Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Maximum Likelihood Estimation for Totally Positive Log-Concave Densities

by

Elina Robeva, Bernd Sturmfels, Ngoc Tran, and Caroline Uhler

Preprint no.: 42 2018

MAXIMUM LIKELIHOOD ESTIMATION FOR TOTALLY POSITIVE LOG-CONCAVE DENSITIES

By Elina Robeva*, Bernd Sturmfels†,‡, Ngoc Tran§ and Caroline Uhler*

*Massachusetts Institute of Technology**, University of California at Berkeley*†*, Max Planck Institute Leipzig*‡*, University of Texas at Austin*§

We study nonparametric density estimation for two classes of multivariate distributions that imply strong forms of positive dependence; namely *multivariate totally positive distributions of order 2* (*MTP_2*, a.k.a. *log-supermodular*) and the subclass of *log-$L^\natural$-concave* (*LLC*) distributions. In both cases we impose the additional assumption of log-concavity in order to ensure boundedness of the likelihood function. Given n independent and identically distributed random vectors from a d-dimensional MTP_2 distribution (LLC distribution, respectively), we show that the maximum likelihood estimator (MLE) exists and is unique with probability one when $n \geq 3$ ($n \geq 2$, respectively), independent of the number d of variables. The logarithm of the MLE is a tent function in the binary setting and in $\mathbb{R}^2$ under MTP_2 and in the rational setting under LLC. We provide a conditional gradient algorithm for computing it, and we conjecture that the same convex program also yields the MLE in the remaining cases.

1. Introduction. Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be independent and identically distributed (i.i.d.) samples from a distribution with density function f_0. Nonparametric methods are attractive for computing a density estimate for f_0 since they do not impose any parametric assumptions on f_0. Popular such methods include kernel density estimation, adaptive smoothing and neighbor-based techniques. For details we refer to the surveys [Ize91, Tur93, Sco15, Sil18, WJ94, Che17, Was16] and references therein. These techniques require choosing a smoothing parameter, either in the form of a bandwidth for kernel density estimation, or a regularization penalty and clustering parameters for adaptive smoothing and neighbor-based techniques.

An approach gaining popularity in recent years that does not require the choice of a tuning parameter is shape-constrained density estimation. Here

MSC 2010 subject classifications: Primary 62G07, 62H12; secondary 52B55

Keywords and phrases: Nonparametric density estimation; Shape-constrained density estimation; Total positivity; Log-concavity; Supermodularity; $L^\natural$-concavity.

one seeks to maximize the (possibly weighted) log-likelihood function

$$\ell(f) \;=\; \sum_{i=1}^{n} w_i \log f(X_i),$$

subject to a shape constraint $f \in \mathcal{F}$. The weights w_i are fixed, positive, and satisfy $\sum_{i=1}^{n} w_i = 1$. They can be interpreted as the relative importance or confidence among different samples. If the class $\mathcal{F}$ of shapes is unrestricted, i.e., it contains all density functions $f : \mathbb{R}^d \to \mathbb{R}$, then the maximum likelihood estimator (MLE) does not exist. The likelihood function is unbounded, even if the density function is constrained to be unimodal. Thus it is of interest to identify shape constraints that restrict $\mathcal{F}$ sufficiently to make the MLE well-defined with good numerical and theoretical properties, but at the same time keep $\mathcal{F}$ large enough to be relevant for applications.

Different shape constraints that satisfy these requirements are monotonicity (studied in [Gre56] for $d = 1$ and extended in [Pol98] to $d > 1$), convexity (studied in [GJW01] for $d = 1$ and extended in [SW10] to $d > 1$), and, most prominently, log-concavity. The log-concave MLE $\hat{f}_n$ was first introduced by Walther [Wal02] and has been studied in great detail recently. In particular, the MLE exists and is unique with probability one when $n \geq d$ [CSS10, DR09]. Furthermore, Cule, Samworth and Stewart [CSS10] showed that $\log(\hat{f}_n)$ is a piecewise linear concave function; namely, it is a *tent function* with tent poles at the observations $X_i \in \mathbb{R}^d$. Such a tent function induces a regular subdivision of the configuration $X = \{X_1, \ldots, X_n\} \subset \mathbb{R}^d$, and in fact any regular subdivision can arise with positive probability as the MLE for some set of weights [RSU17]. For a recent review of shape-constrained density estimation see Groeneboom and Jongbloed [GJ14].

In this paper, we consider the problem of maximum likelihood estimation under *total positivity*, a special form of positive dependence. A distribution defined by a density function f over $\mathcal{X} = \prod_{i=1}^{d} \mathcal{X}_i$, where each set $\mathcal{X}_i$ is totally ordered, is *multivariate totally positive of order 2* (MTP_2) if

$$f(x)f(y) \quad \leq \quad f(x \wedge y)f(x \vee y) \qquad \text{for all } x, y \in \mathcal{X},$$

where $x \wedge y$ and $x \vee y$ denote the element-wise minimum and maximum. Note that if f is strictly positive, then f is MTP_2 if and only if f is *log-supermodular*, i.e., $\log(f)$ is supermodular. MTP_2 was introduced in [FKG71]. It implies *positive association*, an important property in probability theory and statistical physics, which is usually difficult to verify. In fact, most notions of positive dependence are implied by MTP_2; see for example [CSS05] for a recent overview. The special case of Gaussian MTP_2 distributions was studied by Karlin and Rinott [KR83] and more recently in [SH15, LUZ17] from the perspective of maximum likelihood estimation and optimization.

Unfortunately, maximum likelihood estimation under MTP_2 is ill-defined, since the likelihood function is unbounded. For this reason, in this paper we consider the problem of nonparametric density estimation, where $\mathcal{F}$ consists of all *MTP_2 and log-concave* density functions $f : \mathbb{R}^d \to \mathbb{R}$. A strictly positive function $f : \mathbb{R}^d \to \mathbb{R}$ is log-concave if $\log(f)$ is a concave function, i.e.,

$$f(x)f(y) \;\leq\; f\Big(\frac{x+y}{2}\Big)^2 \quad \text{for all } x, y \in \mathbb{R}^d. \tag{1.1}$$

The class $\mathcal{F}$ of log-concave MTP_2 densities contains many interesting distributions, such as totally positive Gaussian distributions [LUZ17] and in particular Brownian motion tree models, which are used for evolutionary processes [Fel73]. The densities in $\mathcal{F}$ satisfy many desirable properties. For instance, the class $\mathcal{F}$ is closed under marginalization and conditioning [FLS⁺17].

We also consider a subclass of MTP_2 distributions where the logarithm of the density function is *$L^\natural$-concave*. A function $g : \mathbb{R}^d \to \mathbb{R}$ is $L^\natural$-concave if

$$g(x) + g(y) \leq g((x+\alpha\mathbf{1}) \wedge y) + g(x \vee (y - \alpha\mathbf{1})) \quad \text{for all } \alpha \geq 0 \text{ and } x, y \in \mathbb{R}^d,$$

where $\mathbf{1}$ denotes the all ones vector [Mur03, Mur09]. The set $\mathcal{L}$ of log-$L^\natural$-concave (LLC) densities is an appealing subclass of MTP_2 densities. For example, a d-dimensional Gaussian distribution is MTP_2 if and only if its inverse covariance matrix K is an *M-matrix*, i.e. $K_{ij} \leq 0$ for all $i \neq j$, and it is LLC if and only if K is also *diagonally dominant*, i.e.

$$K_{ij} \leq 0 \;\text{ for all }\; i \neq j \quad \text{and} \quad \sum_{j=1}^{d} K_{ij} \geq 0 \;\text{ for all }\; i = 1, \ldots d.$$

In the setting of Gaussian graphical models, if the density is LLC, loopy belief propagation converges and hence marginal distributions can be computed efficiently [WF01, MJW06]. Such functions also appear naturally in economics [Tam04, DKM01, FY03], network flow problems [Nar97], phylogenetics [Zwi16] and combinatorics [JK10, LP07]. In these cases, the i-th coordinate of a vector x represents the unit price of item i or the potential at node i. Since only the relative differences $(x_i - x_j : i, j = 1, \ldots, d)$ matter for the application, $L^\natural$-concavity is the appropriate notion. In the discrete setting, when points have integer coordinates, $L^\natural$-concavity is equivalent to discrete concavity [Mur03, Sections 1 and 2]. For more on $L^\natural$-concavity, see [Mur03, Mur09] and references therein.

