} \in \mathcal{P}_{n+1,d}$.

Let $\boldsymbol{p} = [p_1 : \ldots : p_s]$ be a solution of the ideal $I_{\mathsf{a}}^d \setminus \mathcal{H}$. It corresponds to the a-critical points $\boldsymbol{q} = [q_0 : \ldots : q_n]$ having a_i coordinates equal to p_i (up to a fixed scalar) for all i. As the symmetric group $\mathcal{S}_{n+1}$ acts on the solutions of $I^{\mathbf{1}}_{n,d} \setminus \mathcal{H}$, the number of points $\boldsymbol{q}$ corresponding to $\boldsymbol{p}$ is equal to the number of ways of partitioning the set of indices $\{0,\ldots,n\}$ in s subsets with cardinality given by the entries of a. This is exactly c_{a}.

To complete the proof we need to determine how many "different" solutions the ideal $I_{\mathsf{a}}^d \setminus \mathcal{H}$ has. For instance, if $a_1 = a_2$ and $\boldsymbol{p} = [p_1 : p_2 : \ldots : p_s]$ is a solution of the ideal $I_{\mathsf{a}}^d \setminus \mathcal{H}$, then also $\boldsymbol{p}' = [p_2 : p_1 : \ldots : p_s]$ is a solution. But $\boldsymbol{p}$ and $\boldsymbol{p}'$ correspond to the same subset of a-critical points with $n+1$ coordinates since we are considering the action of $\mathcal{S}_{n+1}$ on them. Looking at the definition (12) of $I_{\mathsf{a}}^d \setminus \mathcal{H}$, we notice that the ideals do not change if we swap two variables z_i, z_j such that $a_i = a_j$. The repetitions of the entries of a are counted by the sequence α_{a} and $I_{\mathsf{a}}^d \setminus \mathcal{H}$ is invariant under the action of the group $\mathcal{S}_{\alpha_1} \times \cdots \times \mathcal{S}_{\alpha_\sigma}$. Hence,

Table 1: The a-critical points of the quartic Fermat hypersurface $F_{n,4}$ for $n = 2, 3, 4$.

$a \in \mathcal{P}_{3,4}$	$\deg I_a^4$	$\deg\left(I_a^4 \cap \mathcal{H}\right)$	c_a	$\alpha_a!$	a-critical points
$(2,1)$	4	0	3	1	12
$(1,1,1)$	8	2	6	6	6

				MLdeg $F_{2,4}$	18

$a \in \mathcal{P}_{4,4}$	$\deg I_a^4$	$\deg\left(I_a^4 \cap \mathcal{H}\right)$	c_a	$\alpha_a!$	a-critical points
$(3,1)$	4	0	4	1	16
$(2,2)$	4	0	6	2	12
$(2,1,1)$	8	0	12	2	48
$(1,1,1,1)$	8	8	24	24	0

				MLdeg $F_{3,4}$	76

$a \in \mathcal{P}_{5,4}$	$\deg I_a^4$	$\deg\left(I_a^4 \cap \mathcal{H}\right)$	c_a	$\alpha_a!$	a-critical points
$(4,1)$	4	0	5	1	20
$(3,2)$	4	0	10	1	40
$(3,1,1)$	8	0	20	2	80
$(2,2,1)$	8	0	30	2	120
$(2,1,1,1)$	8	2	60	6	60

				MLdeg $F_{4,4}$	320

the solutions we are interested in are the solutions of $I_a^d \setminus \mathcal{H}$ modulo the action of this group of symmetries, i.e. $\deg\left(I_a^d \setminus \mathcal{H}\right)/\alpha_a!$. The equality $\deg\left(I_a^d \setminus \mathcal{H}\right) = \deg I_a^d - \deg\left(I_a^d \cap \mathcal{H}\right)$ follows directly from the equality $\deg\left(I_{n,d}^{\mathbf{1}} \setminus \mathcal{H}\right) = \deg I_{n,d}^{\mathbf{1}} - \deg\left(I_{n,d}^{\mathbf{1}} + (x_0 + \cdots + c_n)\right)$.

