

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Grosdos, Alexandros; Heaton, Alexander; Kubjas, Kaie; Kuznetsova, Olga; Scholten, Georgy; Sorea, Miruna Ștefana

## Exact solutions in log-concave maximum likelihood estimation

*Published in:*  
Advances in Applied Mathematics

*DOI:*  
[10.1016/j.aam.2022.102448](https://doi.org/10.1016/j.aam.2022.102448)

Published: 01/02/2023

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Grosdos, A., Heaton, A., Kubjas, K., Kuznetsova, O., Scholten, G., & Sorea, M. Ș. (2023). Exact solutions in log-concave maximum likelihood estimation. *Advances in Applied Mathematics*, 143, 1-32. Article 102448.  
<https://doi.org/10.1016/j.aam.2022.102448>

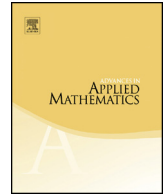


ELSEVIER

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama



# Exact solutions in log-concave maximum likelihood estimation



Alexandros Grosdos<sup>a,1</sup>, Alexander Heaton<sup>b,2</sup>, Kaie Kubjas<sup>c,\*</sup>,  
Olga Kuznetsova<sup>c</sup>, Georgy Scholten<sup>d,3</sup>, Miruna-Ștefana Sorea<sup>b,4</sup>

<sup>a</sup> Institute for Mathematics, Osnabrück University, Osnabrück, Germany

<sup>b</sup> Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

<sup>c</sup> Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland

<sup>d</sup> Department of Mathematics, North Carolina State University, Raleigh, NC, USA

## ARTICLE INFO

### Article history:

Received 1 September 2020

Received in revised form 30 August 2022

Accepted 11 October 2022

Available online xxxx

### MSC:

52B99

62G07

### Keywords:

Log-concavity

Maximum likelihood estimation

Polynomial-exponential systems

Certified solutions

## ABSTRACT

We study probability density functions that are log-concave. Despite the space of all such densities being infinite-dimensional, the maximum likelihood estimate is the exponential of a piecewise linear function determined by finitely many quantities, namely the function values, or heights, at the data points. We explore in what sense exact solutions to this problem are possible. First, we show that the heights given by the maximum likelihood estimate are generically transcendental. For a cell in one dimension, the maximum likelihood estimator is expressed in closed form using the generalized  $W$ -Lambert function. Even more, we show that finding the log-concave maximum likelihood estimate is equivalent to solving a collection of polynomial-exponential systems of a special form. Even in the case of two equations, very little is known about solutions to these systems. As an alternative, we use Smale's

\* Corresponding author.

E-mail addresses: alex.grosdos@tum.de (A. Grosdos), alexheaton2@gmail.com (A. Heaton), kaie.kubjas@aalto.fi (K. Kubjas), olga.kuznetsova@aalto.fi (O. Kuznetsova), georgy.scholten@lip6.fr (G. Scholten), msorea@sissa.it (M.-Ș. Sorea).

<sup>1</sup> Department of Mathematics, TUM School of Computation, Information and Technology, Technical University of Munich, Munich, Germany.

<sup>2</sup> Department of Mathematics, Lawrence University, Appleton, WI, USA.

<sup>3</sup> LIP6, LJLL, Sorbonne Université, Paris, France.

<sup>4</sup> SISSA - Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy and RCMA, Lucian Blaga University Sibiu, Sibiu, Romania.

Transcendence theory  
 Smale's  $\alpha$ -theory  
 Lambert functions  
 Polyhedral subdivisions

$\alpha$ -theory to refine approximate numerical solutions and to certify solutions to log-concave density estimation.

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nonparametric methods in statistics emerged in the 1950-1960s [26,50,44,3] and fall into two main streams: smoothing methods and shape constraints. Examples of smoothing methods include delta sequence methods such as kernel, histogram and orthogonal series estimators [58], and penalized maximum likelihood estimators, e.g., spline methods [25]. Their defining feature is the need to choose the smoothing or tuning parameters. It is a delicate process because smoothing parameters depend on the unknown probability density function. In contrast to smoothing methods, shape constrained nonparametric density estimation is fully automatic and does not depend on the underlying probability distribution, though this comes at the expense of worse  $L^1$  convergence rates for smooth densities [24]. Some previously studied classes of functions include non-increasing [27], convex [29],  $k$ -monotone [7] and  $s$ -concave [20]. We refer the reader to [55,53,57,28] for general references on nonparametric statistics. The definitions of  $k$ -monotone and  $s$ -concave can be found in [6] and [18], respectively.

In this paper we focus on the class of log-concave densities, which is an important special case of  $s$ -concave densities. The choice of log-concavity is attractive for several reasons. First of all, most common univariate parametric families are log-concave, including the normal, Gamma with shape parameter greater than one, Beta densities with parameters greater than 1, Weibull with parameter greater than 1 and others. Furthermore, log-concavity is used in reliability theory, economics and political science [4]. In addition to this, log-concave densities have several desirable statistical properties. For example, log-concavity implies unimodality but log-concave density estimation avoids the spiking phenomenon common in general unimodal estimation [21]. Moreover, this class is closed under convolutions and taking pointwise limits [13]. We refer the reader to [52] for an overview of the recent progress in the field.

Let  $X = (x_1, x_2, \dots, x_n)$  be a point configuration in  $\mathbb{R}^d$  with weights  $w = (w_1, w_2, \dots, w_n)$  such that  $w_i \geq 0$  and  $w_1 + w_2 + \dots + w_n = 1$ . The log-concave *maximum likelihood estimation* (MLE) problem aims to find a Lebesgue density that solves

$$\max \sum_{i=1}^n w_i \log(f(x_i)) \text{ s.t. } \log(f) \text{ is concave and } \int_{\mathbb{R}^d} f(x) dx = 1. \quad (1.1)$$

It has been shown that the solution exists with probability 1 and is unique, and its logarithm is a tent function, i.e., a piecewise linear function with regions of linearity inducing a subdivision of the convex hull of  $X$  [59,43,15,49], see Fig. 1 for an example.

While MLE is the most widely studied estimator in this setting, it is not the only one, for examples see [19,16].

The maximum likelihood estimator is attractive because of its consistency under general assumptions [43,21,13,23] and superior performance compared to kernel-based methods with respect to mean integrated squared error, as observed in simulations [15]. At the same time, the convergence rate is still an open question and only lower [33,35] and upper [33,10] bounds are known. Further theoretical properties have been studied for some special cases of log-concave densities, e.g.,  $k$ -affine densities [34] and totally positive densities [48]. Several algorithms have been developed to compute the log-concave MLE in one dimension [51] and in higher dimensions [15,2,46]. Software implementations include R packages such as `logcondens` [22] and `cnm1cd` [37] in one dimension, and `LogConcDEAD` [14] and `fmlogcondens` [45] in higher dimensions.

**Example 1.1.** The starting point of this paper is the following problem. Consider the sample of 14 points in  $\mathbb{R}^2$  with uniform weights:

$$X = ((0, 1), (0, 9), (1, 4), (2, 4), (2, 6), (3, 3), (5, 5), (6, 3), (6, 9), (7, 6), (7, 8), (8, 9), (9, 5), (9, 9)).$$

How many cells does the subdivision induced by the logarithm of the optimal log-concave density have?

Using the R package `LogConcDEAD` with default parameters, one obtains that the logarithm of the maximum likelihood estimate is a piecewise linear function with seven unique linear pieces. However, when one investigates the optimal density more closely, it appears that several linear pieces are similar. For example, a visual inspection of the optimal density depicted in Fig. 1 makes it impossible to distinguish all 7 regions and suggests that there are only four unique linear pieces. Using `LogConcDEAD` one also obtains the two triangles, but according to the `LogConcDEAD` output the quadrangle consists of two linear pieces and the hexagon consists of three linear pieces. The subdivision corresponding to the `LogConcDEAD` result is depicted in Fig. 8a. What is the true number of unique linear pieces of the optimal density? Is it four, seven or another value?

Theoretically, the algorithm used in `LogConcDEAD` finds the true optimal density, however, in practice, the answer is a numerical approximation. By changing the parameter `sigmatol` from default value  $10^{-8}$  to  $10^{-10}$ , `LogConcDEAD` outputs four unique linear pieces, exactly as we observed in Fig. 1. Although it might seem obvious that four is the correct number of linear pieces, in reality the situation is more complicated, see Example 4.16. How do we find the correct number of linear pieces?

The goal of this paper is to study exact solutions to log-concave maximum likelihood estimation. An exact solution will have three different meanings in this paper. First, one might hope that it is an algebraic number. This would enable exact symbolic computations by way of storing a floating point approximation of a number along with a

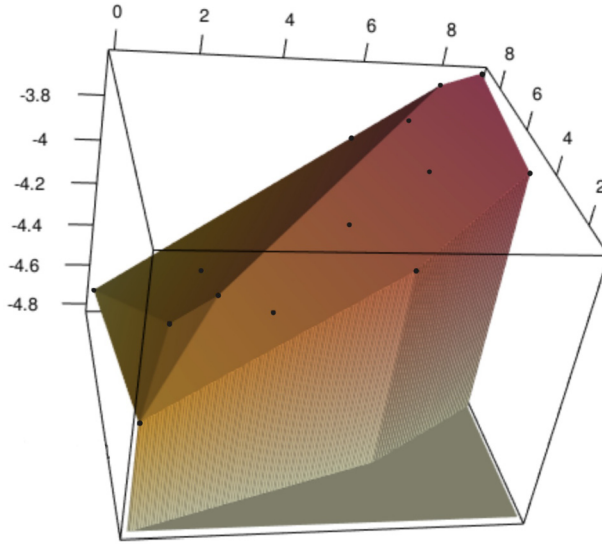


Fig. 1. The optimal tent function for the sample of 14 points in Example 1.1.

polynomial that vanishes on it. Such computations are not possible for transcendental numbers. Thus, the first main result of our paper is Theorem 3.7, which states that the heights at the sample points of the logarithm of the log-concave density estimate are, in general, transcendental.

Second, in light of Theorem 3.7, we would like to express the maximum likelihood estimator in closed form using well-known mathematical operations and functions, although not necessarily elementary functions. In the simplest case of one cell in one dimension, we derive the log-concave density estimator in closed form using the generalized  $W$ -Lambert function, see Proposition 3.9. It is known that the generalized  $W$ -Lambert function is not an elementary function. More generally, solving the MLE can be restated as a collection of polynomial-exponential systems of equations, which have been studied in the literature. However, even in the case of two equations, only bounds on solutions are known [38]. This suggests that it might be difficult to express the log-concave maximum likelihood estimator in closed form. As an alternative, we turn to Smale's  $\alpha$ -theory, which we describe briefly now.

Third, given a sufficiently close floating point solution to the MLE problem, one hopes that it can be refined to any desired precision using Newton iteration or other techniques. A natural question arises: when is the approximate solution good enough for these methods to succeed? A way to make this mathematically rigorous is Smale's  $\alpha$ -theory [9,56], which we discuss in Section 4. We obtain the  $\alpha$ -certified solutions to log-concave density estimation. This allows us to test and compare numerical solvers, as well as rigorously decide the certified, correct subdivision for a given log-concave density estimation problem. Our methods are especially relevant when the precision of

the log-concave density estimate is important. This opens new pathways to answering the motivating question: what is the correct number of cells?

The code for computations in this paper can be found at [30].

## 2. Geometry of log-concave maximum likelihood estimation

We start by reviewing the geometry of log-concave maximum likelihood estimation mostly following [49].

**Definition 2.1.** Let  $P$  be the convex hull of a point configuration  $X = (x_1, x_2, \dots, x_n) \subset \mathbb{R}^d$ . For a fixed real vector  $y \in \mathbb{R}^n$ , we define a function  $h_{X,y}$  on  $\mathbb{R}^d$ , called the *tent function*, as the smallest concave function such that  $h_{X,y}(x_i) \geq y_i$  for  $i = 1, \dots, n$ . Here the term smallest means that for any other concave function  $\bar{h}$  on  $\mathbb{R}^d$  such that  $\bar{h}(x_i) \geq y_i$  for  $i = 1, \dots, n$ , one must have  $\bar{h}(x) \geq h_{X,y}(x)$  for all  $x \in \mathbb{R}^d$ . The tent function  $h_{X,y}$  is piecewise linear on  $P$  with linear pieces equal to upper facets of the convex hull of the points  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  in  $\mathbb{R}^{d+1}$ . We have  $h_{X,y}(x) = -\infty$  at all points  $x \in \mathbb{R}^d$  outside  $P$ . If  $h_{X,y}(x_i) = y_i$  for  $i = 1, \dots, n$ , then  $y$  is called *relevant*.

It was shown by Cule, Samworth and Stewart for uniform weights [15] and by Robeva, Sturmfels and Uhler in general [49] that the constrained optimization problem (1.1) of finding the log-concave maximum likelihood estimate is equivalent to the unconstrained optimization problem

$$\max_{y \in \mathbb{R}^n} w \cdot y - \int_P \exp(h_{X,y}(t)) dt. \quad (2.1)$$

Moreover, the log-concave maximum likelihood estimate is a tent function with tent poles at some of the  $x_i$ . Therefore finding the log-concave density which maximizes the likelihood of  $(X, w)$  is equivalent to finding an optimal height vector  $y^*$ .