Our aim is to study the MLE in two settings: log-concavity combined with MTP_2 and log-concavity combined with LLC. In other words, we study properties of the solutions to the optimization problems

$$\begin{array}{ll} \text{maximize} & w_1 \log f(X_1) + \cdots + w_n \log f(X_n) \\ \text{such that} & f \text{ is a log-concave and } \mathrm{MTP}_2 \text{ density,} \end{array} \tag{1.2}$$

and

$$\begin{array}{ll} \text{maximize} & w_1 \log f(X_1) + \cdots + w_n \log f(X_n) \\ \text{such that} & f \text{ is a log-concave and LLC density.} \end{array} \tag{1.3}$$

Results. Our main result, Theorem 1.1, concerns the minimum number of samples needed for the MLE to exist almost surely. If the support of the underlying density is full-dimensional, that number is three for MTP_2, and two for LLC. This does not depend on the dimension d. We also show that the MLE is unique and consistent in this setting. The proof is given in Appendix A. It builds on work of Royset and Wetts [RW17] which provides a general framework for proving existence and consistency of the MLE in shape constrained density estimation. A direct application of [RW17] to log-concave MTP_2 or LLC densities would require an extra technical assumption (see [RW17, Proposition 4.6]). In comparison, our theorem requires no such assumption, and has the optimal bound on the minimum number of samples.

THEOREM 1.1 (Existence, uniqueness, and consistency of the MLE). *Let $X_1, \ldots, X_n$ be i.i.d samples from a distribution with density f_0 supported on a full-dimensional subset of $\mathbb{R}^d$. The following hold with probability one:*

- *If $n \geq 3$, the MTP_2 log-concave MLE exists and is unique.*
- *If $n \geq 2$, the LLC log-concave MLE exists and is unique.*

In both cases, when the MLE $\hat{f}_n$ exists, it is consistent in the sense that $\hat{f}_n$ converges almost surely to f_0 in the Attouch-Wets metric of [RW17].

The remaining results of our paper concern the shape and computation of the MLE. We describe the support of the MLE and give algorithms for computing it (cf. Section 2). We provide conditions on the samples X which ensure that the MLE under MTP_2 and LLC, respectively, can be computed by solving a convex optimization problem. Under these conditions, which we call *tidy*, the MTP_2 and LLC MLEs behave like the log-concave MLE. They are piecewise-linear and can be computed by solving a finite-dimensional convex optimization problem (cf. Theorem 3.3). For MTP_2, these conditions include $X \subset \mathbb{R}^2$ or $X \subseteq \{0,1\}^d$, and for LLC, they include $X \subset \mathbb{Q}^d$. Since numerical computations are usually performed in $\mathbb{Q}$ by rounding points in $\mathbb{R}$, the LLC MLE can always be computed using the optimization problem in Theorem 3.3. For computing the MLE, we use the conditional gradient method combined with subgradient optimization (cf. Algorithms 2 and 3).

We conjecture that the MLE is always piecewise linear, and can be computed via an explicit finite dimensional convex program (cf. Conjectures 5.5 and 5.6). As steps towards proving these conjectures, we show that a tent function is concave and MTP_2 if and only if it induces a bimonotone subdivision. This result (cf. Theorem 5.1) is the analogue of [Mur03, Theorem 7.45] for the LLC case. Section 5 raises research questions in geometric combinatorics and discrete convex analysis that are of independent interest.

Organization. Our paper is organized as follows. Section 2 characterizes the support of the MTP_2 and LLC MLEs, and it offers algorithms for computing these. Sections 3 develops the convex optimization problem associated to the tidy case, and Section 4 develops algorithms for solving this optimization problem. Section 5 analyzes the estimation problem in the general case. We conclude with a short discussion in Section 6.

2. Support of the MLE. According to [CSS10], the log-concave MLE exists with probability one if there are at least $d+1$ samples, and its support is the convex polytope $\mathrm{conv}(X)$. The support of the log-concave and MTP_2/LLC estimates always contains $\mathrm{conv}(X)$, but is in general larger. In this section we develop the relevant geometric theory to compute it.

We begin with the log-concave MTP_2 case. In the course of the proof of Theorem 1.1 (see Appendix A), we show that if an MLE exists, then its support is the *min-max convex hull* of the samples $X_1, \ldots, X_n$ in $\mathbb{R}^d$.

DEFINITION 2.1. A subset of $\mathbb{R}^d$ is *min-max closed* if, for any two elements u and v in the set, the vectors $u \wedge v$ and $u \vee v$ are also in the set. For a finite set $X \subset \mathbb{R}^d$, its *min-max closure* $\overline{X}$ is the smallest min-max closed set that contains X. Its *min-max convex hull* $\mathrm{MMconv}(X)$ is the smallest min-max closed and convex set containing X. In discrete geometry [FK11], min-max closed convex sets are also known as *distributive polyhedra*.

For a finite set X in $\mathbb{R}^d$, its min-max closure $\overline{X}$ can be computed by adding points iteratively. That is, we set $X^{(0)} = X$, and we define

$$X^{(i)} = \{u \wedge v \,:\, u, v \in X^{(i-1)}\} \cup \{u \vee v \,:\, u, v \in X^{(i-1)}\} \quad \text{for } i \geq 1. \tag{2.1}$$

Since the j-th coordinate of each point in $X^{(i)}$ is among the j-th coordinates of the points in X, equation (2.1) defines an increasing nested sequence of sets which stabilizes in at most $d+1$ steps, and the final set is $\overline{X}$.

For a finite set X, $\mathrm{MMconv}(X)$ is a convex polytope. At first, one might think that when X is finite, $\mathrm{MMconv}(X) = \mathrm{conv}(\overline{X})$. This is true only when the dimension d equals 2. However, when $d \geq 3$, $\mathrm{MMconv}(X)$ is generally larger than $\mathrm{conv}(\overline{X})$. We refer to Example 2.4 and [QT06, Example 17].

The *2-D Projections Theorem* below gives the linear inequality representation of the polytope $\mathrm{MMconv}(X)$. This result was published in the 1970's by Baker and Pickley [BP75] and Topkis [Top76], but they attribute it to George Bergman, who discovered it in the context of universal algebra. This theorem was extended to a characterization of min-max closed polytopes by Queyranne and Tardella [QT06]. For $i, j \in \{1, \ldots, d\}, i \neq j$, let $\pi_{ij} : \mathbb{R}^d \to \mathbb{R}^2$ denote the projection map onto the i-th and j-th coordinate.

Theorem 2.2 (2-D Projections Theorem). *For any finite subset $X \subset \mathbb{R}^d$,*

$$\mathrm{MMconv}(X) \;=\; \bigcap_{i \neq j} \pi_{ij}^{-1}\big(\,\mathrm{conv}\big(\,\pi_{ij}(\overline{X})\,\big)\,\big).$$

Theorem 2.2 reduces the computation of min-max convex hulls to the case $d = 2$, and yields Algorithm 1 for computing $\mathrm{MMconv}(X)$ for a finite set X.