We conclude this section with a more detailed description of the ideal I_a^d, where we try to make the saturation by the linear forms $z_i - z_j$ more explicit. Let us introduce some notation that will be useful later. For any set of indices $\mathcal{I} = \{i_1, \ldots, i_r\}$ of variables z_j and for any non-negative integer m, let

$$\boldsymbol{z}_{\mathcal{I}}^{(0)} := 1 \qquad \text{and} \qquad \boldsymbol{z}_{\mathcal{I}}^{(m)} := \sum_{\substack{(e_1,\ldots,e_r)\in\mathbb{Z}_{\geqslant 0}^r \\ e_1+\cdots+e_r=m}} z_{i_1}^{e_1} \cdots z_{i_r}^{e_r}, \text{ for } m > 0.$$

Lemma 5. *Consider a set of s indices $\mathcal{I}$ and fix two distinct elements h and k in $\mathcal{I}$. For every positive integer m,*

$$\boldsymbol{z}_{\mathcal{I}\setminus k}^{(m)} - \boldsymbol{z}_{\mathcal{I}\setminus h}^{(m)} = (z_h - z_k)\boldsymbol{z}_{\mathcal{I}}^{(m-1)}.$$

Proof. By looking at $\boldsymbol{z}_{\mathcal{I}\setminus k}^{(m)}$ as an univariate polynomial in the variable z_h and $\boldsymbol{z}_{\mathcal{I}\setminus h}^{(m)}$ as

an univariate polynomial in the variable z_k, we can write

$$\boldsymbol{z}_{\mathcal{I}\backslash k}^{(m)} = \sum_{e=0}^{m} \left(\boldsymbol{z}_{\mathcal{I}\backslash h,k}^{(m-e)}\right) z_h^e \qquad \text{and} \qquad \boldsymbol{z}_{\mathcal{I}\backslash h}^{(m)} = \sum_{e=0}^{m} \left(\boldsymbol{z}_{\mathcal{I}\backslash h,k}^{(m-e)}\right) z_k^e.$$

Hence,

$$\boldsymbol{z}_{\mathcal{I}\backslash k}^{(m)} - \boldsymbol{z}_{\mathcal{I}\backslash h}^{(m)} = \sum_{e=0}^{m} \left[\left(\boldsymbol{z}_{\mathcal{I}\backslash h,k}^{(m-e)}\right)(z_h^e - z_k^e)\right] =$$

$$= (z_h - z_k) \sum_{e=1}^{m} \left[\left(\boldsymbol{z}_{\mathcal{I}\backslash h,k}^{(m-e)}\right) \boldsymbol{z}_{h,k}^{(e-1)}\right] = (z_h - z_k)\boldsymbol{z}_{\mathcal{I}}^{(m-1)}.$$

Proposition 2. *Let* $\mathsf{a} = (a_1, \ldots, a_s)$ *be a partition in* $\mathcal{P}_{n+1,d}$.

   (i) *If* $s = 2$, *then* $I_{\mathsf{a}}^d = (a_1 z_1^d + a_2 z_2^d)$.

   (ii) *If* $s > 2$, *then* I_{a}^d *contains the polynomials* $\boldsymbol{z}_{\mathcal{I}}^{(d-w+1)}$ *for all multi-indices* $\mathcal{I} \subset \{1, \ldots, s\}$ *with* w *elements, for all* $w = 3, \ldots, s$.

   Proof. (i) If the coordinates of a critical point have only two distinct values, then the 3×3 minors of $\mathcal{M}_{n,d}^{\mathbf{1}}$ automatically vanish because all 3×3 submatrices of $\mathcal{M}_{n,d}^{\mathbf{1}}$ have at most 2 linearly independent columns. Moreover, the unique solution to $a_1 z_1^d + a_2 z_2^d$ with $z_1 = z_2$ is given by the pair $(0,0)$ that does not define a point. Therefore, we can avoid the saturation with respect to $z_1 - z_2$.

   (ii) Let us proceed iteratively: we start with the generators of I_{a}^d

$$\mathrm{e}_{ijk}\left(\mathcal{M}_{s-1,d}^{\mathbf{1}}\right) = (z_i - z_j)(z_j - z_k)(z_k - z_i)\boldsymbol{z}_{i,j,k}^{(d-2)}.$$

Since we compute these polynomials for every 3-tuple i, j, k and we saturate with respect to all linear forms given by the difference of two variables, the statement is true for $w = 3$. For $3 < w \leqslant s$, it suffices to apply Lemma 5.

5. Closed formulas for the ML degree of special Fermat hypersurfaces

   In this section, we give two closed formulas for computing the ML degree of the Fermat hypersurfaces $F_{n,2}$ and $F_{2,d}$. In the case of $F_{n,2}$, the set $\mathcal{P}_{n+1,2}$ contains only partitions a of length 2 and the ideal I_{a}^2 is very simple (Proposition 2(i)), while in the case $F_{2,d}$ we can successfully apply a topological criterion which works because we have only 3 variables.

5.1. The ML degree of $F_{n,2}$

We start counting the number of solutions of the ideal $I_\mathsf{a}^d \setminus \mathcal{H}$ for partitions $\mathsf{a} = (a_1, a_2)$ of length 2.