**Definition 2.2.** We follow the definitions in [17]. Given a point configuration  $X$  in  $\mathbb{R}^d$ , a *subdivision*  $\Delta$  of  $X$  is a collection of  $d$ -polytopes, denoted  $\sigma_i$ , such that the union of polytopes in  $\Delta$  equals  $\text{conv}(X)$ , the vertex set of polytopes in  $\Delta$  is contained in  $X$  and the intersection of polytopes in  $\Delta$  can only happen along lower dimensional faces. A subdivision  $\Delta$  is called a *triangulation*, if all polytopes in  $\Delta$  are simplices. A triangulation  $\Delta$  of the point configuration  $X$  is called *maximal*, if every element of  $X$  is a vertex of a simplex in  $\Delta$ . A subdivision is called *regular* if its full dimensional cells  $\sigma_i$  are combinatorially equivalent to the regions of linearity of a tent function on  $X$  for some height vector  $y \in \mathbb{R}^n$ .

**Corollary 2.3.** [49, Corollary 2.6] To find the optimal height vector  $y^*$  in (2.1) is to maximize the following rational-exponential objective function over  $y \in \mathbb{R}^n$ :

$$S(y_1, \dots, y_n) = w \cdot y - \sum_{\sigma \in \Delta} \sum_{i \in \sigma} \frac{\text{vol}(\sigma) \cdot \exp(y_i)}{\prod_{\alpha \in \sigma \setminus i} (y_i - y_\alpha)}, \quad (2.2)$$

where  $\Delta$  is any regular triangulation that refines the regular subdivision induced by the tent function  $h_{X,y}$ .

If  $y$  induces a regular subdivision  $\Delta$  that is not a maximal regular triangulation, then we can consider any maximal regular triangulation that refines  $\Delta$ . Thus if there are  $m$  maximal regular triangulations of  $X$ , then to find the optimal  $y^*$  we must compare the optimal values  $y_{\Delta_1}^*, y_{\Delta_2}^*, \dots, y_{\Delta_m}^*$  which are obtained by solving the optimization problem (2.2)  $m$  times, once for each maximal regular triangulation  $\Delta_1, \Delta_2, \dots, \Delta_m$ .

**Notation 2.4.** We will denote by  $S_\Delta$  the function given by the right hand side of (2.2) for a fixed triangulation  $\Delta$ .

**Example 2.5.** Fix  $d = 1$ ,  $n = 3$  and  $X = (2, 5, 7)$ . The configuration  $X$  has two triangulations  $\Delta_1 = \{\{1, 3\}\}$  and  $\Delta_2 = \{\{1, 2\}, \{2, 3\}\}$ , which are both regular triangulations. Only  $\Delta_2$  is a maximal triangulation. Hence solving the optimization problem (2.1) is equivalent to maximizing the objective function

$$S_{\Delta_2} = w \cdot y - 3 \frac{e^{y_1} - e^{y_2}}{y_1 - y_2} - 2 \frac{e^{y_2} - e^{y_3}}{y_2 - y_3}. \quad (2.3)$$

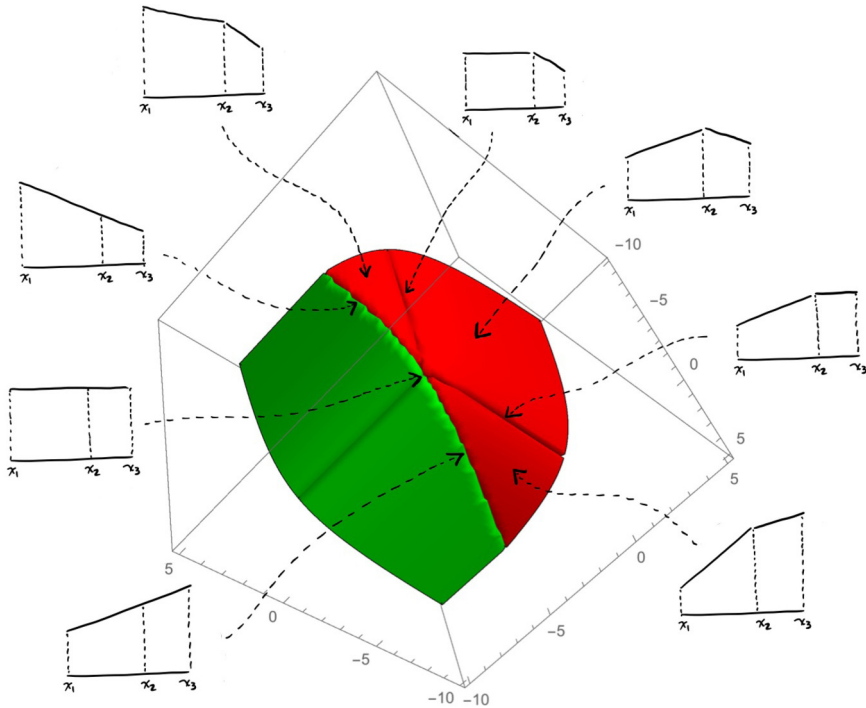
If  $y_1 = y_2$  or  $y_2 = y_3$ , then a denominator on the right hand side of (2.3) becomes zero. However, the objective function in the formulation (2.1) can be still simplified to

$$w \cdot y - 3e^{y_2} - 2 \frac{e^{y_2} - e^{y_3}}{y_2 - y_3} \quad \text{or} \quad w \cdot y - 3 \frac{(e^{y_1} - e^{y_2})}{y_1 - y_2} - 2e^{y_2}.$$

To visualize the situation, we consider the Samworth body

$$\mathcal{S}(X) = \left\{ y \in \mathbb{R}^3 : \int_P \exp(h_{X,y}(t)) dt \leq 1 \right\},$$

which was introduced in [49]. The unconstrained optimization problem (2.1) is equivalent to the constrained optimization problem of maximizing the linear function  $w \cdot y$  over the Samworth body. For different choices of weight vector  $w = (w_1, w_2, w_3)$ , we obtain different optimal height vectors  $y = (y_1, y_2, y_3)$  on the surface of the Samworth body, and the height vector determines the triangulation. The Samworth body consists of two regions that can be seen in Fig. 2. The green region comes from the one-simplex triangulation  $\Delta_1 = \{\{1, 3\}\}$ , while the red region comes from the two-simplex triangulation  $\Delta_2 = \{\{1, 2\}, \{2, 3\}\}$ . Moreover, one can see lines separating the green region into two pieces and the red region into three pieces (ignore the curve separating the green and



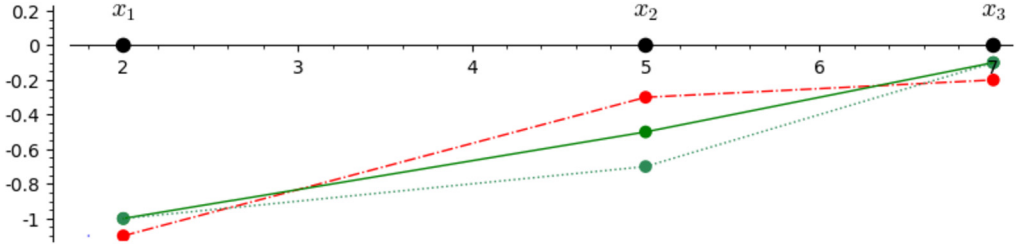
**Fig. 2.** The Samworth body for  $X = (2, 5, 7)$ . (For interpretation of the colors in the figures, the reader is referred to the web version of this article.)

the red regions for now). These lines correspond to the degenerate cases where  $y_1 = y_3$ ,  $y_1 = y_2$  or  $y_2 = y_3$ , and hence the right hand side of (2.2) is not defined. Therefore those lines are simply artifacts of the reformulation (2.2) since in the original unconstrained setting (2.1) these points present no difficulty. The intersection of the three lines is the point  $(-\log 5, -\log 5, -\log 5)$ .

Consider the curve separating the green and red regions of the Samworth body. This curve is made of all the points  $y$  that form a relevant tent function, inducing the subdivision  $\Delta_1$ . To understand the green region, see the piecewise linear functions drawn in Fig. 3. Since the lowest (dotted) function is not concave, it is invalid as a tent function. Therefore, if the height  $y_2$  is too low, the optimal tent function will be the (solid-line) linear function. In effect, the optimal tent-function ignores heights  $y_i$  if they are too low. This basic phenomenon is responsible for the green part of the Samworth body being flat in the  $y_2$  direction, meaning that it is a pencil of half-lines parallel to the  $y_2$ -axis.

The transition from the red region to the green region is not smooth. For every  $y$  on the curve between the green and red regions, there is a two-dimensional cone of weight vectors that give  $y$  as an optimal solution. The generators of this cone are described in [49, Theorem 3.7]. The optimal height vector  $y^*$  for  $w = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$  lies on the curve between the red and green regions. It is not a critical point of the function (2.3), because  $w$  is not a normal vector to the red region at the point  $y^*$ .





**Fig. 3.** The red tent function corresponds to a vector  $y$  in the red region of the Samworth body. The solid green tent function corresponds to a vector  $y$  on the curve separating red and green regions of the Samworth body. The dotted green function is not concave. Its height vector  $y$  belongs to the green region of the Samworth body and both green sets of heights give the same tent function.

We now return to the general situation and consider the specific approach of critical equations for solving the optimization problem (2.2). Let  $X = (x_1, \dots, x_n)$  be a configuration of  $n$  points  $x_i \in \mathbb{R}^d$ . Fixing a maximal regular triangulation  $\Delta$  of our point configuration  $X$ , we can find the optimal  $y_\Delta^*$  for  $S_\Delta$  in (2.2) over  $y \in \mathbb{R}^n$  by solving the system of critical equations  $\partial S_\Delta / \partial y_i = 0$ . These partial derivatives take the form (see [49, Proof of Lemma 3.4]):

$$\begin{aligned} \frac{\partial S_\Delta}{\partial y_i} = & w_i - \sum_{\substack{\sigma \in \Delta, \\ i \in \sigma}} \text{vol}(\sigma) \exp(y_i) \frac{1}{\prod_{\alpha \in \sigma \setminus i} (y_i - y_\alpha)} \left( 1 - \sum_{\alpha \in \sigma \setminus i} \frac{1}{(y_i - y_\alpha)} \right) \\ & - \sum_{\substack{\sigma \in \Delta, \\ i \in \sigma}} \text{vol}(\sigma) \sum_{j \in \sigma \setminus i} \exp(y_j) \frac{1}{\prod_{\alpha \in \sigma \setminus j} (y_j - y_\alpha)} \frac{1}{(y_j - y_i)}. \end{aligned} \quad (2.4)$$

**Definition 2.6.** For a fixed maximal regular triangulation  $\Delta$  of  $X$ , let  $A$  be the matrix such that the system of  $n$  critical equations (2.4) can be written in the form

$$Ae^y = w, \quad (2.5)$$

where  $e^y$  is a column vector of exponentials  $(e^{y_1}, e^{y_2}, \dots, e^{y_n})^T$ , and  $w$  is a column vector of weights  $(w_1, \dots, w_n)^T$ . The matrix  $A$  is called the *score equation matrix*.

The entries of  $A$  are in the field of rational functions in the variables  $y_1, \dots, y_n$ . Diagonal entries of  $A$  are

$$A_{j,j} = \sum_{\substack{\sigma \in \Delta, \\ j \in \sigma}} \text{vol}(\sigma) \frac{1}{\prod_{\alpha \in \sigma \setminus j} (y_j - y_\alpha)} \left( 1 - \sum_{\alpha \in \sigma \setminus j} \frac{1}{(y_j - y_\alpha)} \right)$$

and off-diagonal entries of  $A$  are

$$A_{i,j} = \sum_{\substack{\sigma \in \Delta, \\ i,j \in \sigma}} \text{vol}(\sigma) \frac{1}{\prod_{\alpha \in \sigma \setminus j} (y_j - y_\alpha)} \frac{1}{(y_j - y_i)}.$$

The matrix  $A$  can be written as a sum of matrices over maximal simplices  $\sigma \in \Delta$ . This will be described explicitly in the proof of Theorem 3.1.

There are two caveats when solving the optimization problem (2.2) using the method of critical equations. First, it is not enough to consider the system of critical equations  $\partial S_\Delta / \partial y_i = 0$  only for each of the maximal regular triangulations  $\Delta$ , since the optimization problem (2.2) is not smooth. One has to consider a system of critical equations for each subdivision of  $X$ . For a general subdivision  $\Delta$  of  $X$ , this system is constructed in the following way. We consider  $S_{\Delta'}(y_1, \dots, y_n)$  for any maximal triangulation  $\Delta'$  that refines  $\Delta$ , substitute  $y_i$  that can be expressed in terms of other  $y$ 's in the subdivision  $\Delta$  and construct the system of critical equations  $\partial \tilde{S}_\Delta / \partial y_i = 0$  for the resulting function  $\tilde{S}_\Delta$ . For maximal triangulations, we have  $\tilde{S}_\Delta = S_\Delta$  and the system of critical equations is given by (2.4). We will demonstrate this phenomenon on the point configuration from Example 2.5.