Each edge of a min-max closed polygon in $\mathbb{R}^2$ has non-negative slope, so the line it spans has the form $\{a_1 z_1 + a_2 z_2 = b\}$ where $a_1 a_2 \leq 0$. Extending this to higher dimensions $d \geq 2$, we say that a linear inequality on $\mathbb{R}^d$ is *bimonotone* if it has the form $a_i z_i + a_j z_j \geq b$, where $a_i a_j \leq 0$. Theorem 2.2 implies the following representation for polytopes that are min-max closed.

Corollary 2.3 (Theorem 5 in [QT06]). *A polytope P in $\mathbb{R}^d$ is min-max closed if and only if P is defined by a set of bimonotone linear inequalities.*

Algorithm 1: Computing the min-max closed convex hull

Input : A finite set of points X in $\mathbb{R}^d$.
Output: The polytope $\mathrm{MMconv}(X)$ in $\mathbb{R}^d$.

Compute the finite set $\overline{X}$ by iterating the completion process in (2.1);
for *each pair of distinct indices* i, j **do**
    Compute the convex hull of the planar point configuration $\pi_{ij}(\overline{X})$;
    List the bimonotone inequalities that minimally describe this polygon in $\mathbb{R}^2$;
end
Collect all bimonotone inequalities. Their solution set in $\mathbb{R}^d$ is $\mathrm{MMconv}(X)$;

EXAMPLE 2.4. Set $X = \{(0,0,0),(6,0,0),(6,4,0),(8,4,2)\} \subset \mathbb{R}^3$. This set is a chain in the coordinatewise partial order on $\mathbb{R}^3$, thus $\overline{X} = X$. Hence, $\mathrm{conv}(X) = \mathrm{conv}(\overline{X})$ is a tetrahedron with the four given vertices. This tetrahedron is not min-max closed: $(6,4,0)$ and $(6,3,3/2)$ are both in $\mathrm{conv}(\overline{X})$, but their maximum is the point $(6,4,3/2)$ which is not in $\mathrm{conv}(\overline{X})$. In fact,

$$\mathrm{MMconv}(X) \;=\; \mathrm{conv}\left(X \cup \{\,(6,4,\tfrac{3}{2})\,\}\right).$$

The bimonotone inequalities that describe $\mathrm{MMconv}(X)$ are

$$z \geq 0, \quad y - 2z \geq 0, \quad x - 4z \geq 0, \quad 2x - 3y \geq 0, \quad y - 4 \leq 0, \quad x - z \leq 6.$$

These halfspaces furnish the description of $\mathrm{MMconv}(X)$ in Theorem 2.2. This polytope is a bipyramid, discussed in more detail in Example 5.7. □

We now turn to log-concave LLC distributions. Here the support of the MLE is generally larger than $\mathrm{MMconv}(X)$. In the proof of Theorem 1.1 we show that, if an MLE exists, its domain is given by the $L^\natural$-extension of X.

DEFINITION 2.5. Let $X \subset \mathbb{R}^d$ be a finite set. We say that X is *$L^\natural$-closed* if there exists some $r > 0$ such that

(2.2) $$x, y \in X \text{ implies } (x + \alpha \cdot r\mathbf{1}) \wedge y,\; x \vee (y - \alpha \cdot r\mathbf{1}) \in X \quad \text{for all } \alpha \in \mathbb{Z}_+.$$

We say that X is *$L^\natural$-convex* if

$$x, y \in X \text{ implies } (x + \alpha \cdot \mathbf{1}) \wedge y,\; x \vee (y - \alpha \cdot \mathbf{1}) \in X \quad \text{for all } \alpha \in \mathbb{R}_+.$$

The *discrete $L^\natural$-extension* of X, denoted $\tilde{X}$, is the smallest finite $L^\natural$-closed set containing X. The (continuous) *$L^\natural$-extension* of X, denoted $P(X)$, is the smallest $L^\natural$-convex set containing X.

Unlike the MTP_2 case, a finite set X may not have a discrete $L^\natural$-extension. However, it always has a continuous $L^\natural$-extension $P(X)$. This is a polytope in $\mathbb{R}^d$. The following proposition gives the inequality description for $P(X)$, and it characterizes when $\tilde{X}$ exists. The proof is given in Appendix B.

PROPOSITION 2.6. *Let $X \subset \mathbb{R}^d$ be a finite set. Then*

(2.3) $$P(X) = \{y \in \mathbb{R}^d : y_i - y_j \leq \max_{x \in X}(x_i - x_j), \quad \min_{x \in X} x_i \leq y_i \leq \max_{x \in X} x_i\}.$$

The set X admits a discrete $L^\natural$-extension $\tilde{X}$ if and only if $X - v \subset r \cdot \mathbb{Z}^d$ for some $r > 0$ and $v \in X$. In this case, there is a unique smallest constant $r^* > 0$, independent of the choice of v, such that $X - v \subset r^* \cdot \mathbb{Z}^d$, and

$$\tilde{X} \;=\; v + r^* \cdot \left(P\left(\frac{1}{r^*}(X - v)\right) \cap \mathbb{Z}^d \right). \tag{2.4}$$

If X has a discrete $L^\natural$-extension $\tilde{X}$, then $P(X) = P(\tilde{X}) = \mathrm{conv}(\tilde{X})$.

Any finite set of rational points admits a discrete $L^\natural$-extension; so for practical purposes, $\tilde{X}$ always exists. In that case $P(X) = P(\tilde{X}) = \mathrm{conv}(\tilde{X})$, a key difference compared to the MTP_2 setting. Computation of $\tilde{X}$ and $P(X)$ is immediate from their definitions. We summarize the main results of this section in the following proposition, which is proven in Appendix B.

PROPOSITION 2.7. *Let $X \subset \mathbb{R}^d$ be a finite set of points. If the log-concave MTP_2 MLE exists, then its support a.s. equals the min-max convex hull* $\mathrm{MMconv}(X)$. *If the LLC MLE exists, then its support a.s. equals $P(X)$.*

3. The MLE for tidy configurations. Fix $X = \{X_1, \dots, X_n\} \subset \mathbb{R}^d$. For a vector $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, the *tent function* $h_{X,y} : \mathbb{R}^d \to \mathbb{R}$ is the smallest concave function satisfying $h_{X,y}(X_i) \geq y_i$ for all i. This function is piecewise linear, it is greater than $-\infty$ on $\mathrm{conv}(X)$, and equals $-\infty$ outside $\mathrm{conv}(X)$. In this section, we provide sufficient conditions for the MLE under log-concavity and MTP_2/LLC to equal the exponential of a tent function. We show that this is the case when $d = 2$ or when $X \subseteq \prod_{i=1}^d \{a_i, b_i\}$ for some $a_i \leq b_i$, $i = 1, \dots, d$ under MTP_2, and when $X \subset \mathbb{Q}^d$ under LLC. If we know that the MLE equals the exponential of a tent function then we only need to find the heights $y \in \mathbb{R}^n$. This is a finite-dimensional problem.

DEFINITION 3.1. Given X, a vector of heights $y \in \mathbb{R}^n$ is *supermodular* if

$$y_i + y_j \leq y_k + y_\ell \quad \text{whenever} \quad X_k = X_i \wedge X_j \text{ and } X_\ell = X_i \vee X_j.$$

The vector $y \in \mathbb{R}^n$ is $L^\natural$*-concave* if, for all $\alpha > 0$,

$$y_i + y_j \leq y_k + y_\ell \;\text{ whenever }\; X_k = (X_i + \alpha\mathbf{1}) \wedge X_j \text{ and } X_\ell = X_i \vee (X_j - \alpha\mathbf{1}).$$

The vector $y \in \mathbb{R}^n$ is *tight* or *concave* if $h_{X,y}(X_i) = y_i$ for all $i \in \{1, \dots, n\}$.