Lemma 6. *Let* $\mathsf{a} = (a_1, a_2)$ *be a partition of* $\mathcal{P}_{n+1,d}$.

$$\deg\left(I_\mathsf{a}^d \setminus \mathcal{H}\right) = \begin{cases} d-1, & \text{if } a_1 = a_2 \text{ and } d \text{ odd,} \\ d, & \text{otherwise.} \end{cases}$$

Proof. We start considering the solutions of the equation $a_1 z_1^d + a_2 z_2^d = 0$. Without loss of generality, we can assume $z_2 = 1$, so that we have d solutions

$$\left[\sqrt[d]{\frac{a_2}{a_1}}\, \mathrm{e}^{\mathrm{i}\frac{2k+1}{d}\pi} : 1\right], \qquad k = 0, \ldots, d-1.$$

We need to discard every solution lying on the line $a_1 z_1 + a_2 z_2 = 0$. Such a point exists only in the case $a_1 = a_2$ and $\mathrm{e}^{\mathrm{i}\frac{2k+1}{d}\pi} = -1$, i.e. $d = 2k+1$.

We now apply the strategy used in the proof of Theorem 5 to count the (a_1, a_2)-critical points of $I_{n,d}^1 \setminus \mathcal{H}$ for all $(a_1, a_2) \in \mathcal{P}_{n+1,d}$.

Proposition 3. *The total number of* a-*critical points of the ideal* $I_{n,d}^1 \setminus \mathcal{H}$ *as* a *varies over all pairs* $(a_1, a_2) \in \mathcal{P}_{n+1,d}$ *is*

$$d(2^n - 1) - \frac{1}{2}\binom{n+1}{\frac{n+1}{2}} \text{ if } d \text{ and } n \text{ are odd,} \qquad d(2^n - 1) \text{ otherwise.} \tag{13}$$

Proof. If $n+1$ is odd, for each partition $(a_1, a_2) \in \mathcal{P}_{n+1,d}$ of length 2, we have $\alpha_{(a_1,a_2)}! = 1$ (as $a_1 > a_2$) and $\deg\left(I_{(a_1,a_2)}^d \setminus \mathcal{H}\right) = d$ (Lemma 6), so that the total number of (a_1, a_2)-solutions of $I_{n,d}^1 \setminus \mathcal{H}$ is

$$\sum_{(a_1,a_2)\in\mathcal{P}_{n+1,d}} c_{(a_1,a_2)} \frac{\deg\left(I_{(a_1,a_2)}^d \setminus \mathcal{H}\right)}{\alpha_{(a_1,a_2)}!} = \sum_{\substack{a_1+a_2=n+1 \\ a_1 > a_2 > 0}} d\binom{n+1}{a_1} = d\sum_{i=1}^{n/2}\binom{n+1}{i} =$$

$$= d\frac{1}{2}\sum_{i=1}^{n}\binom{n+1}{i} = d\frac{1}{2}\left(\sum_{i=0}^{n+1}\binom{n+1}{i} - 2\right) = d(2^n - 1).$$

If $n+1$ is even, there is also the partition $(\frac{n+1}{2}, \frac{n+1}{2})$ for which $\alpha_{(\frac{n+1}{2}, \frac{n+1}{2})}! = 2$. Assume

d even, so that $\deg\left(I^d_{(\frac{n+1}{2},\frac{n+1}{2})}\setminus\mathcal{H}\right)=d$. The total number of (a_1,a_2)-critical points is

$$\sum_{(a_1,a_2)\in\mathcal{P}_{n+1,d}}c_{(a_1,a_2)}\frac{\deg\left(I^d_{(a_1,a_2)}\setminus\mathcal{H}\right)}{\alpha_{(a_1,a_2)}!}=\sum_{\substack{a_1+a_2=n+1\\a_1>a_2>0}}d\binom{n+1}{a_1}+\frac{d}{2}\binom{n+1}{\frac{n+1}{2}}=$$

$$=d\sum_{i=1}^{\frac{n-1}{2}}\binom{n+1}{i}+\frac{d}{2}\binom{n+1}{\frac{n+1}{2}}=\frac{d}{2}\sum_{i=1}^{n}\binom{n+1}{i}=$$

$$=\frac{d}{2}\left(\sum_{i=0}^{n}\binom{n+1}{i}-2\right)=d(2^n-1).$$

Finally, if d is odd, we have one less solution of $I^d_{\left(\frac{n+1}{2},\frac{n+1}{2}\right)}\setminus\mathcal{H}$ so that the total number of (a_1,a_2)-critical points is decreased by $c_{\left(\frac{n+1}{2},\frac{n+1}{2}\right)}/\alpha_{\left(\frac{n+1}{2},\frac{n+1}{2}\right)}!=\frac{1}{2}\binom{n+1}{\frac{n+1}{2}}$.