**Example 2.7.** Recall that  $d = 1$ ,  $n = 3$  and  $X = (2, 5, 7)$ . The configuration  $X$  has two triangulations  $\Delta_1 = \{\{1, 3\}\}$  and  $\Delta_2 = \{\{1, 2\}, \{2, 3\}\}$ . Let  $w = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ . The output from LogConcDEAD suggests that the optimal tent function is supported on one cell, with heights given by  $y_1^* = -1.816665$ ,  $y_2^* = -1.576024$  and  $y_3^* = -1.415597$ . However, the vector  $y^*$  is neither a critical point of  $S_{\Delta_2}$  nor of the function

$$S_{\Delta_1} = w \cdot y - 5 \frac{e^{y_1} - e^{y_3}}{y_1 - y_3}.$$

This can be seen by taking partial derivatives of these functions with respect to  $y_1, y_2, y_3$  and substituting  $y_1^*, y_2^*, y_3^*$ . In the case of  $\partial S_{\Delta_1} / \partial y_i = 0$ , it is particularly easy to see that there are no solutions, since  $\partial S_{\Delta_1} / \partial y_2 = w_2 \neq 0$ . In the case of  $\partial S_{\Delta_2} / \partial y_i = 0$ , the system of critical equations fails to certify in the sense of Section 4.

The points  $(x_1, y_1^*), (x_2, y_2^*), (x_3, y_3^*)$  being collinear is equivalent to  $(x_2, y_2^*) = \lambda_1(x_1, y_1^*) + \lambda_3(x_3, y_3^*)$  where  $\lambda_1, \lambda_3 \geq 0$ ,  $\lambda_1 + \lambda_3 = 1$ . Since  $x_1 = 2, x_2 = 5, x_3 = 7$ , we have  $\lambda_1 = \frac{2}{5}, \lambda_3 = \frac{3}{5}$ . Hence  $y_2 = \frac{2}{5}y_1 + \frac{3}{5}y_3$ . Substituting this expression into the objective function (2.3) we get

$$\tilde{S}_{\Delta_2} = \left(w_1 + \frac{2}{5}w_2\right)y_1 + \left(w_3 + \frac{3}{5}w_2\right)y_3 - 5 \frac{e^{y_1} - e^{y_3}}{y_1 - y_3}$$

which for uniform weights  $w = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$  becomes

$$\tilde{S}_{\Delta_2} = \frac{7}{15}y_1 + \frac{8}{15}y_3 - 5 \frac{e^{y_1} - e^{y_3}}{y_1 - y_3}. \quad (2.6)$$

We will verify in Example 4.13 that  $y^*$  is a critical point of the function  $\tilde{S}_{\Delta_2}$ .

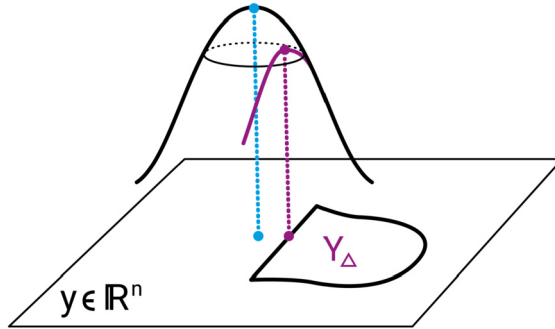


Fig. 4. Maximizing  $S_\Delta$  over  $y$  restricted to  $Y_\Delta$ .

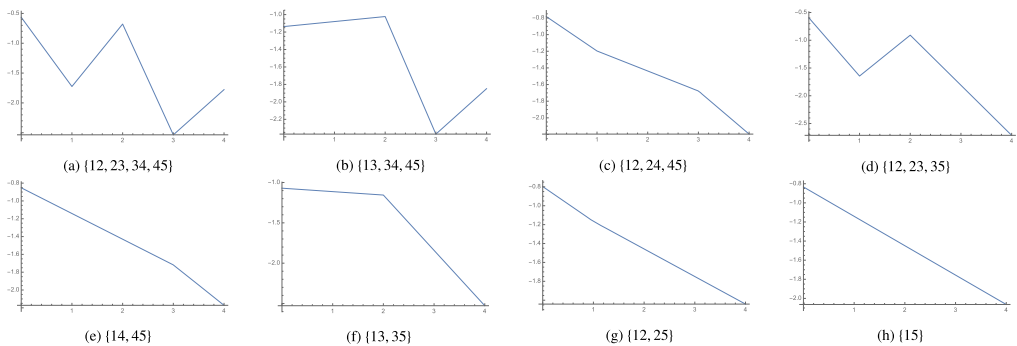


Fig. 5. Piecewise-linear functions induced by  $y_\Delta^*$  maximizing  $S_\Delta$  for each subdivision  $\Delta$  in Example 2.8. The notation  $ij$  in the subcaptions refers to the set  $\{i, j\}$ .

The second caveat is that to find the optimal tent function, it is not enough to merely compare the optimal critical points  $y_\Delta^*$  of  $\partial S_\Delta / \partial y_i = 0$  for each subdivision  $\Delta$ . Denote by  $Y_\Delta$  the set of  $y$  that induce a subdivision that is equal to or coarser than  $\Delta$ . For each  $\Delta$ , it also has to be checked that  $y_\Delta^*$  is in  $Y_\Delta$ . Thus if  $y_\Delta^*$  is not in  $Y_\Delta$ , then  $y_\Delta^*$  should be discarded. If the maximum of  $S_\Delta$  over  $Y_\Delta$  is not a critical point of  $S_\Delta$ , then the maximum must be on the boundary of  $Y_\Delta$ , see Fig. 4 for an illustration. The boundary of  $Y_\Delta$  is stratified into regions  $Y_{\tilde{\Delta}}$  corresponding to the various subdivisions  $\tilde{\Delta}$  which are refined by  $\Delta$ . Hence one should consider critical points for strictly coarser subdivisions  $\tilde{\Delta}$ .

**Example 2.8.** We consider the point configuration  $X = \{0, 1, 2, 3, 4\} \subseteq \mathbb{R}$  and the weight vector  $w = (3/15, 4/15, 5/15, 2/15, 1/15)$ . This point configuration has exactly eight subdivisions. For each subdivision  $\Delta$ , we use the `Mathematica` command `NMaximize` to find the maximum  $y_\Delta^*$  of the function  $S_\Delta$ . For each subdivision  $\Delta$ , the smallest piecewise-linear function  $f_\Delta^*$  such that  $f_\Delta^*(x_i) \geq y_{\Delta,i}^*$  for  $i = 1, \dots, 5$  is depicted in Fig. 5. We have  $\int_P \exp(f_\Delta^*(t)) dt = 1$  for all subdivisions  $\Delta$ . This implies that if  $y_\Delta^*$  is not relevant, then  $\exp(h_{X,y_\Delta^*})$  is not a distribution.

The optimal tent function is supported on the subdivision  $\{\{1, 3\}, \{3, 5\}\}$ . Also subdivisions  $\{\{1, 4\}, \{4, 5\}\}$  and  $\{\{1, 5\}\}$  give concave piecewise-linear functions  $f_\Delta^*$ , however, the value of  $S_\Delta$  at  $y_\Delta^*$  is less for these subdivisions (respectively  $-2.32524$  and  $-2.32556$ ) than for the optimal subdivision ( $-2.31007$ ). Moreover, only for the optimal subdivision we obtain  $y_\Delta^*$  that is close to the optimal  $y_\Delta^*$  obtained by **LogConcDEAD**. In this example, **LogConcDEAD** gives  $y_1^* = -1.070377$ . For the eight subdivisions in Fig. 5, we get the following values for the first coordinate of  $y_\Delta^*$  using **Mathematica**: (a)  $-0.564769$  (b)  $-1.13722$  (c)  $-0.783036$  (d)  $-0.595576$  (e)  $-0.852468$  (f)  $-1.07045$  (g)  $-0.797148$  (h)  $-0.833582$ . Similarly for other coordinates of  $y^*$ , only  $y_{\{\{1,3\},\{3,5\}\}}^*$  agrees with  $y^*$  when rounded to the third decimal digit. This suggests a method for checking whether a subdivision supports the optimal tent function: The piecewise-linear function  $f_\Delta^*$  should be concave and the height vector  $y_\Delta^*$  should be close to  $y^*$  obtained by **LogConcDEAD**.

We see from this example, if a subdivision  $\Delta$  is incompatible with the optimal subdivision, then  $f_\Delta^*$  might or might not be concave. The subdivisions  $\{\{1, 4\}, \{4, 5\}\}$  and  $\{\{1, 2\}, \{2, 5\}\}$  are both incompatible with the subdivision  $\{\{1, 3\}, \{3, 5\}\}$ , and  $f_{\{\{1,4\},\{4,5\}\}}^*$  is concave whereas  $f_{\{\{1,2\},\{2,5\}\}}^*$  is not concave. In all examples that we have done, if a subdivision  $\Delta$  refines the optimal subdivision, then  $f_\Delta^*$  is not concave and if a subdivision  $\Delta$  is coarser than the optimal subdivision, then  $f_\Delta^*$  is concave. Whether this is true in general, is left as an open question.

### 3. Transcendentality and closed-form solutions

In this section we use notions from geometric combinatorics to study the structure of (2.6). In particular, we will prove that the matrix  $A$  is invertible. This will be our main tool in proving the transcendentality of log-concave MLE and deriving closed form solutions in the one-dimensional one cell case using Lambert functions.

#### 3.1. Score equation matrix invertibility and transcendentality

Towards proving transcendentality, we first investigate the invertibility of the matrix  $A$ .

**Theorem 3.1.** *Consider a point configuration  $X = (x_1, \dots, x_n)$  in  $\mathbb{R}^d$ , let  $\Delta = \{\sigma_1, \dots, \sigma_m\}$  be a maximal regular triangulation of  $X$ . The score equation matrix  $A$  from (2.5) is invertible.*

**Definition 3.2.** Given a triangulation  $\Delta$ , we define the *neighborhood*  $\mathcal{N}(j)$  of a vertex  $j$  in  $\Delta$  to be the set of vertices

$$\mathcal{N}(j) = \{i : (i, j) \in \sigma_k \text{ for some } k\}.$$

Before giving the proof of Theorem 3.1, we illustrate the construction in the proof with a small example.

**Example 3.3.** Let  $X = (x_1, x_2, x_3, x_4)$  be a four point configuration in  $\mathbb{R}^2$  with  $\Delta = \{\sigma_1, \sigma_2\}$ , where  $\sigma_1 = \{1, 2, 3\}$  and  $\sigma_2 = \{2, 3, 4\}$ . Let  $A$  be the score equation matrix for the entire regular triangulation  $\Delta$ . Let us denote the difference  $y_i - y_j$  by  $y_{ij}$ . Then  $A = A(\sigma_1) + A(\sigma_2)$ , where

$$\frac{A(\sigma_1)}{\text{vol}(\sigma_1)} = \begin{bmatrix} \frac{1}{y_{12}y_{13}} - \frac{1}{y_{12}^2y_{13}} - \frac{1}{y_{12}y_{13}^2} & -\frac{1}{y_{21}^2y_{23}} & -\frac{1}{y_{31}^2y_{32}} & 0 \\ \frac{1}{y_{12}^2y_{13}} & \frac{1}{y_{21}y_{23}} - \frac{1}{y_{21}^2y_{23}} - \frac{1}{y_{21}y_{23}^2} & -\frac{1}{y_{31}y_{32}^2} & 0 \\ \frac{1}{y_{12}y_{13}^2} & -\frac{1}{y_{21}^2y_{23}^2} & \frac{1}{y_{31}y_{32}} - \frac{1}{y_{31}^2y_{32}} - \frac{1}{y_{31}y_{32}^2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\frac{A(\sigma_2)}{\text{vol}(\sigma_2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{y_{23}y_{24}} - \frac{1}{y_{23}^2y_{24}} - \frac{1}{y_{23}y_{24}^2} & -\frac{1}{y_{32}^2y_{34}} & -\frac{1}{y_{42}^2y_{43}} \\ 0 & \frac{1}{y_{23}^2y_{24}} & \frac{1}{y_{32}y_{34}} - \frac{1}{y_{32}^2y_{34}} - \frac{1}{y_{32}y_{34}^2} & -\frac{1}{y_{42}y_{43}^2} \\ 0 & \frac{1}{y_{23}y_{24}^2} & -\frac{1}{y_{32}y_{34}^2} & \frac{1}{y_{42}y_{43}} - \frac{1}{y_{42}^2y_{43}} - \frac{1}{y_{42}y_{43}^2} \end{bmatrix}.$$

We define matrix  $B$  to be the matrix  $A$  with its  $j$ -th column multiplied by  $\prod_{i \in \mathcal{N}(j)} y_{ji}^2$ , for all  $j$  from 1 to 4. We obtain the following matrices

$$\frac{B(\sigma_1)}{\text{vol}(\sigma_1)} = \begin{bmatrix} y_{13}y_{12} - y_{12} - y_{13} & y_{24}^2y_{23} & y_{34}^2y_{32} & 0 \\ y_{13} & y_{21}y_{23}y_{24}^2 - y_{24}^2y_{21} - y_{23}y_{24}^2 & y_{34}^2y_{31} & 0 \\ y_{12} & y_{24}^2y_{21} & y_{31}y_{32}y_{34}^2 - y_{34}^2y_{31} - y_{34}^2y_{32} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\frac{B(\sigma_2)}{\text{vol}(\sigma_2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & y_{21}^2y_{23}y_{24} - y_{21}^2y_{23} - y_{21}^2y_{24} & y_{31}^2y_{34} & y_{43} \\ 0 & y_{21}^2y_{24} & y_{32}y_{31}^2y_{34} - y_{31}^2y_{32} - y_{31}^2y_{34} & y_{42} \\ 0 & y_{21}^2y_{23} & y_{31}^2y_{32} & y_{43}y_{42} - y_{42} - y_{43} \end{bmatrix}.$$

The product of the diagonal entries of  $B = B(\sigma_1) + B(\sigma_2)$  is a polynomial of degree 12. Whereas a term in the expansion of the determinant of  $B$  with off-diagonal entries has at most degree 10.