In the following, we show that the MLE is a tent function for special configurations X for which supermodularity/$L^\natural$-concavity of the heights y already implies that the tent function $h_{X,y}$ is supermodular/$L^\natural$-concave.

DEFINITION 3.2. Let $X = \{X_1, X_2, \ldots, X_n\}$ be a subset of $\mathbb{R}^d$. We say that X is *MTP_2-tidy* if, for any set of tight supermodular heights $y \in \mathbb{R}^n$, the tent function $h_{X,y}$ is supermodular. We say that X is *LLC-tidy* if, for any set of tight $L^\natural$-concave heights $y \in \mathbb{R}^n$, the tent function $h_{X,y}$ is $L^\natural$-concave.

Note that MTP_2-tidy/LLC-tidy implies that X is min-max closed/equals its discrete $L^\natural$-extension. We now state our main result of this section.

THEOREM 3.3. *If X is MTP_2-tidy/LLC-tidy, then the optimal solution to the MLE problem is the exponential of a tent function on X. The tent pole heights y can be found by solving the finite-dimensional convex program*

$$\begin{aligned} &\text{(3.1)} \qquad \textit{minimize} \quad -w\cdot y + \int_{\mathbb{R}^d} exp(h_{X,y}(z))dz \\ &\qquad\qquad\quad \textit{s.t.} \qquad y \in \mathcal{S}, \end{aligned}$$

where $\mathcal{S}$ is the set of tight supermodular heights y on X for the MTP_2 case, and it is the set of tight $L^\natural$-concave heights on X for the LLC case.

PROOF. We here prove the MTP_2 case. The LLC case is the same, with 'tight supermodular' replaced by 'tight $L^\natural$-concave'. Suppose X is MTP_2-tidy, and let $\mathcal{S}$ be the set of tight supermodular heights on X. Note that $\mathcal{S}$ is convex. Using Lagrange multipliers, one can see that (3.1) is equivalent to

$$\begin{aligned} &\text{(3.2)} \qquad \text{minimize} \quad -w\cdot y \\ &\qquad\qquad\quad \text{s.t.} \qquad y \in \mathcal{S}, \quad \int_{\mathbb{R}^d} exp(h_{X,y}(z))dz = 1. \end{aligned}$$

Any optimizer of (3.2) is a feasible solution to (1.2), because (3.2) minimizes the same objective function over a subset of the MTP_2 densities on X. Let f be any feasible solution to (1.2). Define $y'_i = \log(f(X_i))$ for $i = 1, 2, \ldots, n$.

We shall exhibit a feasible solution y of (3.2) whose objective function is at least as large as that of f, that is, $-w\cdot y \le -w\cdot y'$. This will prove the desired statement. We abbreviate $g = \exp(h_{X,y'})$. As $h_{X,y}$ is defined to be the smallest concave function with the given values on X, we have $g \le f$ pointwise, and hence $\int_{\mathbb{R}^d} g(x)dx \le 1$. Let $c = 1/\int_{\mathbb{R}^d} g(x)dx$ and define y by setting $y_i := y'_i + \log c$ for all i. Since X is tidy and f is MTP_2, we have $y \in \mathcal{S}$. Furthermore, $\int \exp(h_{X,y}(z))\,dz = c\int \exp(h_{X,y'}(z))\,dz = 1$. Therefore, y is a feasible solution of (3.2), and $-w\cdot y \le -w\cdot y'$, as desired. □

It is hence of interest to characterize the configurations that are MTP_2-tidy or LLC-tidy. In the MTP_2 case, a necessary condition is that X is min-max closed, that is, $X = \overline{X}$. We show that in dimension two or when X is binary, this is also sufficient. The proof is provided in Appendix C.

THEOREM 3.4. *Let $X = \overline{X}$ be a finite set of points in $\mathbb{R}^d$ that is min-max closed. If $d = 2$ or if $X \subseteq \prod_{i=1}^{d}\{a_i, b_i\}$ where $a_i \leq b_i$ for all $i = 1, \ldots, d$, then X is MTP$_2$-tidy.*

Outside the cases covered by Theorem 3.4, there is an abundance of min-max closed configurations X that fail to be MTP$_2$-tidy. This happens even if $\operatorname{conv}(X)$ is min-max closed, as in Example 5.3. We conjecture that two-dimensional and binary are essentially the only MTP$_2$-tidy configurations.

CONJECTURE 3.5. *A min-max closed configuration X of points in $\mathbb{R}^d$ for which $\operatorname{conv}(X) = MMconv(X)$ is MTP$_2$-tidy if and only if $d = 2$ or $X \subseteq \prod_{i=1}^{d}\{a_i, b_i\}$, where $a_i \leq b_i \in \mathbb{R}$ for $i = 1, \ldots, d$.*

In other words, if three points X_1, X_2, X_3 in X have distinct i-th coordinates, for some i, then we conjecture that the configuration X is not tidy. The situation is much better for LLC-tidiness. Similar to the MTP$_2$ case, a necessary condition for X to be LLC-tidy is that it admits a discrete $L^\natural$-extension. The following states that this condition is also sufficient.

THEOREM 3.6 (Theorem 7.26 in [Mur03]). *Let X be a finite configuration in $\mathbb{R}^d$ that admits a discrete $L^\natural$-extension $\tilde{X}$. Then $\tilde{X}$ is LLC-tidy. In particular, if $X \subset \mathbb{Q}^d$, then its discrete $L^\natural$-extension is LLC-tidy.*

For computational purposes, numbers in $\mathbb{R}$ are usually rounded to numbers in $\mathbb{Q}$. Theorem 3.6 hence means that, for practical purposes, any given sample X can be extended to an LLC-tidy configuration. Theorem 3.3 implies that the LLC MLE is the exponential of a tent function which can be computed by solving a finite-dimensional convex optimization problem.

Note that MTP$_2$-tidiness does not imply LLC-tidiness: by Theorem 3.4, $X \subset \mathbb{R}^2$ is MTP$_2$-tidy, but not every finite set of points in $\mathbb{R}^2$ can be extended to a finite LLC-tidy configuration by Proposition 2.6. Neither does being LLC-tidy imply MTP$_2$-tidy; see Example 5.4. One instance that is both MTP$_2$-tidy and LLC-tidy is the cube $X = \{0,1\}^d$. Here, the tight supermodular function $h_{X,y}$ is known as the *Lovász extension* of y [Lov83].

4. Algorithms for computing the MLE for tidy configurations. As shown in Theorem 3.3, for tidy configurations, the optimization problem (3.1) finds the MLE in both the MTP$_2$ and LLC cases. Note that this problem has a convex but not everywhere differentiable objective function and a convex polyhedral feasible set $\mathcal{S}$ consisting of all tight supermodular/$L^\natural$-concave heights. We implemented the *Frank-Wolfe* (also known as *conditional gradient*) method for solving this problem [NW06], using subgradients instead of gradients at points of non-differentiability. The Frank-Wolfe algorithm is nevertheless guaranteed to find the global optimum [Whi93].

Algorithm 2 computes the MTP$_2$ MLE for configurations X in $\mathbb{R}^2$. It easily extends to other tidy configurations. Its first step is Algorithm 4 for computing the min-max closure $\overline{X}$ of the set X. The algorithm uses an alternative characterization of MMconv(X) (cf. Lemma D.1 in Appendix D). As a speedup, our algorithm iterates through the points (a_1, b_2) in $\mathbb{R}^2$, for all pairs $a, b \in X$, and checks whether they lie in MMconv(X). The resulting set X' gives the tent pole locations. Then the algorithm computes the set $\mathcal{S}$ of tight supermodular heights, as described in Algorithm 6, and uses the conditional gradient method (Algorithm 8) to find the optimal solution.