Corollary 1. *The ML degree of the Fermat hypersurface $F_{n,2}$ is $2^{n+1}-2$.*

Proof. If $d=2$, all the partitions of $\mathcal{P}_{n+1,2}$ have length 2, so that the number of (a_1,a_2)-critical points determined in Proposition 3 is exactly the ML degree of $F_{n,d}$.

Notice that this corollary proves that the Fermat hypersurface of degree 2 has the same ML degree of the generic hypersurface of degree 2 (cf. [16, Theorem 1.10, Example 1.11]). This is the expected result, as the quadric hypersurfaces are classified by the rank of the associated symmetric matrix and the quadric Fermat hypersurface has maximal rank as the generic quadric hypersurface. In general, this property does not hold. For instance, in the case discussed in Example 1, the Fermat surface $F_{3,3}$ has ML degree 30, while the ML degree of the generic cubic surface is 39. Notice that in order to obtain a cubic surface with ML degree equal to 39 it suffices to randomly choose the coefficients of the monomials in the equation of the Fermat surface.

5.2. The ML degree of $F_{2,d}$

Applying Lemma 6 and Theorem 5 to the case $n=2$, we obtain

$$\text{MLdeg }F_{2,d}=c_{(2,1)}\frac{\deg\left(I^d_{(2,1)}\setminus\mathcal{H}\right)}{\alpha_{(2,1)}!}+c_{(1,1,1)}\frac{\deg\left(I^d_{(1,1,1)}\setminus\mathcal{H}\right)}{\alpha_{(1,1,1)}!}=3d+\deg\left(I^d_{(1,1,1)}\setminus\mathcal{H}\right).$$

Furthermore, since in the case of partitions of length 3 there is a unique 3×3 minor, the solutions of $I^d_{(1,1,1)}\setminus\mathcal{H}$ are contained in the intersection of $z^{(d-2)}_{1,2,3}=0$ with $z^d_1+z^d_2+z^d_3=0$. By Bezout's theorem, we know that there are $d(d-2)$ solutions (counted with multiplicity). This gives the upper bound of the ML degree $3d+d(d-2)=d^2+d$.

In this case, instead of trying to determine explicitly the solutions of $I^d_{(1,1,1)}\setminus\mathcal{H}$, we resort to a special case of a theorem that correlates the ML degree of a variety with

its signed topological Euler characteristic (see [3, 14, 16]). To make the paper as self-contained as possible we report the theorem below. Recall that $\mathcal{H}$ denotes the distinguished arrangement of hyperplanes.

Theorem 6 ([16, Theorem 1.1]). *Let X be a smooth curve of degree d in $\mathbb{P}^2$, and $a = \#(X \cap \mathcal{H})$ the number of its points on the distinguished arrangement. Then the ML degree of X equals $d^2 - 3d + a$.*

As the Fermat curve is smooth, we need to calculate the number $a_d = \#(F_{2,d} \cap \mathcal{H})$ for $d > 1$.

Proposition 4. *The ML degree of the Fermat curve $F_{2,d}$ is*

$$\text{MLdeg}\, F_{2,d} = \begin{cases} d^2 + d, & \text{if } d \equiv 0, 2 \bmod 6, \\ d^2 + d - 3, & \text{if } d \equiv 3, 5 \bmod 6, \\ d^2 + d - 2, & \text{if } d \equiv 4 \bmod 6, \\ d^2 + d - 5, & \text{if } d \equiv 1 \bmod 6, \end{cases}$$

Proof. Let us examine the intersection of $F_{n,d}$ with $x_0 = 0$, $x_1 = 0$, $x_2 = 0$ and $x_0 + x_1 + x_0 = 0$. Notice that the points $[1:0:0]$, $[0:1:0]$ and $[0:0:1]$ (lying on two different lines $x_i = x_j = 0$) are not points of $F_{2,d}$. By Bezout's theorem, the intersection $F_{n,d} \cap \{x_0 = 0\}$ consists of d points counted with multiplicity. In this case, all the points are simple as they correspond to the solutions of the equation $x_1^d + x_2^d = 0$. By symmetry, we conclude that $\#(F_{2,d} \cap \{x_i = 0\}) = d$ for all $i = 0, 1, 2$.