**Proof of Theorem 3.1.** The score equation matrix  $A$  associated to a maximal regular triangulation  $\Delta$  can be written as

$$A = \sum_{\sigma \in \Delta} A(\sigma),$$

where the entries of  $A(\sigma)$  for  $i \neq j$  are

$$A(\sigma)_{i,j} = \text{vol}(\sigma) \left( \prod_{\alpha \in \sigma \setminus \{j\}} \frac{1}{(y_j - y_\alpha)} \right) \left( \frac{1}{y_j - y_i} \right),$$

$$A(\sigma)_{j,j} = \text{vol}(\sigma) \left( \prod_{\alpha \in \sigma \setminus \{j\}} \frac{1}{(y_j - y_\alpha)} \right) \left( 1 - \sum_{\alpha \in \sigma \setminus \{j\}} \frac{1}{(y_j - y_\alpha)} \right).$$

The matrix  $A(\sigma)$  is sparse: If  $i$  or  $j$  does not belong to  $\sigma$  then  $A_{i,j}(\sigma) = 0$ .

Let  $B$  (resp.  $B(\sigma)$ ) be the matrix that is obtained by multiplying the  $j$ -th column of  $A$  (resp.  $A(\sigma)$ ) by  $\left( \prod_{\alpha \in \mathcal{N}(j)} (y_j - y_\alpha)^2 \right)$  for  $j = 1, \dots, n$ :

$$B_{.,j} = A_{.,j} \left( \prod_{\alpha \in \mathcal{N}(j)} (y_j - y_\alpha)^2 \right) = \sum_{\sigma \in \Delta} A(\sigma)_{.,j} \left( \prod_{\alpha \in \mathcal{N}(j)} (y_j - y_\alpha)^2 \right). \quad (3.1)$$

Fix  $\sigma \in \Delta$ . We describe separately the off-diagonal and diagonal entries of  $B(\sigma)$ . For  $i, j \in \sigma$  and  $i \neq j$  we get

$$\begin{aligned} B(\sigma)_{i,j} &= A(\sigma)_{i,j} \left( \prod_{\alpha \in \sigma \setminus \{j\}} (y_j - y_\alpha)^2 \right) \left( \prod_{\alpha \in \mathcal{N}(j) \setminus \sigma} (y_j - y_\alpha)^2 \right) \\ &= \frac{\text{vol}(\sigma)}{y_j - y_i} \left( \prod_{\alpha \in \sigma \setminus \{j\}} \frac{1}{(y_j - y_\alpha)} \prod_{\alpha \in \sigma \setminus \{j\}} (y_j - y_\alpha)^2 \right) \left( \prod_{\alpha \in \mathcal{N}(j) \setminus \sigma} (y_j - y_\alpha)^2 \right) \\ &= \text{vol}(\sigma) \left( \prod_{\alpha \in \sigma \setminus \{i,j\}} (y_j - y_\alpha) \right) \left( \prod_{\alpha \in \mathcal{N}(j) \setminus \sigma} (y_j - y_\alpha)^2 \right). \end{aligned}$$

And for the diagonal entries

$$\begin{aligned} B(\sigma)_{j,j} &= A(\sigma)_{j,j} \left( \prod_{\alpha \in \mathcal{N}(j)} (y_j - y_\alpha)^2 \right) \\ &= \text{vol}(\sigma) \left( \prod_{\alpha \in \sigma \setminus \{j\}} \frac{1}{(y_j - y_\alpha)} \right) \left( 1 - \sum_{\alpha \in \sigma \setminus \{j\}} \frac{1}{(y_j - y_\alpha)} \right) \left( \prod_{\alpha \in \mathcal{N}(j)} (y_j - y_\alpha)^2 \right) \\ &= \text{vol}(\sigma) \left( \prod_{\alpha \in \sigma \setminus \{j\}} (y_j - y_\alpha) - \sum_{k \in \sigma \setminus \{j\}} \prod_{\alpha \in \sigma \setminus \{j,k\}} (y_j - y_\alpha) \right) \left( \prod_{\alpha \in \mathcal{N}(j) \setminus \sigma} (y_j - y_\alpha)^2 \right). \end{aligned}$$

Given a polynomial  $f \in \mathbb{R}[y_1, \dots, y_n]$ , we can rewrite  $f = \sum_{i=0}^{d_j} f_i y_j^i$  as a univariate polynomial in  $y_j$  of degree  $d_j$ , where  $f_i \in \mathbb{R}[y_i : i \neq j]$  is a constant with respect to  $y_j$ . We then define the *initial form* of  $f$  with respect to  $j$  to be

$$\text{in}_j(f) = f_{d_j} y_j^{d_j}.$$

We observe that for the off-diagonal entries  $B(\sigma)_{i,j}$ , the initial form with respect to  $j$  is

$$\text{in}_j(B(\sigma)_{i,j}) = y_j^{2\gamma_j-d-1},$$

where  $\gamma_j = |\mathcal{N}(j)|$  is the number of vertices adjacent to  $j$  in  $\Delta$ . Whereas for the diagonal entry  $B(\sigma)_{j,j}$ , the initial form is

$$\text{in}_j(B(\sigma)_{j,j}) = y_j^{2\gamma_j-d}.$$

In both cases, the degree of the initial form is the degree of the polynomial. We sum the matrices  $B(\sigma)$  for  $\sigma \in \Delta$ , to get  $B$  and note that the coefficient of the monomial  $y_j^{2\gamma_j-d}$  in  $B_{j,j}$  is the number of simplices in  $\Delta$  containing vertex  $j$ . Hence, using the Leibniz formula to compute the determinant of  $B$ , we get that the product of diagonal entries is a polynomial of degree  $\left(\sum_{j=1}^n 2\gamma_j - d\right)$ . All off-diagonal entries in that column of  $B$  are of degree one smaller, thus any monomial in the expanded form of the determinant with off-diagonal entries must have degree at least two smaller than the product of diagonal entries. The following equality is a direct consequence of (3.1)

$$\det(B) = \det(A) \prod_{j=1}^n \left( \prod_{\alpha \in \mathcal{N}(j)} (y_j - y_\alpha)^2 \right).$$

Since  $\det(B)$  is not identically 0,  $\det(A)$  is not identically zero, hence  $A$  is invertible over the field of rational functions.  $\square$

The proof of Theorem 3.1 inspires the following conjecture about the combinatorial properties of the determinant.

**Conjecture 3.4.** *The sum over terms of highest total degree of the numerator of  $\det(A)$  is*

$$\prod_{j=1, \dots, n} \left( \sum_{\sigma \in \Delta \text{ s.t. } j \in \sigma} \text{vol}(\sigma) \prod_{\alpha \in \mathcal{N}(j): \alpha \notin \sigma} (y_j - y_\alpha) \right).$$

Since  $A$  is invertible, (2.5) can be rewritten as

$$e^y = A^{-1}w$$

where entries of  $A$  are rational functions in  $\mathbb{R}(y_1, \dots, y_n)$ .

**Corollary 3.5.** *Fix a maximal triangulation  $\Delta$ . Then the critical equations (2.4) can be written in the form*

$$\begin{aligned}
 \exp(y_1) &= p_1(y_1, y_2, \dots, y_n) \\
 \exp(y_2) &= p_2(y_1, y_2, \dots, y_n) \\
 &\vdots \\
 \exp(y_n) &= p_n(y_1, y_2, \dots, y_n)
 \end{aligned} \tag{3.2}$$

where  $p_1, \dots, p_n \in \mathbb{R}(y_1, \dots, y_n)$ . If  $x_1, \dots, x_n \in \mathbb{Q}^d$ , then  $p_1, \dots, p_n \in \mathbb{Q}(y_1, \dots, y_n)$ .

We will explore rational-exponential systems of the form (3.2) further in Sections 3.2–3.3. The following is a result from transcendental number theory, for a textbook reference see Theorem 1.4 of [5].

**Theorem 3.6** (Lindemann-Weierstrass). *If  $y_1, \dots, y_r$  are distinct algebraic numbers then the numbers  $\exp(y_1), \dots, \exp(y_r)$  are linearly independent over the algebraic numbers.*

A special case of the Lindemann-Weierstrass theorem is the Lindemann theorem which states that  $\exp(y)$  is transcendental for algebraic  $y \neq 0$ .

**Theorem 3.7.** *Assume that  $X \subset \mathbb{Q}^d$ . If  $\text{vol}(\text{conv}(X)) \neq 1$  then at least one coordinate of the optimal height vector  $y^*$  is transcendental. If  $\text{vol}(\text{conv}(X)) = 1$ , then all coordinates of  $y^*$  are algebraic if and only if  $w$  is in the cone over the secondary polytope  $\Sigma(X)$ .*

**Proof.** It follows from the proof of [49, Lemma 3.4] that any relevant  $y^* \in \mathbb{R}^n$  such that  $\exp(h_{X,y^*})$  is a density, is a critical point of  $S_\Delta(y_1, \dots, y_n)$  for a maximal regular triangulation  $\Delta$  and some weight vector  $w$ . We consider the rational-exponential system (3.2) for this choice of  $\Delta$  and  $w$ . Then we have  $\exp(y_1) = p_1(y_1, \dots, y_n)$  where  $p_1$  is a rational function in  $\mathbb{Q}(y_1, \dots, y_n)$ . Assume that  $y_1, \dots, y_n$  are algebraic. By Lindemann's theorem  $\exp(y_1)$  is algebraic if and only if  $y_1 = 0$ .

However,  $p(y_1, \dots, y_n)$  is always algebraic, since  $y_1, \dots, y_n$  are algebraic and the algebraic numbers form a field. Hence  $y_1 = 0$ . We can argue similarly that  $y_i = 0$  for all  $i$ . The vector  $y = (0, \dots, 0)$  belongs to the boundary of the Samworth body if and only if the volume of the convex hull of  $X$  is 1. In this case,  $y$  is the optimal solution for any  $w$  in the cone over the secondary polytope  $\Sigma(X)$  by [49, Corollary 3.9].  $\square$

### 3.2. One cell in one dimension

In this section we apply the invertibility of the score equation matrix to give a closed form solution to log-concave maximum likelihood estimator in case the logarithm of the optimal density is a linear function on the real line. If  $X = (x_1, x_2) \subset \mathbb{R}$ , then

$$A = \text{vol}(\sigma) \begin{bmatrix} \frac{1}{y_1 - y_2} - \frac{1}{(y_1 - y_2)^2} & \frac{1}{(y_1 - y_2)^2} \\ \frac{1}{(y_1 - y_2)^2} & -\frac{1}{y_1 - y_2} - \frac{1}{(y_1 - y_2)^2} \end{bmatrix}$$



and

$$A^{-1} = \frac{1}{\text{vol}(\sigma)} \begin{bmatrix} 1 + y_1 - y_2 & 1 \\ 1 & 1 - y_1 + y_2 \end{bmatrix}.$$

Hence the polynomial-exponential system (3.2) has the form

$$\exp(y_1) = \frac{1}{\text{vol}(\sigma)} ((1 + y_1 - y_2) w_1 + w_2) \quad (3.3)$$

$$\exp(y_2) = \frac{1}{\text{vol}(\sigma)} (w_1 + (1 - y_1 + y_2) w_2) \quad (3.4)$$

Dividing (3.3) by (3.4) and setting  $y_{12} = y_1 - y_2$ , gives

$$\exp(y_{12}) = \frac{(1 + y_{12}) w_1 + w_2}{w_1 + (1 - y_{12}) w_2}. \quad (3.5)$$

In the rest of the section we will discuss how to solve Equation (3.5) using Lambert functions. The solutions for  $y_1$  and  $y_2$  can then be obtained from Equations (3.3) and (3.4) by solving for  $y_{12}$ .

**Definition 3.8** (Section 2 in [42]). For  $x, t_i, s_j \in \mathbb{R}$ , consider the function

$$\exp(x) \frac{(x - t_1)(x - t_2) \dots (x - t_n)}{(x - s_1)(x - s_2) \dots (x - s_m)}.$$

We denote its (generally multi-valued) inverse function at the point  $a \in \mathbb{R}$  by

$$W(t_1, t_2, \dots, t_n; s_1, s_2, \dots, s_m; a)$$

and call it *the generalized W-Lambert function*. The function  $W(a) := W(0; ; a)$  is called the usual *W-Lambert function*.