The LLC analogue is given in Algorithm 3. It computes the LLC MLE for rational configurations $X \subset \mathbb{Q}^d$. The $L^\natural$-extension of a set X is computed using Algorithm 5. Then Algorithm 7 computes $\mathcal{S}$ based on an alternative characterization of $L^\natural$-concavity given in [Mur03, Proposition 7.5]. The proof of correctness is given in Appendix D. Finally, the Frank-Wolfe method (Algorithm 8) solves the optimization problem using subgradient optimization.

We now illustrate how Algorithms 2 and 3 perform on Gaussian samples.

Example 4.1. Let X consist of 55 i.i.d. samples from a standard Gaussian distribution in $\mathbb{R}^2$. In Figure 1, we show the corresponding log-concave density estimator on the left, the log-concave MTP$_2$ density estimator in the middle, and the log-concave LLC density estimator on the right. The

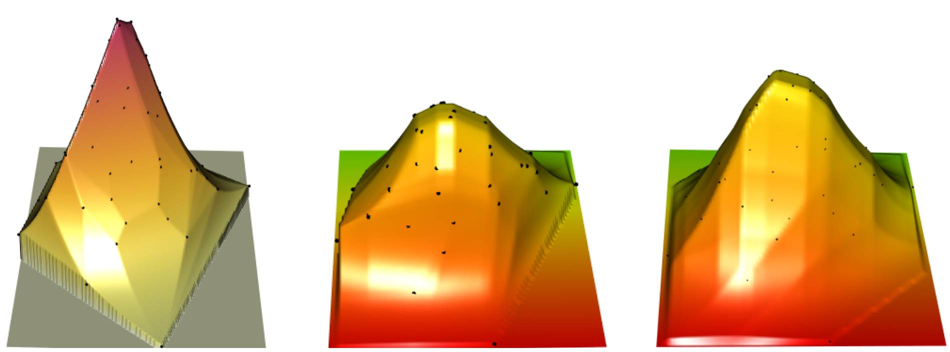


Fig 1. *Density estimates for 55 samples from a standard Gaussian distribution in* $\mathbb{R}^2$*; Log-concave using [CGS09] (left), log-concave MTP*$_2$ *(middle), log-concave LLC (right).*

Algorithm 2: Computing the MTP_2 MLE in $\mathbb{R}^2$

Input : n samples $X \subset \mathbb{R}^2$ with weights $w \in \mathbb{R}^n$.
Output: the optimal heights $\overline{y} \in \mathbb{R}^N$, $N = |\overline{X}|$, corresponding to the points in $\overline{X}$.

Compute the min-max closure $\overline{X}$ of X using Algorithm 4;
Set the new weights $\overline{w}$ equal to the old weights w padded with zeros for each of the extra points in $\overline{X}$;
Compute the set $\mathcal{S}$ of inequalities for all supermodular heights $\overline{y}$ using Algorithm 6;
Compute $\overline{y}$ as the optimum for (3.1) using Algorithm 8.

Algorithm 3: Computing the LLC MLE for $X \subset \mathbb{Q}^d$

Input : n samples $X \subset \mathbb{Q}^d$ with weights $w \in \mathbb{R}^n$.
Output: the optimal heights $\tilde{y} \in \mathbb{R}^N$, $N = |\tilde{X}|$, corresponding to the points in $\tilde{X}$.

Compute the discrete $L^\natural$-extension $\tilde{X}$ of X using Algorithm 5
Set the new weights $\tilde{w}$ equal to the old weights w padded with zeros for each of the extra points in $\tilde{X}$;
Compute the set $\mathcal{S}$ of inequalities for all supermodular heights $\tilde{y}$ using Algorithm 7;
Compute $\tilde{y}$ as the optimum for (3.1) using Algorithm 8.

Algorithm 4: Computing the min-max closure of a finite set $X \subset \mathbb{R}^2$

Input : a finite set of points $X \subset \mathbb{R}^2$;
Output: the min-max closure $\overline{X}$ of X.

```
Create an n1 × n2 index table T filled with zeros corresponding to grid(X).
For each X_i ∈ X, find its place in T and set the corresponding value of T to i.
Set X̄ = X;
Let C = conv(X ∪ {min(X), max(X)}) and counter = #X + 1;
For i from 1 to n1 do:
    For j from 1 to n2 do:
        If p[i, j] ∈ C then set
            T[i, j] = counter; counter = counter +1; Add p[i, j] to X̄;
Output X̄.
```

log-concave density estimator was computed using the package `LogConcDEAD` described in [CGS09]. The log-concave MTP_2 density estimator was computed using Algorithm 2 for solving the optimization problem (3.1). The min-max closure $\overline{X}$ was computed using Algorithm 4 and consisted of the original 55 samples plus 2691 additional points. The LLC density estimator was computed using Algorithm 3 with input X' obtained by rounding each point in X up to the first decimal place. The discrete $L^\natural$-extension of X' was computed using Algorithm 5 and consisted of the original 55 samples

Algorithm 5: The discrete $L^\natural$-extension of a finite set $X \subset \mathbb{Q}^d$

Input : a finite set of points $X \subset Q^d$;
Output: the discrete $L^\natural$-extension $\tilde{X}$ of X.

Let $m \in \mathbb{N}$ be the smallest integer such that $m \cdot X \subset \mathbb{Z}^d$;
Compute $P(m \cdot X)$ via (2.3);
Set $X' = P(m \cdot X) \cap \mathbb{Z}^d$;
Output $\tilde{X} = \{\frac{1}{m}x' : x' \in X'\}$.

Algorithm 6: Computing the inequalities $\mathcal{S}_{\text{supermodular}}$ in $\mathbb{R}^2$

Input : the index table T generated for X in Algorithm 4
Output: the set of inequalities defining the cone of supermodular heights $\mathcal{S}$

Set $\mathcal{S}$ to be an empty set of inequalities
For i from 1 to $n_1 - 1$:
  For j from 1 to $n_2 - 1$:
    If $T[i,j] \neq 0, T[i+1,j] \neq 0, T[i,j+1] \neq 0, T[i+1,j+1] \neq 0$, then
      add the inequality $y_{T[i,j]} + y_{T[i+1,j+1]} - y_{T[i+1,j]} - y_{T[i,j+1]} \geq 0$ to $\mathcal{S}$;
Output $\mathcal{S}$.

Algorithm 7: Computing the inequalities $\mathcal{S}_{L^\natural\text{-concave}}$ in $\mathbb{Z}^d$

Input : Finitely many points $X \subset \mathbb{Z}^d$ that are LLC-tidy
Output: the set of inequalities defining the cone of $L^\natural$-concave heights $\mathcal{S}$

Let $e_i = i$-th coordinate vector, $e_0 = -\sum_{i=1}^d e_i$, $\mathcal{S}$ empty set of inequalities;
For each $x \in X$:
  For each $i, j = 0, 1, \ldots, d$, $i < j$:
    If $x + e_i + e_j, x + e_i, x + e_j \in X$ then
      add the inequality $y(x) + y(x + e_i + e_j) - y(x + e_i) - y(x + e_j) \geq 0$ to $\mathcal{S}$
Output $\mathcal{S}$.

Algorithm 8: Frank-Wolfe update

Input : objective function $f : \mathbb{R}^n \to \mathbb{R}$, function $g : \mathbb{R}^n \to \mathbb{R}^n$ producing a subgradient of f, feasible set $\mathcal{S} \subseteq \mathbb{R}^d$.
Output: solution $y^* = \text{argmin}_{y \in \mathcal{S}} f(y)$.

Initialize $y^{(0)}$ in the interior of $\mathcal{S}$
For k from 1 to N do:
  Set $s^{k-1} = \text{argmin}_{s \in \mathcal{S}} g(y^{(k-1)})^T s$ via a linear program;
  Set $y^{(k)} = (1 - \gamma_k) y^{(k-1)} + \gamma_k s^{(k-1)}$, where $\gamma_k = \frac{2}{k+1}$;
Output $y^{(N)}$.

plus 2496 additional points. An implementation is provided in

`http://github.com/erobeva/MTP_2_density_estimation` .