Now consider the line $\{x_0 + x_1 + x_2 = 0\}$ and assume $x_2 \neq 0$. We have

$$\begin{cases} x_0^d + x_1^d + 1 = 0 \\ x_0 + x_1 + 1 = 0 \end{cases} \qquad \begin{cases} x_0^d + (-1)^d (x_0 + 1)^d + 1 = 0 \\ x_1 = -1 - x_0 \end{cases}.$$

As multiple roots of the intersection $F_{2,d} \cap \{x_0 + x_1 + x_2 = 0\}$ corresponds to multiple roots of the polynomial $f_d(x_0) := x_0^d + (-1)^d (x_0 + 1)^d + 1 \in \mathbb{C}[x_0]$, we try to determine the common factors of $f_d(x_0)$ and its derivative $f_d'(x_0) = d\big(x_0^{d-1} + (-1)^d (x_0 + 1)^{d-1}\big)$:

$$\begin{cases} x_0^d + (-1)^d (x_0 + 1)^d + 1 = 0 \\ x_0^{d-1} + (-1)^d (x_0 + 1)^{d-1} = 0 \end{cases} \begin{cases} x_0^d + (x_0 + 1)(-x_0^{d-1}) + 1 = 0 \\ (-1)^d (x_0 + 1)^{d-1} = -x_0^{d-1} \end{cases} \begin{cases} x_0^{d-1} = 1 \\ (x_0 + 1)^{d-1} = (-1)^{d-1} \end{cases}.$$

Now we observe that if $x_0 \in \mathbb{C}$ is a solution to the system, then $|x_0| = |x_0 + 1| = 1$, where $|\xi|$ is the usual Euclidean norm of a complex number ξ. The only complex numbers satisfying such relations are the third root of unity $\omega_3 = e^{i\frac{2\pi}{3}}$ and $\overline{\omega_3}$ (see Figure 1). Thus, we can check directly whether ω_3 and $\overline{\omega_3}$ are solutions. Substituting $x_0 = \omega_3$, we obtain

$$\begin{cases} \omega_3^{d-1} = 1 \\ (\omega_3 + 1)^{d-1} = (-1)^{d-1} \end{cases} \qquad (\omega_3 + 1 = -\overline{\omega_3}) \qquad \begin{cases} \omega_3^{d-1} = 1 \\ \overline{\omega_3}^{d-1} = 1 \end{cases}.$$

These equations are satisfied if, and only if, 3 divides $d - 1$, i.e. $d \equiv 1 \bmod 3$. In this case the multiple roots are exactly two, each with multiplicity two, since the second derivative
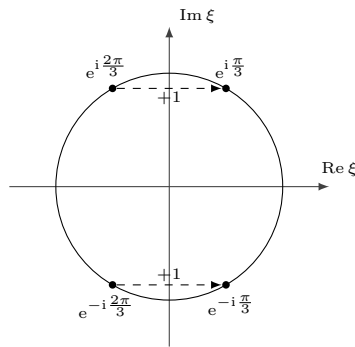
Figure 1: The complex numbers ξ such that $|\xi| = |\xi + 1| = 1$.

$f_d''(x_0)$ does not vanish at ω_3 and $\overline{\omega_3}$. Since the three coordinates of the multiple roots are the three solutions to the equation $\xi^3 = 1$, we notice that changing the affine chart does not lead to any other multiple roots. It follows that

$$\#(F_{2,d} \cap \{x_0 + x_1 + x_2 = 0\}) = \begin{cases} d, & \text{if } d \not\equiv 1 \bmod 3, \\ d - 2, & \text{if } d \equiv 1 \bmod 3. \end{cases}$$

To conclude the proof, we need to count how many simple points of $F_{2,d} \cap \{x_0 + x_1 + x_2 = 0\}$ lie also on a line $x_i = 0$. The points lying on $F_{2,d} \cap \{x_0 + x_1 + x_2 = 0\} \cap \{x_0 = 0\}$ satisfy the system

$$\begin{cases} x_0 = 0 \\ x_1 = -x_2 \\ \left((-1)^d + 1\right) x_2^d = 0 \end{cases}.$$

If d is even, there are no solutions, while if d is odd, we get the point $[0 : 1 : -1]$. We conclude that

$$\begin{aligned} \#(F_{2,d} \cap \mathcal{H}) = 3\#\left(F_{2,d} \cap \{x_0 = 0\}\right) \\ + \#\left(F_{2,d} \cap \{x_0 + x_1 + x_2 = 0\}\right) \\ - 3\#\left(F_{2,d} \cap \{x_0 + x_1 + x_2 = 0\} \cap \{x_0 = 0\}\right) = \end{aligned}$$