We have  $W(; ; a) = \log(a)$ .

**Proposition 3.9.** *The tent poles corresponding to a single-cell triangulation in 1 dimension are given by:*

$$\begin{aligned} y_1 &= \log(w_1 W(\rho + 1; -\rho^{-1} - 1; -\rho) + w_1 + w_2) - \log(\text{vol}(\sigma)), \\ y_2 &= \log(-w_2 W(\rho + 1; -\rho^{-1} - 1; -\rho) + w_1 + w_2) - \log(\text{vol}(\sigma)), \end{aligned}$$

where  $\rho = w_1/w_2$  and  $W(\rho + 1; -\rho^{-1} - 1; -\rho)$  is a value of the multi-valued generalized Lambert  $W$  function if  $y_1 \neq y_2$ . Otherwise  $y = (-\log(\text{vol}(\sigma)), -\log(\text{vol}(\sigma)))$ .

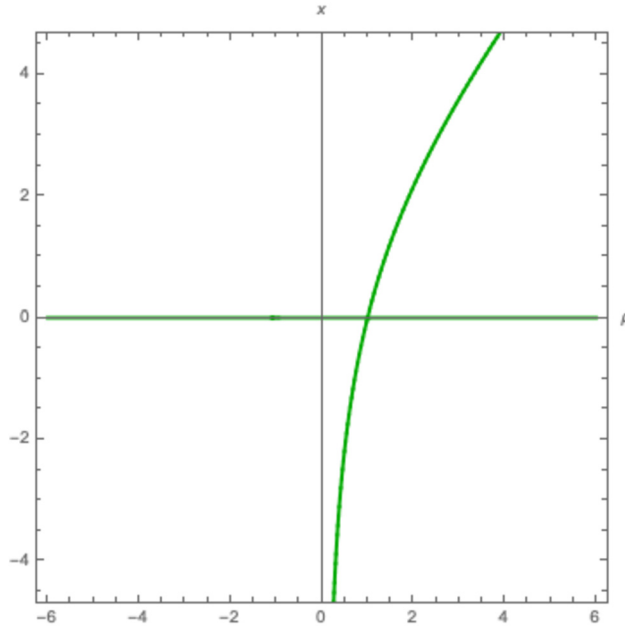


Fig. 6. Generalized Lambert function  $W(\rho + 1; -\rho^{-1} - 1; -\rho)$ .

**Proof.** Recall from Equation (3.5):

$$\exp(y_{12}) = \frac{w_1 y_{12} + w_1 + w_2}{-w_2 y_{12} + w_1 + w_2}$$

or, by setting  $\rho = w_1/w_2$ , equivalently

$$\frac{y_{12} - \rho - 1}{y_{12} + \rho^{-1} + 1} \exp(y_{12}) = -\rho.$$

Seen as an equation in  $y_{12}$  this has solutions given by the generalized Lambert function  $W(\rho + 1; -\rho^{-1} - 1; -\rho)$ . The solutions for  $y_1$  and  $y_2$  can then be obtained from (3.3) and (3.4) by solving  $y_{12}$ .  $\square$

**Remark 3.10.** Proposition 3.9 generalizes to the case when we have  $n$  points on a line and the optimal tent function is supported on one cell.

The generally multi-valued generalized  $W$ -Lambert function  $W(\rho + 1; -\rho^{-1} - 1; -\rho)$  is plotted in Fig. 6. We explore its branches, i.e., single-valued functions of  $\rho$ , using  $r$ -Lambert functions.

**Definition 3.11** (Section 3.2 in [42]). If  $r \in \mathbb{R}$ , consider the function

$$x \exp(x) + rx.$$

We denote its inverse function in the point  $a \in \mathbb{R}$  by  $W_r(a)$  and call it the  $r$ -Lambert function.

The following theorem makes the connection between the generalized Lambert function and the  $r$ -Lambert function:

**Theorem 3.12** (Theorem 3 in [42]). *If  $t, s, a \in \mathbb{R}$ , the following equality holds:*

$$W(t; s; a) = t + W_{-a \exp(-t)}(a \exp(-t)(t - s)).$$

Hence

$$W(\rho + 1; -\rho^{-1} - 1; -\rho) = \rho + 1 + W_{\rho \exp(-\rho-1)}(-\rho \exp(-\rho-1)(\rho + \rho^{-1} + 2)).$$

The number of branches of the  $r$ -Lambert function is classified in [42, Theorem 4] and [39, Theorem 4]. For  $r = \rho \exp(-\rho - 1)$ , it translates to

1. two branches, if  $\rho \exp(-\rho - 1) < 0$ ;
2. three branches, if  $0 < \rho \exp(-\rho - 1) < \exp(-2)$ ;
3. one branch, if  $\rho \exp(-\rho - 1) \geq \exp(-2)$ .

The second case happens when  $\rho > 0$ , in which case we have the double branch of constant zero function and an additional branch. This is the branch that is relevant to us in the context of Proposition 3.9. The first case happens when  $\rho < 0$ , in which case there exists a double branch of the constant zero function. This cannot appear for positive weights  $w_i$ . The third case does not happen.

The  $r$ -Lambert function can be computed with the C++ implementation [41]. Alternatively, one can use results about computing roots of polynomial-exponential equations. In [38], a symbolic-numeric algorithm is proposed for constructing explicitly an interval containing all the real roots of a single real polynomial-exponential equation, and counting how many roots are contained in a non-bounded interval. In [47], the decision problem of the existence of positive roots of such functions is discussed. This subject is strongly related to quantifier elimination [60], and to transcendental problems [40,11,12]. The latter problem of the transcendence theory appears in our Theorem 3.7.

### 3.3. Two cells in one dimension

Let  $X = (x_1, x_2, x_3) \subset \mathbb{R}$ . Then

$$A = \begin{bmatrix} \frac{v_1}{(y_1-y_2)^2} - \frac{v_1}{y_1-y_2} & -\frac{v_1}{(y_1-y_2)^2} & \frac{v_1}{y_1-y_2} + \frac{v_2}{(y_2-y_3)^2} - \frac{v_2}{y_2-y_3} & 0 \\ -\frac{v_1}{(y_1-y_2)^2} & \frac{v_1}{(y_1-y_2)^2} - \frac{v_1}{y_1-y_2} + \frac{v_2}{(y_2-y_3)^2} - \frac{v_2}{y_2-y_3} & -\frac{v_2}{(y_2-y_3)^2} - \frac{v_2}{y_2-y_3} & 0 \\ 0 & -\frac{v_2}{(y_2-y_3)^2} & \frac{v_2}{(y_2-y_3)^2} - \frac{v_2}{y_2-y_3} & 0 \end{bmatrix}.$$

Recall  $y_{12} = y_1 - y_2$  and  $y_{23} = y_2 - y_3$ . Then

$$A^{-1} = \frac{1}{v_1(1+y_{23}) + v_2(1-y_{12})} \times \begin{bmatrix} -(1+y_{12})(1+y_{23}) + \frac{v_2}{v_1} y_{12}^2 & -1-y_{23} & -1 \\ -1-y_{23} & (-1+y_{12})(1+y_{23}) & -1+y_{12} \\ -1 & -1+y_{12} & -(-1+y_{12})(-1+y_{23}) + \frac{v_1}{v_2} y_{23}^2 \end{bmatrix}.$$

Consider the polynomial-exponential system  $\exp(y) = A^{-1}w$  as in (3.2). Dividing the first equality with the second one and the second one with the third one gives:

$$\begin{cases} \exp(y_{12}) = \frac{(-(1+y_{12})(1+y_{23}) + \frac{v_2}{v_1} y_{12}^2)w_1 + (-1-y_{23})w_2 - w_3}{(-1-y_{23})w_1 + (-1+y_{12})(1+y_{23})w_2 + (-1+y_{12})w_3}, \\ \exp(y_{23}) = \frac{(-1-y_{23})w_1 + (-1+y_{12})(1+y_{23})w_2 + (-1+y_{12})w_3}{-w_1 + (y_{12}-1)w_2 - ((y_{12}-1)(y_{23}-1) + \frac{v_1}{v_2} y_{23}^2)w_3}. \end{cases} \quad (3.6)$$

Hence we could reduce a polynomial-exponential system with three equations and three variables to a polynomial-exponential system with two equations and two variables. Systems of two rational bivariate polynomial-exponential equations such as (3.6) are studied in [38]. An algorithm giving the number of solutions of such a system is provided, where all the solutions are contained in a generalized open rectangle of type  $I_1 \times I_2 \subset \mathbb{R}^2$ , under the hypothesis that at least one of the intervals  $I_1$  or  $I_2$  is bounded.

**Remark 3.13.** Let  $X \subset \mathbb{R}$ . If we consider tent functions  $h_{X,y}$  that are supported on two cells such that  $h_{X,y}$  is a constant function on one of the two cells, then one can use methods similar to the one cell case (see Section 3.2) to give the optimal solution using the Lambert function.

#### 4. Certifying solutions with Smale's $\alpha$ -theory

As explained in Section 2, our task is to maximize the objective function  $S(y_1, \dots, y_n)$  defined in Corollary 2.3. For a subdivision  $\Delta$ , we can find the optimal  $y_\Delta^*$  by considering  $S_{\Delta'}(y_1, \dots, y_n)$  for any maximal triangulation  $\Delta'$  that refines  $\Delta$ , substituting  $y_i$  that can be expressed in terms of other  $y$ 's for the subdivision  $\Delta$  and solving the system of critical equations  $\partial \tilde{S}_\Delta / \partial y_i = 0$  for the resulting function  $\tilde{S}_\Delta$ . For maximal triangulations, we have  $\tilde{S}_\Delta = S_\Delta$  and the system of critical equations is given by (2.4). We will write  $S_\Delta$  instead of  $\tilde{S}_\Delta$  also when talking about general subdivisions and for brevity we denote the system of critical equations by  $\nabla S_\Delta(y) = 0$ . We say the system is square because we have  $n$  equations  $\partial S_\Delta / \partial y_i = 0$  in  $n$  variables  $y_1, \dots, y_n$ . Usually it will be impossible to write down exact solutions to these systems, but there is a way forward. In what follows we discuss the computation of certified solutions to this system of equations. To do so, we discuss Smale's  $\alpha$ -theory, which makes mathematically rigorous the idea of approximate zeros in the sense of quadratic convergence of Newton iterations. The following influential definition was given in [9,56].

**Definition 4.1** (Chapter 8 of [9]). Let  $Df(x)$  be the  $n \times n$  Jacobian matrix of the square system of complex-analytic equations  $f(x) = 0 \in \mathbb{C}^n$ , where  $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$  is written as a column vector of its component functions

$$f(x) = [f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)]^T.$$

A point  $z \in \mathbb{C}^n$  is an *approximate zero* of  $f$  if there exists a zero  $z^* \in \mathbb{C}^n$  of  $f$  such that the sequence of Newton iterates

$$z_{k+1} = z_k - Df(z_k)^{-1}f(z_k)$$

satisfies

$$\|z_{k+1} - z^*\| \leq \frac{1}{2} \|z_k - z^*\|^2$$

for all  $k \geq 1$  where  $z_0 = z$ . If this holds, then we call  $z^*$  the *associated zero* of  $z$ . Here  $\|x\| := (\sum_{i=1}^n x_i \bar{x}_i)^{\frac{1}{2}}$  is the standard norm in  $\mathbb{C}^n$ , and the zero  $z^*$  is assumed to be nonsingular, meaning that  $\det Df(z^*) \neq 0$ .

Therefore the problem becomes two-fold. Given a system of equations  $f$ , we need a way to (1) generate approximate solutions, and (2) certify their quadratic convergence under Newton iterations. The methods of Smale's  $\alpha$ -theory solve exactly this second problem. This is accomplished using the constants  $\alpha(f, x)$ ,  $\beta(f, x)$  and  $\gamma(f, x)$ , which we will discuss in Section 4.1. Typically  $\gamma$  is difficult to compute, since it is defined as the supremum of infinitely many quantities depending on higher-order derivatives of our system of equations. However, explicit upper bounds on  $\gamma$  were calculated in [31] which we can specialize to the system required for log-concave density estimation. These upper bounds have the advantage that they are easily computed from our system  $\nabla S = 0$ , and can therefore be used to  $\alpha$ -certify approximate solutions coming from numerical software. In Section 4.1, we make this precise, discussing recent work on the subject [31,32,54,56] and how it applies in our context.

**Remark 4.2.** One might wonder why we do not directly evaluate the equations in question to the approximate height values given by statistical packages. The reason is that we want to have a measure of how accurate this solution is, which is also very sensitive to the system. Consider for example the system consisting of the single polynomial  $f(x) = x$ . We would not accept  $1/2$  as a solution. But if we consider the system  $f(x) = x^{10}$  and we evaluate at  $x = 1/2$ , we get a value that is less than 0.001. This could have been tempting, but note that in both cases the difference between actual solution and approximation is the same.