This example indicates the power of MTP_2 and LLC constraints: with only 55 samples, the log-concave MTP_2 and LLC estimators look close to a Gaussian distribution, while the log-concave density estimator is still quite rough. This is reflected in the ℓ_2, Hellinger and ℓ_∞ distances between the estimates and the true distribution, shown in the following table.

	log-concave	log-concave MTP_2	log-concave LLC
ℓ_2-loss	0.0115093	0.00539397	0.00627162
Hellinger loss	0.1210729	0.09313234	0.06755705
ℓ_∞-loss	0.0981327	0.03811975	0.04853513

5. The MLE in the general case. We saw in Theorem 3.3 that in the tidy case, the MLE always equals the exponential of a tent function and can be found by solving a finite-dimensional convex optimization problem. In this section we explore the non-tidy case. A naïve approach is to start by computing the log-concave MLE. Following [CSS10, RSU17], this is always the exponential of a tent function. If that MLE is MTP_2/LLC, then it is also the MLE under log-concavity plus MTP_2/LLC. It is therefore of interest to characterize supermodular/$L^\natural$-concave tent functions.

5.1. *Characterization of supermodular/$L^\natural$-concave tent functions.* Without loss of generality we assume that our configuration X satisfies $X = \overline{X}$ and $\mathrm{MMconv}(X) = \mathrm{conv}(X)$. If this is not the case, Algorithm 1 can be used to add points to X so that those two conditions are satisfied. Every piecewise linear function $h_{X,y}$ induces a regular polyhedral subdivision of the point configuration X. A *polyhedral subdivision of a point configuration* X is a polyhedral complex $\Delta \subseteq \mathbb{R}^d$ whose set of vertices is a subset of X and the union of whose cells, denoted $|\Delta|$, equals $\mathrm{conv}(X)$. The cells in Δ are themselves convex polyhedra of various dimensions, and any two cells intersect in a polyhedron that is a face of each. If all cells in Δ are simplices then Δ is called a *triangulation*. A polyhedral subdivision Δ of X is *regular* if there exists a set of heights $y \in \mathbb{R}^n$ such that the polyhedra in Δ are the regions of linearity of the piecewise linear function $h_{X,y}$. We refer to the textbook [LRS10] for all relevant basics on subdivisions and triangulations.

We are now ready for the main result in this section which is the following characterization, in terms of polyhedral subdivisions, of the functions $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ that are concave, supermodular, and piecewise linear.

THEOREM 5.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a piecewise linear concave function, and Δ the subdivision of* $\mathrm{supp}(f)$ *whose cells are the regions of linearity of f. Then, f is supermodular if and only if each of the cells of Δ is min-max closed, i.e. it is defined by bimonotone linear inequalities.*

The proof of Theorem 5.1 is given in Appendix E. The analogous theorem for $L^\natural$-concave tent functions is the following result of Murota.

THEOREM 5.2 (Theorem 7.45, [Mur03]). *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a piecewise linear concave function, and Δ the induced polyhedral subdivision of* $\mathrm{supp}(f)$*. Then, f is $L^\natural$-concave if and only if each cell of Δ is $L^\natural$-concave.*

Theorem 5.1 can be used to construct simple examples of configurations that are not MTP_2-tidy.

EXAMPLE 5.3. Let $X = \{(0,0,0),(6,0,0),(6,4,0),(6,4,\frac{3}{2}),(8,4,2)\}$ in $\mathbb{R}^3$. This is a chain in $(\mathbb{Z}^3, \leq)$. Thus, X is min-max closed, and any height vector $y \in \mathbb{R}^5$ is supermodular. The configuration X has precisely two triangulations. Neither of them is bimonotone. Hence X is not MTP_2-tidy.

Next, we study the set $\mathcal{S} \subset \mathbb{R}^n$ of height vectors y which induce a supermodular tent function. In the tidy setting, $\mathcal{S}$ is a convex set, since it is the set of all tight supermodular heights. However, in the non-tidy setting, $\mathcal{S}$ is often not convex, as illustrated in the following example.

EXAMPLE 5.4 (Two cubes in $\mathbb{R}^3$). Consider the $n = 12$ points in $\mathbb{R}^3$ given by

$$X = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{pmatrix}.$$

Thus X consists of the lattice points in a $1 \times 1 \times 2$ box, obtained by adjoining two 3-cubes. Consider the height vectors $y = (5,6,1,1,4,1,1,4,1,1,6,5)$ and $y' = (11,9,1,1,0,0,10,12,\ 16,11,21,27)$. These induce two distinct regular triangulations Δ and Δ' of X. In terms of column labels for X, they are

$$\begin{aligned} \Delta = \big\{ & \{1,2,5,11\},\{1,2,8,11\},\{1,4,5,11\},\{1,4,10,11\},\{1,7,8,11\},\{1,7,10,11\}, \\ & \{2,3,6,12\},\{2,3,9,12\},\{2,5,6,12\},\{2,5,11,12\},\{2,8,9,12\},\{2,8,11,12\}\big\}, \end{aligned}$$

$$\begin{aligned} \Delta' = \big\{ & \{1,2,6,12\},\{1,2,9,12\},\{1,4,6,12\},\{1,4,10,11\},\{1,4,11,12\}, \\ & \{1,7,9,12\},\{1,7,10,11\},\{1,7,11,12\},\{2,3,6,12\},\{2,3,9,12\}\big\}. \end{aligned}$$

A computation reveals that a subdivision of X is bimonotone if and only if each of its walls is spanned by points indexed by subsets in

$$\{1,4,7,10\},\{1,4,8,11\},\{1,4,9,12\},\{1,5,7,11\},\{1,6,7,12\},\{2,5,8,11\},\\ \{2,5,9,12\},\{2,6,8,12\},\{3,6,9,12\},\{1,2,3,10,11,12\},\{1,2,3,4,5,6\},\\ \{1,2,3,7,8,9\},\{4,5,6,10,11,12\},\{7,8,9,10,11,12\}.$$

Hence, both Δ and Δ' are bimonotone, so y and y' induce supermodular tent functions. However, the convex combination $y'' = \frac{5}{12}y + \frac{7}{12}y'$ induces

$$\Delta'' = \big\{\{\mathbf{1},2,\mathbf{5},\mathbf{12}\},\{1,2,9,12\},\{\mathbf{1},4,\mathbf{5},\mathbf{12}\},\{1,4,10,11\},\{1,4,11,12\},\{1,7,8,12\},\\ \{1,7,10,11\},\{1,7,11,12\},\{1,8,9,12\},\{2,3,6,12\},\{2,3,9,12\},\{2,5,6,12\}\big\}.$$

The regular triangulation Δ'' is not bimonotone because the wall given by $\{1,5,12\}$ is not bimonotone. Therefore, the set of heights y for which $h_{X,y}$ is supermodular is not convex. This in particular implies that the configuration X is not MTP_2-tidy. Further, X is also an example of a configuration that is LLC-tidy (as can be checked from Definition 2.5) but not MTP_2-tidy. □

5.2. *Computing the MTP_2 MLE for non-tidy configurations.* Since rational X are always LLC-tidy, we focus on configurations that are not MTP_2-tidy. In this case, one may hope to find a finite set $X' \supseteq X$ such that X' is MTP_2-tidy. However, Conjecture 3.5 suggests that this is difficult. Algorithm 9 provides a finite set $X' \supseteq X$ together with a particular candidate bimonotone subdivision Δ' on X'. Though X' is not necessarily tidy, we conjecture that the optimization problem (3.1) solved over tent functions on X' which induce a coarsening of the subdivision Δ' yields the solution to the MTP_2 MLE problem. The intuition for Algorithm 9 is to add a minimal number of points to Δ so that the new subdivision Δ' is bimonotone.