$$= \begin{cases} 3d & + & d, & \text{if } d \not\equiv 1 \bmod 3 \text{ and } d \text{ even}, \\ 3d & + & d & - & 3, & \text{if } d \not\equiv 1 \bmod 3 \text{ and } d \text{ odd}, \\ 3d & + & d - 2, & \text{if } d \equiv 1 \bmod 3 \text{ and } d \text{ even}, \\ 3d & + & d - 2 & - & 3, & \text{if } d \equiv 1 \bmod 3 \text{ and } d \text{ odd}, \end{cases}$$

$$= \begin{cases} 4d, & \text{if } d \equiv 0, 2 \bmod 6, \\ 4d - 3, & \text{if } d \equiv 3, 5 \bmod 6, \\ 4d - 2, & \text{if } d \equiv 4 \bmod 6, \\ 4d - 5, & \text{if } d \equiv 1 \bmod 6. \end{cases}$$

The formula for the ML degree follows by Theorem 6.

6. Computational results

This last section is intended to report the computational results we obtained. We provide a comparison among the possible procedures to get the ML degree of Fermat hypersurfaces that we explored and developed throughout the paper. We discussed three main methods: determining the multidegree of the likelihood correspondence, determining the degree of the projection $\pi_{\boldsymbol{y}} : \mathcal{L}_{n,d} \to \mathbb{P}^n_{\boldsymbol{y}}$ considering the fiber of a generic point of $\mathbb{P}^n_{\boldsymbol{y}}$ and partitioning the solutions of $I^{\mathbf{1}}_{n,d} \setminus \mathcal{H}$ according to the number of distinct coordinates (Theorem 5). To all these methods, Lemma 1 applies, so that we can choose to compute the saturation of an ideal with respect to a linear form or to compute the degree of two ideals, the one not saturated and the one intersected with the linear form. Hence, we compare the following six strategies (see Table 2).

Strategy 1 (simple "multidegree"): determine the ML degree of $F_{n,d}$ as the leading coefficient of the multidegree polynomial (5) of the likelihood correspondence $\mathcal{L}_{n,d}$.

Strategy 2 ("multidegree" by difference): apply Lemma 1 and Proposition 1 and determine the multidegree polynomial of $\mathcal{L}_{n,d}$ as the difference between the multidegree polynomial of the variety defined by $I^{\boldsymbol{y}}_{n,d}$ and the multidegree polynomial of the variety defined by $I^{\boldsymbol{y}}_{n,d} + (x_0 + \cdots + x_n)$.

Strategy 3 (simple "random data"): determine the ML degree of $F_{n,d}$ as the degree of the projection $\pi_{\boldsymbol{y}} : \mathcal{L}_{n,d} \to \mathbb{P}^n_{\boldsymbol{y}}$, i.e. computing the degree of the ideal $I^{\boldsymbol{u}}_{n,d} \setminus \mathcal{H}$ for a randomly chosen $\boldsymbol{u} \in \mathbb{P}^n_{\boldsymbol{y}}$.

Strategy 4 ("random data" by difference): apply Lemma 1 and Proposition 1 and compute $\deg\left(I^{\boldsymbol{u}}_{n,d} \setminus \mathcal{H}\right)$ as $\deg I^{\boldsymbol{u}}_{n,d} - \deg\left(I^{\boldsymbol{u}}_{n,d} + (x_0 + \ldots + x_n)\right)$ for a randomly chosen $\boldsymbol{u} \in \mathbb{P}^n_{\boldsymbol{y}}$.

Strategy 5 (simple "partitioning"): apply Theorem 4 and Theorem 5 and determine $\text{MLdeg}\, F_{n,d}$ from the degree of the ideals $I^d_{\mathsf{a}} \setminus \mathcal{H}$ defining the a-critical points.

Strategy 6 ("partitioning" by difference): apply Lemma 1, Proposition 1, Theorem 4 and Theorem 5 and compute $\deg\left(I^d_{\mathsf{a}} \setminus \mathcal{H}\right)$ as $\deg I^d_{\mathsf{a}} - \deg\left(I^d_{\mathsf{a}} \cap \mathcal{H}\right)$.