Another example that illustrates the potential difficulties involved in judging a numerical solution based on its evaluation into the original system of equations comes from [8]. Consider the univariate polynomial

$$f(z) = z^{10} - 30z^9 + 2.$$

A solution which is accurate within  $9.4 \times 10^{-12}$  of the true solution is

$$z^* = 30.000000000000142 - 0.00000000000047i,$$

but evaluating the polynomial at this solution yields a complex number  $f(z^*)$  with norm  $|f(z^*)| = 31.371$ , which certainly seems far from zero. However, refining the accuracy of this solution to

$$z^{**} = 29.9999999999998983894731343124 + 0.000000000000000000000062i,$$

we find that  $|f(z^{**})| = 0.00000000032$ , which is much better.

#### 4.1. Smale's $\alpha$ -theory

The intuition behind  $\alpha$ -theory is as follows. The size of the initial Newton iteration step combined with the size of the derivatives control how quickly Newton iteration converges to a true solution. We can calculate the size of the Newton iteration step, so if we have some control over the higher order derivatives of  $f$ , then we should be able to certify whether a solution satisfies the criterion of Definition 4.1. This motivates the definition of the following constants  $\alpha, \beta, \gamma \in \mathbb{R}$ , associated to a system of equations  $f$  at a point  $x$ . These constants measure quantities relevant to certifying approximate zeros.

**Definition 4.3.** Let  $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$  be a system of complex-analytic functions and let  $x \in \mathbb{C}^n$ . We define  $\alpha(f, x)$  to be the product of  $\beta(f, x)$  and  $\gamma(f, x)$ :

$$\alpha(f, x) = \beta(f, x)\gamma(f, x).$$

The constant  $\beta(f, x)$  measures the size of the Newton iteration step applied at  $x$ , namely:

$$\beta(f, x) = \|Df(x)^{-1}f(x)\|,$$

while  $\gamma(f, x)$  bounds the sizes of the following quantities, involving the higher order derivatives:

$$\gamma(f, x) = \sup_{k \geq 2} \left\| \frac{Df(x)^{-1}D^k f(x)}{k!} \right\|^{\frac{1}{k-1}}.$$

If we can compute these constants  $\beta, \gamma$  for a candidate solution, then we can utilize the following

**Theorem 4.4** (Chapter 8 of [9]). *If  $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$  is a system of complex-analytic functions and  $x \in \mathbb{C}^n$  satisfies*

$$\alpha(f, x) < \frac{13 - 3\sqrt{17}}{4} \approx 0.157671,$$

then  $x$  is an approximate zero of  $f = 0$ .

For polynomial systems, all higher-order derivatives eventually vanish. Exactly this fact was used in [54] to derive an upper bound for  $\gamma(f, x)$  which involves the degrees of the polynomials in the system  $f$ . This is highly convenient since, even for systems of polynomials, calculating  $\gamma(f, x)$  purely based on the definition is quite a difficult task. Yet, if we are to certify candidate solutions to our system of equations, we need to calculate  $\gamma$  and  $\beta$  at our candidate  $x$ , multiply them, and hope they are below  $\approx 0.157671$ .

#### 4.2. Polynomial-exponential systems

For polynomial-exponential systems  $f$ , calculating  $\gamma(f, x)$  is even harder. However, in [31], an upper bound was computed for  $\gamma$  involving quantities more readily apparent in a given system  $f$  than what appears in the bare definition of  $\gamma$ . In fact, an upper bound for  $\gamma$  is calculated which applies to a general class of systems, as well as upper bounds for several special cases. One of these special cases can be further specialized to the system of equations  $\nabla S = 0$  arising in log-concave density estimation (this is Lemma 4.9 below). In [31] an example is given where the bounds for the special cases allowed candidate solutions to be  $\alpha$ -certified despite failure using the more general bounds. In this section we summarize the results of [31] as they relate to log-concave density estimation. First we need a few definitions.

**Definition 4.5.** For a point  $x \in \mathbb{C}^n$  define

$$\|x\|_1^2 = 1 + \|x\|^2 = 1 + \sum_{i=1}^n |x_i|^2.$$

For a polynomial  $g : \mathbb{C}^n \rightarrow \mathbb{C}$  given as  $g(x) = \sum_{|\rho| \leq d} a_\rho x^\rho$  define

$$\|g\|^2 = \frac{1}{d!} \sum_{|\rho| \leq d} \rho! \cdot (d - |\rho|)! \cdot |a_\rho|^2.$$

For a polynomial system  $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$  with  $f(x) = [f_1(x), \dots, f_n(x)]^T$ , we define

$$\|f\|^2 = \sum_{i=1}^n \|f_i\|^2.$$

We now define a quantity  $\mu(f, x)$  associated to a polynomial system which will play a role in bounding  $\gamma$  later.

**Definition 4.6.** Let  $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$  be a polynomial system with  $\deg f_i = d_i$ . Define

$$\mu(f, x) = \max \left\{ 1, \|f\| \cdot \|Df(x)^{-1} C_f(x)\| \right\}$$

where  $C_f(x)$  is the diagonal matrix

$$C_f(x) = \begin{bmatrix} d_1^{1/2} \cdot \|x\|_1^{d_1-1} & & \\ & \ddots & \\ & & d_n^{1/2} \cdot \|x\|_1^{d_n-1} \end{bmatrix}.$$

Following [31], we extend Definition 4.6 to certain polynomial-exponential systems.

**Definition 4.7.** Let  $a \in \mathbb{Z}_{\geq 0}$ ,  $\delta_i \in \mathbb{C}$ , and  $\sigma_i \in \{1, \dots, n\}$ . Consider the polynomial-exponential system

$$G(x_1, \dots, x_n, u_1, \dots, u_a) = \begin{bmatrix} P(x_1, \dots, x_n, u_1, \dots, u_a) \\ u_1 - \exp(\delta_1 x_{\sigma_1}) \\ u_2 - \exp(\delta_2 x_{\sigma_2}) \\ \vdots \\ u_a - \exp(\delta_a x_{\sigma_a}) \end{bmatrix}, \quad (4.1)$$

where  $P : \mathbb{C}^N \rightarrow \mathbb{C}^n$  is a polynomial system with  $N = n + a$  variables. Thus, the system  $G$  is a square system of size  $N$ . We write  $X := (x, u)$ . Define

$$\mu(G, X) = \max \left\{ 1, \left\| DG(x, u)^{-1} \begin{bmatrix} C_P(x, u) \|P\| & \\ & I_a \end{bmatrix} \right\| \right\}.$$

The following specializes Corollary 2.6 of [31].

**Theorem 4.8.** Let  $a \in \mathbb{Z}_{\geq 0}$ ,  $\delta_i \in \mathbb{C}$ , and  $\sigma_i \in \{1, \dots, n\}$  and consider the polynomial-exponential system (4.1). Let  $d_i = \deg P_i$  and  $D = \max d_i$ . For any  $\lambda, \theta \in \mathbb{C}$  define

$$A(\lambda, \theta) = \max \left\{ |\lambda|, \left| \frac{\lambda^2 \exp(\lambda \theta)}{2} \right| \right\}.$$

Then, for any  $X = (x, u) \in \mathbb{C}^N$  such that the Jacobian of  $G$  is invertible,

$$\gamma(G, X) \leq \mu(G, X) \left( \frac{D^{3/2}}{2\|X\|_1} + \sum_{i=1}^a A(\delta_i, x_{\sigma_i}) \right). \quad (4.2)$$

**Proof.** This is a straight-forward specialization of Corollary 2.6 of [31]. We set to zero quantities that deal with functions not relevant to log-concave density estimation.  $\square$



Therefore, reformulating our system of polynomial-exponential equations  $\nabla S_\Delta = 0$  in the format (4.1) will allow us to calculate an upper bound on  $\gamma$ , which will allow us to certify solutions to our critical equations.

**Lemma 4.9.** *Fix a maximal regular triangulation  $\Delta$ . The polynomial-exponential system  $\nabla S_\Delta = 0$  can be reformulated as a system of equations of the form (4.1), demonstrating that Theorem 4.8 applies in the context of log-concave maximum likelihood estimation.*

**Proof.** The partial derivatives  $\partial S_\Delta / \partial y_k$  are rational functions of the  $y_i$  and the  $\exp(y_i)$ . Since we set each partial derivative to zero, we can clear denominators, creating a system of equations, each of which is a polynomial in the  $y_i$  and the  $\exp(y_i)$ . Setting each  $\delta_i = 1$  in (4.1), we can replace each occurrence of  $\exp(y_i)$  with  $u_i$ , creating the polynomial system  $P(y_1, \dots, y_n, u_1, \dots, u_n)$ , hence  $a = n$  as well. Appending the equations  $u_i - \exp(y_i)$  to the system of polynomials  $P$ , we have a system of  $2n$  equations in  $2n$  unknowns. This system is of the required form in order to apply Theorem 4.8.  $\square$

Thus, we have everything we need to compute the upper bound in (4.2) for a system of critical equations  $\nabla S_\Delta = 0$  when  $\Delta$  is a maximal regular triangulation. By calculating this upper bound for a given system of equations, we can certify approximate numerical solutions obtained in any way. When  $\Delta$  is not a maximal regular triangulation, one must impose further linear constraints on some of the  $y_i$ , as was the case in Example 2.7. After simplifications, one might still end up with terms involving exponentials of fractional convex combinations of the  $y_i$ . This poses no threat for the purposes of  $\alpha$ -certification, as one may in fact use products of exponentials of the form  $e^{\beta y_i}$ . In particular, a bound for  $\gamma(G, X)$  also for these more general polynomial-exponential systems is given in [31, Corollary 2.6].

In algebraic statistics, it is common to find algebraic invariants which characterize algebraic complexity. For example, the maximum likelihood degree of a statistical model gives information about the critical points of the likelihood function of a parametric model [1]. Similarly, in nonparametric algebraic statistics, it could be the case that the combinatorial complexity of the optimal subdivision gives us information about the computational complexity of finding a numerical solution.

**Question 4.10.** Does increasing the combinatorial complexity of the optimal subdivision decrease the likelihood that the numerical output from LogConcDEAD is  $\alpha$ -certified?

We study this question experimentally in the next section. In future work, one could hope to precisely describe this phenomenon, should it exist. Of course, higher degrees, more variables, more equations will always increase the bound on  $\gamma$  we calculate, but the combinatorics should still play some role.

### 4.3. A procedure for $\alpha$ -certifying

One of our motivating questions was to determine the correct subdivision for a given data set, as was the case in Example 1.1. In this section we describe a procedure based on Smale's  $\alpha$ -theory that in principle allows us to find the certifiably correct subdivision. Recall that the objective function  $S(y_1, \dots, y_n)$  depends on a subdivision of the convex hull of the data set  $X$ . If there are  $m$  subdivisions, then there are  $m$  different objective functions  $S_1, \dots, S_m$ , and  $m$  different possible systems of equations  $\nabla S_1 = 0, \dots, \nabla S_m = 0$ . Given an estimate of a solution  $y^*$ , perhaps computed numerically using existing software, we can attempt to  $\alpha$ -certify that solution using any of these systems as input to Lemma 4.9 and Theorem 4.8. As we collect  $\alpha$ -certified critical points for the various objective functions, we can use this data to determine the correct subdivision, helping to answer our motivating question.

In practice, we have found that numerically computed solutions  $y^*$  are often not  $\alpha$ -certified, using any of the systems  $\nabla S_i = 0$ . However, using a brute-force search over all possible additional digits, we often can find one system  $\nabla S_j = 0$  to which  $y^* + \varepsilon$  is an  $\alpha$ -certified solution. Here,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$  is a vector providing additional digits of precision to each component of  $y^*$ . As we compute  $\alpha$ -values for each  $y^* + \varepsilon$ , we move in the direction which causes a decrease in the computed  $\alpha$ -value, until we are able to find an  $\alpha$ -certified  $y^* + \varepsilon$ . We describe this in the following

---

**Algorithm 1:** Testing certifiability by digit refinement.

---

**Input:** A system  $\nabla S_i = 0$  coming from the  $i$ th candidate subdivision and a candidate approximate solution  $y^* = (y_1, \dots, y_n)$ .

**Result:** A refinement of the heights  $y^* + \varepsilon$  along with alpha certification of the system, or inability to certify.

- 1 Let  $p$  be the number of trusted significant digits (in binary) of the approximate solution  $y^*$ .
  - 2 Expressing  $y^*$  in binary, compute the  $\alpha$ -value for all  $3^n$  points  $y_i + \varepsilon_i 2^{-p}$ ,  $\varepsilon_i \in \{-1, 0, 1\}$ . Keep the point with the lowest alpha value, and set this as the new  $y_i$ .
  - 3 If the alpha value is below 0.157671 stop and return the solution. If it has decreased between steps or remained the same, increase  $p$  by 1 and go to step 2. If there is no improvement for several loops in a row, stop and declare inability to certify the system.
- 

**Remark 4.11.** Here we collect a few comments on Algorithm 1.

1. We note that this brute-force search over all possible digits could be replaced by any numerical procedure for finding solutions to a given set of equations, see for instance the *refine* command in the Numerical Algebraic Geometry package for Macaulay2 [36]. For example, Newton iteration could be used on the system of equations to produce more accurate solutions, which could then be  $\alpha$ -certified. However, to compare the outputs of LogConcDEAD for problems of increasing combinatorial complexity (see Table 1), we wanted to use a completely “blind” brute-force search as described above.
2. One does not need to stop at Step 3 once a solution is certified. Repeating the loop allows increasing the precision of the solution by moving to lower  $\alpha$  values. This is in

contrast to statistical software like **LogConcDEAD** which only allows up to 7 significant digits.