Algorithm 9: Computing a candidate MLE bimonotone subdivision

Input : n samples $X \subset \mathbb{R}^d$ with $X = \overline{X}$ and weights $w \in \mathbb{R}^n$.
Output: A finite set $X' \supseteq X$ and a bimonotone subdivision Δ' on X'.

1. Compute the logarithm of the log-concave MLE (a tent function). Let Δ be the subdivision it induces;
2. If Δ is a bimonotone subdivision, return $X' = X$, $\Delta' = \Delta$;
3. If Δ is not bimonotone, compute the hyperplanes spanned by each of the bimonotone facets of Δ, and intersect $\mathrm{conv}(X)$ with those hyperplanes; call this new subdivision Δ' and its vertices X';
4. Output (X', Δ').

Conjecture 5.5. *The MTP_2 log-concave MLE is a piecewise linear function whose subdivision is Δ' or any subdivision refined by Δ'.*

This conjecture can be reformulated as follows.

CONJECTURE 5.6. *Let X' be the output of Algorithm 9. Let $\hat{y}$ be the solution to the optimization problem*

$$\text{(5.1)} \qquad \textit{minimize} \quad -w \cdot y + \int_{\mathbb{R}^d} exp(h_{X',y}(z))dz \quad \textit{s.t.} \quad y \in \mathcal{S},$$

where $w \in \mathbb{R}^{|X'|}$ assigns the original weights to the points in X, and weight 0 to all points in $X' \backslash X$. Then, the tent function $h_{X',\hat{y}}$ is supermodular.

We now justify why Conjectures 5.5 and 5.6 are equivalent. The former clearly implies the latter. Now, suppose that Conjecture 5.6 is true. Let $\exp(f)$ be the MTP_2 log-concave MLE, and let $\hat{y}$ be the optimal heights in (5.1). Let $y = f(X')$ be the heights induced by f. Let $c \geq 1$ be the constant such that $\int c \cdot h_{X',y} = 1$. Then, $\ell(\exp(f)) \leq \ell(\exp(h_{X',y+\log(c)})) \leq \ell(\exp(h_{X',\hat{y}}))$, where equality holds if and only if $f = h_{X',\hat{y}}$. Since $\exp(f)$ is the MLE, then equality should hold, and therefore, f is a tent function.

We now give some intuition for Conjecture 5.6. By Theorem 5.1, if $h_{X',\hat{y}}$ is not supermodular, then it induces a non-bimonotone subdivision. Consider a non-bimonotone wall in that subdivision. If it is spanned only by points in X, then it should have been taken care of by the inequalities for $\hat{y} \in \mathcal{S}$. If it involves some points in $X' \backslash X$, then, we should be able to lower the heights at those vertices (without changing the likelihood), and get a subdivision with that wall removed. But now we would have to shift up the whole tent function so that the integral is still equal to 1, thereby getting a higher likelihood. In the following, we verify Conjecture 5.5 for our running example.

EXAMPLE 5.7. As in Example 5.3, fix $d = 3, n = 5$ and let X consist of

$$X_1 = (0,0,0),\ X_2 = (6,0,0),\ X_3 = (6,4,0),\ X_4 = (8,4,2),\ X_5 = (6,4,1.5).$$

We apply Algorithm 9 to $w = \frac{1}{28}(15,1,1,1,10)$. The output is $X' = X \cup \{X_6, X_7\}$, where $X_6 = (6,3,1.5)$, $X_7 = (7.5,4,1.5)$. The solution to (5.1) is

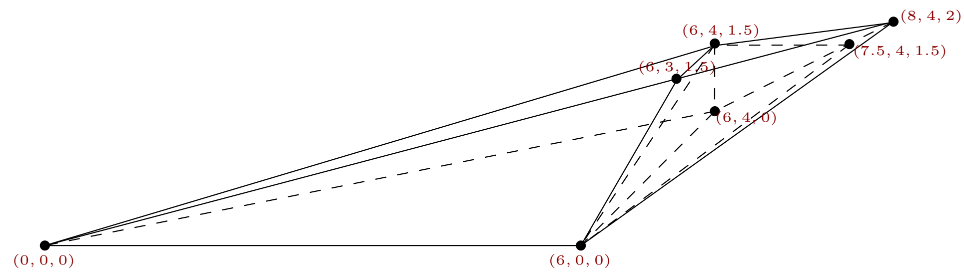
$$y = (2.95,\ -22.05,\ -14.08,\ -5.16,\ 0.40,\ -2.52,\ -6.47) \in \mathbb{R}^7.$$

Let $\exp(f)$ be the log-concave MTP_2 MLE and let $\exp(\phi)$ be the log-concave MLE. Using the software `LogConcDead` [CGS09], we can compute ϕ and show that it induces the following triangulation of $\mathrm{conv}(X)$ with three tetrahedra:

$$\text{(5.2)} \qquad \{X_1, X_2, X_3, X_5\}, \{X_2, X_5, X_3, X_4\}, \{X_2, X_5, X_4, X_1\}$$

The interior faces of this triangulation are

$$\{X_2, X_5, X_1\}, \{X_2, X_5, X_3\} : \text{bimonotone}, \qquad \{X_2, X_5, X_4\} : \text{not bimonotone}.$$



Consider the hyperplanes spanned by the two interior bimonotone faces:

$$\text{affine span}\{X_2, X_5, X_1\} \text{ and } \text{affine span}\{X_2, X_5, X_3\}.$$

Intersecting these with the boundary of $\operatorname{conv}(X)$ yields $X' = X \cup \{X_6, X_7\}$. Now, note that the following inequalities hold for f:

$$\frac{1}{2}f(X_6) + \frac{1}{2}f(X_7) \geq \frac{1}{2}f(X_5) + \frac{1}{2}f(7.5, 3, 1.5) \geq \frac{1}{2}f(X_5) + \frac{3}{8}f(X_4) + \frac{1}{8}f(X_2),$$

where the first inequality follows by supermodularity, and the second follows by concavity of f. Suppose that at least one of these inequalities is strict. Note that for ϕ we have the following reverse inequality:

$$\begin{aligned} \frac{1}{2}\phi(X_6) + \frac{1}{2}\phi(X_7) &= \frac{3}{8}\phi(X_4) + \frac{1}{8}\phi(X_1) + \frac{3}{8}\phi(X_4) + \frac{1}{8}\phi(X_3) \\ &< \frac{1}{2}\phi(X_5) + \frac{3}{8}\phi(X_4) + \frac{1}{8}\phi(X_2), \end{aligned}$$

where the inequality follows because ϕ induces the triangulation (5.2). Hence, there exists $\alpha \in (0,1)$ such that the function $g_\alpha = \alpha f + (1-\alpha)\phi$ satisfies

$$\frac{1}{2}g_\alpha(X_6) + \frac{1}{2}g_\alpha(X_7) = \frac{1}{2}g_\alpha(X_5) + \frac{3}{8}g_\alpha(X_4) + \frac{1}{8}g_\alpha(X_2). \tag{5.3}$$

Moreover, g_α gives a higher likelihood than f, since $\ell(\exp(g_\alpha)) = \alpha\ell(\exp(f)) + (1-\alpha)\ell(\exp(\phi))$, and the piecewise linear function $h_{X',y'}$ on X', induced by the heights of g_α: $y' = g_\alpha(X')$ gives a higher likelihood than g_α. In our experiments, we optimized the likelihood over heights that satisfy the equality (5.3), and the best such tent function induces the bimonotone subdivision

$$\{X_1X_2X_3X_5,\ X_1X_2X_5X_6,\ X_2X_3X_5X_7,\ X_2X_5X_4X_6X_7\}.$$

Thus, we have found a supermodular tent function with a higher likelihood than f, which is a contradiction. As a consequence, equality (5.3) has to hold for f, and f has to be a tent function, namely the one computed above. □

6. Discussion. In this paper, we studied the MLE for nonparametric density estimation under MTP_2 and LLC, two shape constraints that imply strong forms of positive dependence. These shape constraints are of interest for high-dimensional applications since the MLE exists already for 3 samples (under MTP_2) and 2 samples (under LLC), irrespective of the number of variables. We proved that the MTP_2 MLE is a tent function in the two-dimensional or binary setting. We conjectured that this is true in general and we provided an algorithm for computing a candidate MLE. We proved that the MLE under LLC is a tent function when the samples lie in $\mathbb{Q}^d$. It can be computed by solving a finite-dimensional convex optimization problem. Since computations are usually performed in $\mathbb{Q}$ by rounding points in $\mathbb{R}$, in practice, the LLC MLE is always a tent function. Since LLC distributions form a subclass of MTP_2 distributions, the LLC MLE can always be used as an MTP_2 estimator, although it might not be the MTP_2 MLE.