Table 2:   An experimental comparison of the different strategies.   All the algorithms for the different strategies are implemented with the software *Macaulay2* [7]. The code is developed in the package `MLdegreeFermatHypersurface.m2` available at `http://www.paololella.it/EN/Publications.html`. A complete report of the tests can be found at the same address. The algorithms have been run on a Intel(R) Xeon(R) 2.60 GHz processor. We considered the Fermat hypersurfaces in the ranges $2 \leqslant n \leqslant 9$, $2 \leqslant d \leqslant 9$ with a limit for completion of 10^4 seconds of cpu-time. For more efficient computations, polynomial rings with coefficients in a finite field have been exploited; the correctness of the results has been checked by repeating the same process using fields with different characteristic.
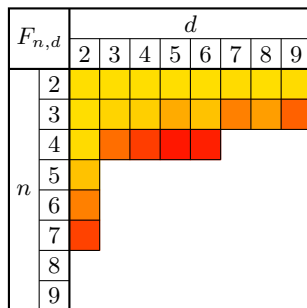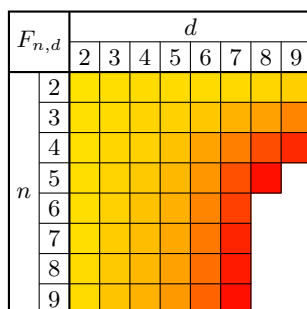

(a) **Simple "multidegree".**
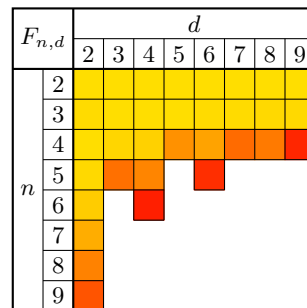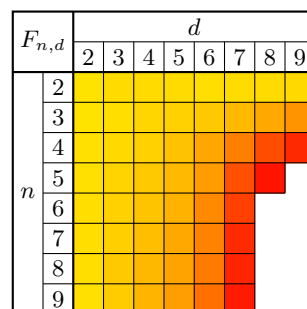

(b) **"Multi-degree" by difference.**
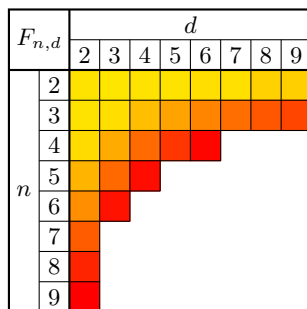
(c) **Simple "random data".**
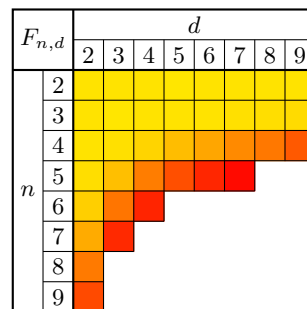
(d) **"Random data" by difference.**

(e) **Simple "partitioning".**

(f) **"Partitioning" by difference.**

CPU-TIME

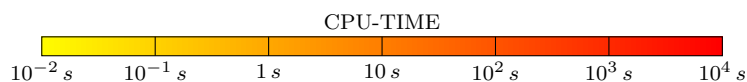$10^{-2}\,s$ $10^{-1}\,s$ $1\,s$ $10\,s$ $10^2\,s$ $10^3\,s$ $10^4\,s$

Table 2 suggests that for the strategies based on the "multidegree" and "random data" methods the dimension n of the projective space affects the computational complexity slightly more than the degree d of the Fermat hypersurface. The effect on the complexity of the number of variables and of the degree totally changes when using the **Strategy 5** or **Strategy 6** which rely on Theorem 5. Table 2 shows that the complexity of the "partitioning" method is heavily affected by the degree d and mildly affected by the dimension of the ambient space n. This behavior occurs because the effective computations in these strategies are performed on a number of variables bounded by $\min\{d, n+1\}$ and so eventually smaller than n. More precisely, assume that $\mathsf{a} \in \mathcal{P}_{n+1,d}$ and $\mathsf{b} \in \mathcal{P}_{n'+1,d}$ are two partitions of $n+1$ and $n'+1$ with same length s. The ideals $I_\mathsf{a}^d \setminus \mathcal{H}$ and $I_\mathsf{b}^d \setminus \mathcal{H}$ defining the critical points are both contained in the polynomial ring $\mathbb{C}[z_1, \ldots, z_s]$ and they are in fact almost the same ideal: they only differ by one generator $(a_1 z_1^d + \cdots + a_s z_s^d$ instead of $b_1 z_1^d + \cdots + b_s z_s^d)$ and one different saturation $(a_1 z_1 + \cdots + a_s z_s$ instead of $b_1 z_1 + \cdots + b_s z_s)$. For such ideals the complexity of degree computation does not change substantially. Hence, for a fixed d and for $n \geqslant d - 1$, the running time of ML degree algorithms grows slowly as n increases, depending only on the number of partitions in $\mathcal{P}_{n+1,d}$. Thus, we need to repeat more times computations for ideals whose complexity is basically fixed.

Finally, Table 3 lists the ML degree of the Fermat hypersurface $F_{n,d}$, for several values of n and d.

Table 3: The ML degree of several Fermat hypersurfaces. The empty entries of the table correspond to ML degrees whose computation following the best strategy could not be completed within 24 hours of cpu-time.