3. Although precision can be added, our (first) goal with Algorithm 1 is to find the correct subdivision induced by the heights. One can test several subdivisions here, therefore we say that we test the (approximate) solution against the corresponding system of equations.
4. It might happen that the  $\alpha$ -value does not immediately decrease from one loop to the next even if we have the correct system of equations. One reason is that if the next significant digit is a zero for all heights, we are computing an  $\alpha$ -value for the same point multiple times.
5. In step 1 of the above algorithm, we let  $p$  be the number of trusted significant digits of the approximate solution  $y^*$ . We have found that several of the last digits of a solution computed with **LogConcDEAD** were incorrect, in the sense that if we start our search (in Algorithm 1) earlier in the significant digits of  $y^*$  we are able to  $\alpha$ -certify some  $y^* + \varepsilon$ . In this way, we can correct for some of the imprecision of a numerical solver.

**Example 4.12.** Consider the data set  $X = (2, 5, 7)$  with weights  $w = (\frac{1}{3}, \frac{1}{2}, \frac{1}{6})$ . With this input, the package **LogConcDEAD** returns the heights

$$y^* = (y_1, y_2, y_3) = (-1.454152, -1.605833, -1.888083),$$

suggesting that there are two regions of linearity (Fig. 7a). Let  $\Delta = \{\{1, 2\}, \{2, 3\}\}$ . We consider critical equations for

$$S_\Delta(y_1, y_2, y_3) = \frac{y_1}{3} + \frac{y_2}{2} + \frac{y_3}{6} - 3 \frac{e^{y_1} - e^{y_2}}{y_1 - y_2} - 2 \frac{e^{y_2} - e^{y_3}}{y_2 - y_3}$$

which lead to the polynomial-exponential system  $\nabla S_\Delta : \mathbb{C}^3 \rightarrow \mathbb{C}^3$  given by

$$\begin{aligned} (y_1 - y_2)^2 \frac{\partial S(y_1, y_2, y_3)}{\partial y_1} &= 0 \\ (y_1 - y_2)^2 (y_2 - y_3)^2 \frac{\partial S(y_1, y_2, y_3)}{\partial y_2} &= 0 \\ (y_2 - y_3)^2 \frac{\partial S(y_1, y_2, y_3)}{\partial y_3} &= 0, \end{aligned}$$

where we have cleared denominators. The numerical solution from **LogConcDEAD** is not immediately  $\alpha$ -certified, but after applying Algorithm 1 we obtain the  $\alpha$ -certified solution:  $y^* + \varepsilon = (y_1, y_2, y_3) = (-1.45415181, -1.60583278, -1.88808307)$ .

**Example 4.13.** We now consider the same sample  $X = (2, 5, 7)$  with uniform weights. As discussed in Example 2.7, **LogConcDEAD** output suggests that the logarithm of the optimal

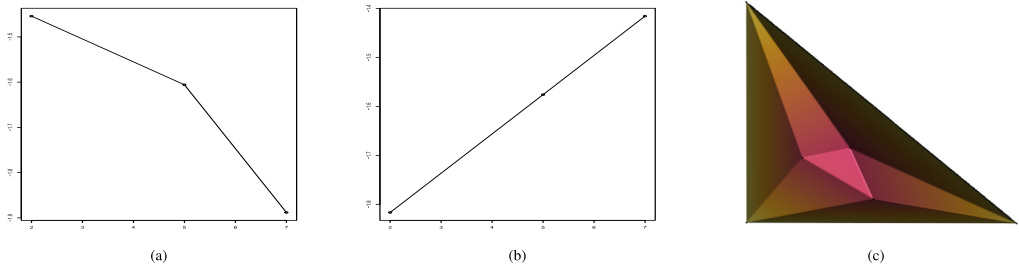


Fig. 7. The height functions for (a) Example 4.12; (b) Example 4.13; (c) Example 4.15.

Table 1

Number of binary digits needed to certify  $n + 1$  points with weights coming from an asymmetric binomial distribution.

n	3	4	5	6	7
binary digits	22	23	27	31	31

density has a single region of linearity (Fig. 7c). Can we certify this assessment? Recall that substituting  $y_2 = \frac{2}{5}y_1 + \frac{3}{5}y_3$  to  $S(y_1, y_2, y_3) = \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_3 - 3\frac{e^{y_1} - e^{y_2}}{y_1 - y_2} - 2\frac{e^{y_2} - e^{y_3}}{y_2 - y_3}$  gives

$$\tilde{S} = \frac{7}{15}y_1 + \frac{8}{15}y_3 - 5\frac{e^{y_1} - e^{y_3}}{y_1 - y_3}.$$

The system of equations  $\nabla \tilde{S} = 0$  does have solutions, and we were able to check that the numerical solution  $y^*$  computed by LogConcDEAD is an  $\alpha$ -certified solution to this amended system of equations.

**Example 4.14.** We used Algorithm 1 to certify the sample  $X = (0, 1, 2, \dots, n) \subset \mathbb{R}$  for weights given by the binomial distribution with  $p = 6/11$ , i.e.,  $w_i = \binom{n}{i}(6/11)^i(5/11)^{n-i}$ . Looking at the LogConcDEAD output, we suspect that the triangulation given by the points consists of all consecutive line segments  $\{i - 1, i\}$  for  $i \in 1, 2, \dots, n$ . We therefore compute  $\alpha$ -values using the system of equations corresponding to the full triangulation. In all cases tested, we were able to certify the system for some refinement of the original LogConcDEAD output. In Table 1, we summarize the number of binary digits required for certification in each case. This table suggests that the complexity of  $\alpha$ -certifying increases when the number of sample points increases.

We now present an example in two dimensions that needs more significant digits than the previous cases.

**Example 4.15.** We consider the point configuration from [49, Example 1.1], given by

$$X = ((0, 0), (0, 100), (22, 37), (36, 41), (43, 22), (100, 0)) \subset \mathbb{R}^2$$

and uniform weights. The package **LogConcDEAD** returns the heights

$$(y_1, y_2, y_3, y_4, y_5, y_6) \\ = (-8.789569, -8.772087, -8.253580, -8.217959, -8.236983, -8.756922)$$

as the optimal solution. This gives rise to a triangulation of the convex hull of the data points with regions of linearity consisting of the triangles

$$\{1, 2, 3\}, \{1, 3, 5\}, \{1, 5, 6\}, \{2, 3, 4\}, \{2, 4, 6\}, \{3, 4, 5\}, \{4, 5, 6\},$$

in Fig. 7b. This data gives an  $\alpha$ -value of  $10^{26}$ , which is much larger than the required 0.157671. However, the system of equations it came from has a relatively high degree and the polynomial equations, when expanded, have between 929 and 1564 terms. We try to decrease the  $\alpha$ -value using the uniform sampling algorithm described above. We create a list of  $729 = 3^6$  points in  $\mathbb{R}^6$ , consisting of all points whose  $i$ -th coordinate is

$$y_i + \epsilon_i 2^{-14}, \epsilon_i \in \{-1, 0, 1\}.$$

After a few repetitions, this finds a point with a lower alpha value. We repeat this process, each time decreasing the exponent of 2 when creating the new test points. After 95 rounds we detect the refined point

$$\begin{bmatrix} -8.789570552675578322471018111262921 \\ -8.772086862481395608253513836856700 \\ -8.253580886913590521217040193671505 \\ -8.217957742357924329528595494315867 \\ -8.236983233544571734253428918807660 \\ -8.756919956247208359690046164738877 \end{bmatrix}$$

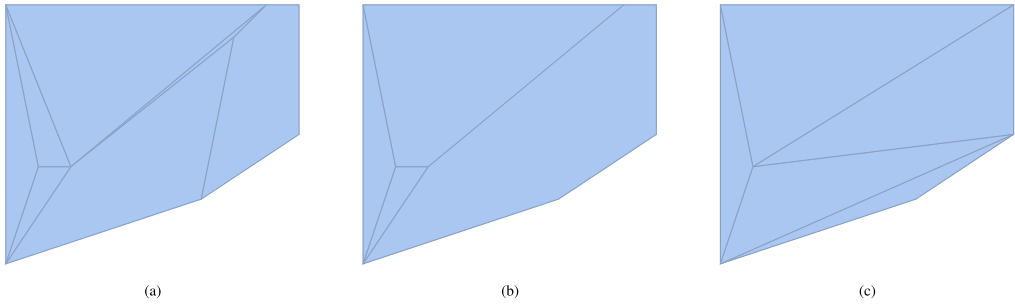
with alpha value 0.125519. Therefore, this new solution is  $\alpha$ -certified. Note that this number has 34 decimal digits; we have rounded digits coming from the conversion from base 2 (109 digits) after this position. Our conclusion is that the triangulation obtained by the heights in the **LogConcDEAD** output is certifiably correct.

**Example 4.16.** We finish our paper by returning to our motivating Example 1.1 from the introduction, and consider two possible subdivisions of  $P = \text{conv}(X)$  for the regions of linearity of the optimal tent function:

$$\Delta_1 = \{\{1, 2, 3\}, \{1, 3, 4\}, \{2, 3, 4\}, \{2, 4, 12\}, \{1, 4, 8, 11\}, \{4, 11, 12\}, \{8, 11, 12, 13, 14\}\}$$

and

$$\Delta_2 = \{\{1, 2, 3\}, \{1, 3, 4\}, \{2, 3, 4, 12\}, \{1, 4, 8, 12, 13, 14\}\}.$$



**Fig. 8.** Subdivisions in Example 4.16. (a) Subdivision  $\Delta_1$  (b) Subdivision  $\Delta_2$  (c) The subdivision induced by  $y_{\Delta_1}^*$ .

The first subdivision  $\Delta_1$  in Fig. 8a arises from the `LogConcDEAD` output with default parameters after using the “unique” function. The second subdivision  $\Delta_2$  in Fig. 8b is given by the four regions of linearity in Fig. 1 that we get by adjusting the precision in `LogConcDEAD` and then using the “unique” function. Unfortunately the objective functions involved have too many summands for  $\alpha$ -certification to be feasible.

As an alternative, we use the `NMaximize` command in `Mathematica` directly on the objective functions  $S_{\Delta_1}$  and  $S_{\Delta_2}$ . The optimal  $y_{\Delta_1}^*$  for the 7-cell subdivision gives a tent function whose regions of linearity are

$$\{\{1, 2, 3\}, \{1, 8, 13\}, \{1, 3, 13\}, \{2, 3, 14\}, \{3, 13, 14\}\},$$

which are depicted in Fig. 8c. This triangulation is not refined by the subdivision  $\Delta_1$ : For example, the triangle  $\{1, 3, 4\}$  in the subdivision  $\Delta_1$  intersects the interiors of triangles  $\{1, 3, 13\}$ ,  $\{2, 3, 14\}$ ,  $\{3, 13, 14\}$ . Thus the 7-cell subdivision  $\Delta_1$  is not the subdivision that we are looking for. In fact, the vector  $y_{\Delta_1}^*$  is not relevant, i.e. there exists  $x_i$  such that  $h_{X, y_{\Delta_1}^*}(x_i) > y_i$ , and as a result  $\int_P \exp(h_{X, y_{\Delta_1}^*}(t)) \neq 1$ .

The command `NMaximize` gives for the 4-cell subdivision

$$y_{\Delta_2}^* = (-4.32285, -4.7141, -4.2737, -4.14495, -4.26961, -4.10156, -3.94188, \\ -3.91671, -3.94162, -3.80042, -3.76397, -3.68413, -3.69541, -3.62252).$$

In comparison, the optimal height vector that we obtain using `LogConcDEAD` is

$$y^* = (-4.322797, -4.714126, -4.273678, -4.144934, -4.269616, -4.101524, -3.941869, \\ -3.916668, -3.941666, -3.800423, -3.764006, -3.684179, -3.695395, -3.622560).$$

A computation in `Polymake` verifies that  $y_{\Delta_2}^*$  gives a tent function whose regions of linearity are exactly the cells of  $\Delta_2$ . This suggests that the 4-cell subdivision  $\Delta_2$  is indeed the subdivision induced by the optimal  $y^*$  in Example 1.1. We conclude with a haiku.

Approximate heights,  
subdivisions inexact.  
A long road ahead.

## Acknowledgments

This project started at the Summer School on Geometric and Algebraic Combinatorics at Sorbonne University in June 2019. We thank Bernd Sturmfels for the guidance with the project, Gleb Pogudin for suggesting to use Lambert functions, Ricky Liu and Cynthia Vinzant for useful discussions and comments. Grosdos was partially supported by the DFG grant GK 1916, Kombinatorische Strukturen in der Geometrie. Kubjas and Kuznetsova were partially supported by the Academy of Finland Grant 323416. Scholten was partially supported by NSF Grant DMS 1620014.

## References

- [1] C. Améndola, M. Drton, B. Sturmfels, Maximum likelihood estimates for Gaussian mixtures are transcendental, in: *Int. Conf. on Math. Asp. of Comput. and Inf. Sci.*, Springer, 2015, pp. 579–590.
- [2] B. Axelrod, I. Diakonikolas, A. Stewart, A. Sidiropoulos, G. Valiant, A polynomial time algorithm for log-concave maximum likelihood via locally exponential families, in: *Adv. Neural Inform. Process. Syst.* 32, Curran Associates, Inc., 2019, pp. 7723–7735.
- [3] M. Ayer, H.D. Brunk, G.M. Ewing, W.T. Reid, E. Silverman, An empirical distribution function for sampling with incomplete information, *Ann. Math. Stat.* 26 (1955) 641–647, <https://doi.org/10.1214/aoms/1177728423>.
- [4] M. Bagnoli, T. Bergstrom, Log-concave probability and its applications, *Econom. Theory* 26 (2) (2005) 445–469, <https://doi.org/10.1007/s00199-004-0514-4>.
- [5] A. Baker, *Transcendental Number Theory*, 2nd edition, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1990.
- [6] F. Balabdaoui, J.A. Wellner, Estimation of a  $k$ -monotone density, part 1: characterizations consistency, and minimax lower bounds, *arXiv preprint*, [arXiv:math/0509080](https://arxiv.org/abs/math/0509080), 2005.
- [7] F. Balabdaoui, J.A. Wellner, Estimation of a  $k$ -monotone density: limit distribution theory and the spline connection, *Ann. Stat.* 35 (6) (2007) 2536–2564, <https://doi.org/10.1214/009053607000000262>.
- [8] D.J. Bates, J.D. Hauenstein, A.J. Sommese, C.W. Wampler, *Numerically Solving Polynomial Systems with Bertini*, Software, Environments, and Tools, vol. 25, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.
- [9] L. Blum, F. Cucker, M. Shub, S. Smale, *Complexity and Real Computation*, Springer-Verlag, New York, 1998, with a foreword by Richard M. Karp.
- [10] T. Carpenter, I. Diakonikolas, A. Sidiropoulos, A. Stewart, Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities, in: S. Bubeck, V. Perchet, P. Rigollet (Eds.), *Proc. 31 Conf. Learn. Theory*, in: *Proceedings of Machine Learning Research*, PMLR, vol. 75, 2018, pp. 1234–1262, <http://proceedings.mlr.press/v75/carpenter18a.html>.
- [11] V. Chonev, J. Ouaknine, J. Worrell, On recurrent reachability for continuous linear dynamical systems, in: *Proc. of the 31st Annual ACM-IEEE Symposium on Logic in Computer Science (LICS 2016)*, ACM, New York, 2016, p. 10.
- [12] V. Chonev, J. Ouaknine, J. Worrell, On the Skolem problem for continuous linear dynamical systems, in: *43rd International Colloquium on Automata, Languages, and Programming*, in: *LIPICs. Leibniz Int. Proc. Inform.*, vol. 55, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, 100.
- [13] M. Cule, R. Samworth, Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density, *Electron. J. Stat.* 4 (2010) 254–270, <https://doi.org/10.1214/09-EJS505>.
- [14] M. Cule, R. Gramacy, R. Samworth, LogConcDEAD: an R package for maximum likelihood estimation of a multivariate log-concave density, *J. Stat. Softw.* 29 (2) (2009), <http://www.jstatsoft.org/v29/i02/>.

- [15] M. Cule, R. Samworth, M. Stewart, Maximum likelihood estimation of a multi-dimensional log-concave density, *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 72 (5) (2010) 545–607, <https://doi.org/10.1111/j.1467-9868.2010.00753.x>.
- [16] A. De, P.M. Long, R.A. Servedio, Density estimation for shift-invariant multidimensional distributions, in: *10th Innovations in Theoretical Computer Science*, in: *LIPICs. Leibniz Int. Proc. Inform.*, vol. 124, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019, 28.
- [17] J.A. De Loera, J. Rambau, F. Santos, *Triangulations. Structures for Algorithms and Applications, Algorithms and Computation in Mathematics*, vol. 25, Springer-Verlag, Berlin, 2010.
- [18] S. Dharmadhikari, K. Joag-Dev, *Unimodality, Convexity, and Applications*, Elsevier, 1988.
- [19] I. Diakonikolas, D.M. Kane, A. Stewart, Learning multivariate log-concave distributions, in: S. Kale, O. Shamir (Eds.), *Proc. 2017 Conf. Learn. Theory*, in: *Proceedings of Machine Learning Research*, vol. 65, PMLR, Amsterdam, Netherlands, 2017, pp. 711–727, <http://proceedings.mlr.press/v65/diakonikolas17a.html>.
- [20] C.R. Doss, J.A. Wellner, Global rates of convergence of the MLEs of log-concave and  $s$ -concave densities, *Ann. Stat.* 44 (3) (2016) 954–981, <https://doi.org/10.1214/15-AOS1394>.
- [21] L. Dümbgen, K. Rufibach, Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency, *Bernoulli* 15 (1) (2009) 40–68, <https://doi.org/10.3150/08-BEJ141>.
- [22] L. Dümbgen, K. Rufibach, logcondens: computations related to univariate log-concave density estimation, *J. Stat. Softw.* 39 (6) (2011) 1–28, <http://www.jstatsoft.org/v39/i06/>.
- [23] L. Dümbgen, R. Samworth, D. Schuhmacher, Approximation by log-concave distributions, with applications to regression, *Ann. Stat.* 39 (2) (2011) 702–730, <https://doi.org/10.1214/10-AOS853>.
- [24] P.P.B. Eggermont, V.N. LaRiccia, Maximum likelihood estimation of smooth monotone and unimodal densities, *Ann. Stat.* 28 (3) (2000) 922–947, <https://doi.org/10.1214/aos/1015952005>.
- [25] P.P.B. Eggermont, V.N. LaRiccia, *Maximum Penalized Likelihood Estimation. Vol. I: Density Estimation*, Springer Series in Statistics, Springer-Verlag, New York, 2001.
- [26] E. Fix, J. Hodges Jr, *Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties*, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
- [27] U. Grenander, On the theory of mortality measurement. II, *Skand. Aktuarietidskr.* 39 (1956) 125–153, <https://doi.org/10.1080/03461238.1956.10414944>, (1957).
- [28] P. Groeneboom, G. Jongbloed, Nonparametric estimation under shape constraints, in: *Estimators, Algorithms and Asymptotics*, in: *Cambridge Series in Statistical and Probabilistic Mathematics*, vol. 38, Cambridge University Press, New York, 2014.
- [29] P. Groeneboom, G. Jongbloed, J.A. Wellner, Estimation of a convex function: characterizations and asymptotic theory, *Ann. Stat.* 29 (6) (2001) 1653–1698, <https://doi.org/10.1214/aos/1015345958>.
- [30] A. Grosdos, A. Heaton, K. Kubjas, O. Kuznetsova, G. Scholten, M.-S. Sorea, Github repository, <https://github.com/agrosdos/Computing-the-Exact-LogConcave-MLE>, 2020.
- [31] J.D. Hauenstein, V. Levandovskyy, Certifying solutions to square systems of polynomial-exponential equations, *J. Symb. Comput.* 79 (part 3) (2017) 575–593, <https://doi.org/10.1016/j.jsc.2016.07.020>.
- [32] J.D. Hauenstein, F. Sottile, Algorithm 921: alphaCertified: certifying solutions to polynomial systems, *ACM Trans. Math. Softw.* 38 (4) (2012), <https://doi.org/10.1145/2331130.2331136> 28.
- [33] A.K.H. Kim, R.J. Samworth, Global rates of convergence in log-concave density estimation, *Ann. Stat.* 44 (6) (2016) 2756–2779, <https://doi.org/10.1214/16-AOS1480>.
- [34] A.K.H. Kim, A. Guntuboyina, R.J. Samworth, Adaptation in log-concave density estimation, *Ann. Stat.* 46 (5) (2018) 2279–2306, <https://doi.org/10.1214/17-AOS1619>.
- [35] G. Kur, Y. Dagan, A. Rakhlin, Optimality of maximum likelihood for log-concave density estimation and bounded convex regression, *arXiv preprint*, [arXiv:1903.05315](https://arxiv.org/abs/1903.05315), 2019.
- [36] A. Leykin, Numerical algebraic geometry, *J. Softw. Algebra Geom.* 3 (2011) 5–10, <https://doi.org/10.2140/jsag.2011.3.5>.
- [37] Y. Liu, Y. Wang, cnmld: maximum likelihood estimation of a log-concave density function, r package version 1.2-0, <https://CRAN.R-project.org/package=cnmld>, 2018.
- [38] A. Maignan, Solving one and two-dimensional exponential polynomial systems, in: *Proc. of the 1998 Int. Symp. Symb. and Algebr. Comput.*, ISSAC '98, ACM, New York, NY, USA, 1998, pp. 215–221.
- [39] A. Maignan, T.C. Scott, Fleshing out the generalized Lambert  $W$  function, *ACM Commun. Comput. Algebra* 50 (2) (2016) 45–60, <https://doi.org/10.1145/3003653.3003662>.
- [40] S. McCallum, V. Weispfenning, Deciding polynomial-transcendental problems, *J. Symb. Comput.* 47 (1) (2012) 16–31, <https://doi.org/10.1016/j.jsc.2011.08.004>.
- [41] I. Mezö, *r-Lambert-function*, <https://github.com/IstvanMezo/r-Lambert-function>, 2017.



- [42] I. Mező, A. Baricz, On the generalization of the Lambert  $W$  function, *Trans. Am. Math. Soc.* 369 (11) (2017) 7917–7934, <https://doi.org/10.1090/tran/6911>.
- [43] J.K. Pal, M. Woodroffe, M. Meyer, Estimating a Polya frequency function<sub>2</sub>, in: *Complex Datasets and Inverse Problems*, in: *IMS Lecture Notes Monogr. Ser.*, Inst. Math., vol. 54, Statist, Beachwood, OH, 2007, pp. 239–249.
- [44] E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Stat.* 33 (1962) 1065–1076, <https://doi.org/10.1214/aoms/1177704472>.
- [45] F. Rathke, Fast multivariate log-concave density estimation in **r**, <https://github.com/FabianRathke/fmlogcondens>, 2018.
- [46] F. Rathke, C. Schnörr, Fast multivariate log-concave density estimation, *Comput. Stat. Data Anal.* 140 (2019) 41–58, <https://doi.org/10.1016/j.csda.2019.04.005>.
- [47] D. Richardson, Roots of real exponential functions, *J. Lond. Math. Soc. (2)* 28 (1) (1983) 46–56, <https://doi.org/10.1112/jlms/s2-28.1.46>.
- [48] E. Robeva, B. Sturmfels, N. Tran, C. Uhler, Maximum likelihood estimation for totally positive log-concave densities, *Scand. J. Stat.* (2018), <https://doi.org/10.1111/sjos.12462>.
- [49] E. Robeva, B. Sturmfels, C. Uhler, Geometry of log-concave density estimation, *Discrete Comput. Geom.* 61 (1) (2019) 136–160, <https://doi.org/10.1007/s00454-018-0024-y>.
- [50] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Ann. Math. Stat.* 27 (1956) 832–837, <https://doi.org/10.1214/aoms/1177728190>.
- [51] K. Rufibach, Computing maximum likelihood estimators of a log-concave density function, *J. Stat. Comput. Simul.* 77 (7–8) (2007) 561–574, <https://doi.org/10.1080/10629360600569097>.
- [52] R.J. Samworth, Recent progress in log-concave density estimation, *Stat. Sci.* 33 (4) (2018) 493–509, <https://doi.org/10.1214/18-STS666>.
- [53] D.W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd edition, *Wiley Series in Probability and Statistics*, John Wiley & Sons, Hoboken, New Jersey, 2015.
- [54] M. Shub, S. Smale, Complexity of Bézout’s theorem. I. Geometric aspects, *J. Am. Math. Soc.* 6 (2) (1993) 459–501, <https://doi.org/10.2307/2152805>.
- [55] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, 1986.
- [56] S. Smale, Newton’s method estimates from data at one point, in: *The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics*, Springer, New York, 1986, pp. 185–196.
- [57] A.B. Tsybakov, *Introduction to Nonparametric Estimation*, *Springer Series in Statistics*, Springer, New York, 2009, revised and extended from the 2004 French original. Translated by Vladimir Zaiats.
- [58] G. Walter, J. Blum, Probability density estimation using delta sequences, *Ann. Stat.* 7 (2) (1979) 328–340, <http://www.jstor.org/stable/2958814>.
- [59] G. Walther, Detecting the presence of mixing with multiscale maximum likelihood, *J. Am. Stat. Assoc.* 97 (458) (2002) 508–513, <https://doi.org/10.1198/016214502760047032>.
- [60] M. Xu, Z.-B. Li, L. Yang, Quantifier elimination for a class of exponential polynomial formulas, *J. Symb. Comput.* 68 (2015) 146–168, <https://doi.org/10.1016/j.jsc.2014.09.015>.