MLdeg $F_{n,d}$		d								
		2	3	4	5	6	7	8	9	10
	2	6	9	18	27	42	51	72	87	108
	3	14	30	76	140	258	370	584	792	1102
	4	30	95	320	725	1530	2635	4680	7265	11090
	5	62	293	1294	3655	9186	18627	37448		
	6	126	896	5180	18494	55482	131320	298872		
	7	254	2726	20892	92972	334578	924154			
n	8	510	6813	84132	467685	2012514	6496251			
	9	1022	25047	337384	2347469	12064506	45627451			
	10	2046	75746	1353110	11781044	72298842	320280400			
	11	4094	228825	5429494	59070599	433840578	2247181471			
	12	8190	690690	21767018	296105784	2605621434	15761369624			
	13	16382	2083370	87215496	1483630894	15650082090	110517144758			
	14	32766	6280649	349452578	7432036277	93935183202	774762908611			
	15	65534	18925046	1397573292	37220018572	563502117618	5430367540394			

Acknowledgments

The content of [16] has been the topic of a series of lectures taught by Bernd Sturmfels at the summer school *Combinatorial algebraic geometry* in Levico Terme in June 2013. This paper arises from an exercise he suggested during a training session. We thank Bernd Sturmfels for having pushed and encouraged us to go into depth on the matter. We also thank the referees for having pointed out some inaccuracies and for the useful comments and suggestions.

References

[1] J.-E. Björk. *Rings of differential operators*, volume 21 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1979.

[2] Max-Louis G. Buot, Serkan Hoşten, and Donald St. P. Richards. Counting and locating the solutions of polynomial systems of maximum likelihood equations. II. The Behrens-Fisher problem. *Statist. Sinica*, 17(4):1343–1354, 2007.

[3] Fabrizio Catanese, Serkan Hoşten, Amit Khetan, and Bernd Sturmfels. The maximum likelihood degree. *Amer. J. Math.*, 128(3):671–697, 2006.

[4] Jan Draisma and Jose Rodriguez. Maximum likelihood duality for determinantal varieties. *Int. Math. Res. Not. IMRN*, 20:5648–5666, 2014.

[5] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on algebraic statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel, 2009.

[6] David Eisenbud. *Commutative algebra*, volume 150 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. With a view toward algebraic geometry.

[7] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. available at `www.math.uiuc.edu/Macaulay2/`.

[8] Elizabeth Gross, Mathias Drton, and Sonja Petrović. Maximum likelihood degree of variance component models. *Electron. J. Stat.*, 6:993–1016, 2012.

[9] Elizabeth Gross and Jose Rodriguez. Maximum likelihood geometry in the presence of data zeros. In *ISSAC 2014 — Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, 2014.

[10] Robin Hartshorne. *Algebraic geometry*. Springer-Verlag, New York, 1977.

[11] Jonathan Hauenstein, Jose Rodriguez, and Bernd Sturmfels. Maximum likelihood for matrices with rank constraints. *J. Alg. Stat.*, 5(1):18–38, 2014.

[12] Serkan Hoşten, Amit Khetan, and Bernd Sturmfels. Solving the likelihood equations. *Found. Comput. Math.*, 5(4):389–407, 2005.

[13] Serkan Hoşten and Seth Sullivant. The algebraic complexity of maximum likelihood estimation for bivariate missing data. In *Algebraic and geometric methods in statistics*, pages 123–133. Cambridge Univ. Press, Cambridge, 2010.

[14] June Huh. The maximum likelihood degree of a very affine variety. *Compos. Math.*, 149(8):1245–1266, 2013.

[15] June Huh. Varieties with maximum likelihood degree one. *J. Alg. Stat.*, 5(1):1–17, 2014.

[16] June Huh and Bernd Sturmfels. Likelihood geometry. In Sandra Di Rocco and Bernd Sturmfels, editors, *Combinatorial Algebraic Geometry*, volume 2108 of *Lecture Notes in Mathematics (C.I.M.E. Foundation Subseries)*. Springer, 2014.

[17] Joachim Jelisiejew. Deformations of zero-dimensional schemes and applications. Master's thesis, University of Warsaw, 2013. Available at `arxiv.org/abs/1307.8108`.

[18] Ezra Miller and Bernd Sturmfels. *Combinatorial commutative algebra*, volume 227 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2005.

[19] Lior Pachter and Bernd Sturmfels. *Algebraic Statistics for Computational Biology*. Cambridge Univ. Press, 2005.

[20] Caroline Uhler. Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist.*, 40(1):238–261, 2012.