We also provided conditional gradient methods for solving the finite-dimensional convex optimization problem to compute the MLE under LLC and MTP_2. Simulations with 55 i.i.d. samples from a standard Gaussian distribution in $\mathbb{R}^2$ indicate fast convergence of these estimators to the true density. A question for future research is to determine these rates as compared to the log-concave MLE. Since MTP_2 is implied by many models including latent tree models in phylogenetics or single factor analysis models in psychology [FLS+17], our work suggests MTP_2 as a strong enough shape constraint to obtain accurate density estimates with relatively few samples, but a large enough class to be of interest for various applications.

Acknowledgments. We thank Steffen Lauritzen and Richard Samworth for helpful discussions. Elina Robeva was supported by the NSF (DMS-170-3821). Bernd Sturmfels was partially supported by the NSF (DMS-1419018) and the Einstein Foundation Berlin. Caroline Uhler was partially supported by the NSF (DMS-1651995), ONR (N00014-17-1-2147), a Sloan Fellowship, and a Guest Professorship at the Department of Biosystems Science and Engineering at ETH Zürich. Ngoc Tran was partially supported by a Bonn Research Fellowship at the Hausdorff Center for Mathematics, Germany.

REFERENCES

[BP75] K. A. Baker and A. F. Pixley. Polynomial interpolation and the Chinese remainder theorem for algebraic systems. *Mathematische Zeitschrift*, 143:165–174, 1975.

[CGS09] M. Cule, R. B. Gramacy, and R. Samworth. LogConcDEAD: an R package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software*, 29, 2009.

[Che17] Y. Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187, 2017.

[CSS05] A. Colangelo, M. Scarsini, and M. Shaked. Some notions of multivariate positive dependence. *Insurance: Mathematics and Economics*, 37(1):13–26, 2005.

[CSS10] M. Cule, R. Samworth, and M. Stewart. Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society, Series B*, 72:545–607, 2010.

[DKM01] V. Danilov, G. Koshevoy, and K. Murota. Discrete convexity and equilibria in economies with indivisible goods and money. *Mathematical Social Sciences*, 41(3):251–273, 2001.

[DR09] L. Dümbgen and K. Rufibach. Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, 15:40–68, 2009.

[Fel73] J. Felsenstein. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, 25(5):471–492, 1973.

[FK11] S. Felsner and K. Knauer. Distributive lattices, polyhedra, and generalized flows. *European Journal of Combinatorics*, 32(1):45–59, 2011.

[FKG71] C. Fortuin, P. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

[FLS+17] S. Fallat, S. Lauritzen, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik. Total positivity in Markov structures. *Annals of Statistics*, 45:1152–1184, 2017.

[FY03] S. Fujishige and Z. Yang. A note on Kelso and Crawford's gross substitutes condition. *Mathematics of Operations Research*, 28(3):463–469, 2003.

[GJ14] P. Groeneboom and G. Jongbloed. *Nonparametric Estimation under Shape Constraints*. Cambridge University Press, Cambridge, 2014.

[GJW01] P. Groeneboom, G. Jongbloed, and J. A. Wellner. Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics*, 29:1653–1698, 2001.

[Gre56] U. Grenander. On the theory of mortality measurement II. *Skandinavisk Aktuarietidskrift*, 39:125–153, 1956.

[Ize91] A. J. Izenman. Review papers: Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, 1991.

[JK10] M. Joswig and K. Kulas. Tropical and ordinary convexity combined. *Advances in Geometry*, 10(2):333–352, 2010.

[KR83] S. Karlin and Y. Rinott. M-matrices as covariance matrices of multinormal distributions. *Linear Algebra and its Applications*, 52:419–438, 1983.

[Lov83] L. Lovász. Submodular functions and convexity. In *Mathematical Programming: The State of the Art*, pages 235–257. Springer, 1983.

[LP07] T. Lam and A. Postnikov. Alcoved polytopes, I. *Discrete & Computational Geometry*, 38(3):453–478, 2007.

[LRS10] J. De Loera, J. Rambau, and F. Santos. *Triangulations. Structures for Algorithms and Applications*, volume 25. Springer-Verlag, Berlin, 2010.

[LUZ17] S. Lauritzen, C. Uhler, and P. Zwiernik. Maximum likelihood estimation in gaussian models under total positivity. *arXiv:1702.04031*, 2017.

[MJW06] D. M. Malioutov, J. K. Johnson, and A. S. Willsky. Walk-sums and belief propagation in Gaussian graphical models. *Journal of Machine Learning Research*, 7:2031–2064, 2006.

[Mur03] K. Murota. *Discrete Convex Analysis*. SIAM, 2003.

[Mur09] K. Murota. Recent developments in discrete convex analysis. In *Research Trends in Combinatorial Optimization*, pages 219–260. Springer, 2009.

[Nar97] H. Narayanan. *Submodular Functions and Electrical Networks*, volume 54. Elsevier, 1997.

[NW06] J. Nocedal and S. Wright. *Numerical Optimization*. Springer-Verlag, 2006.

[Pol98] W. Polonik. The silhouette, concentration functions and ML-density estimation under order restrictions. *Annals of Statistics*, 26:1857–1877, 1998.

[QT06] M. Queyranne and F. Tardella. Bimonotone linear inequalities and sublattices of $\mathbb{R}^n$. *Linear Algebra and its Applications*, 413:100–120, 2006.

[RSU17] E. Robeva, B. Sturmfels, and C. Uhler. Geometry of log-concave density estimation. *arXiv:1704.01910*, 2017.

[RW17] J. Royset and R. Wets. Variational analysis of constrained m-estimators. *arXiv:1702.08109*, 2017.

[Sco15] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 2015.

[SH15] M. Slawski and M. Hein. Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random field. *Linear Algebra and its Applications*, 473:145–179, 2015.

[Sil18] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Routledge, 2018.

[SW10] A. Seregin and J. A. Wellner. Nonparametric estimation of convex-transformed densities. *Annals of Statistics*, 38:3751–3781, 2010.

[Tam04] A. Tamura. Applications of discrete convex analysis to mathematical economics. *Publications of the Research Institute for Mathematical Sciecnes*, 40(3):1015–1037, 2004.

[Top76] D. M. Topkis. The structure of sublattices of the product of n lattices. *Pacific Journal of Mathematics*, 65:525–532, 1976.

[Tur93] B. A. Turlach. Bandwidth selection in kernel density estimation: A review. In *CORE and Institut de Statistique*. Citeseer, 1993.

[Wal02] G. Walther. Detecting the presence of mixing with multiscale maximum likelihood. *Journal of the American Statistical Association*, 97:508–513, 2002.

[Was16] L. Wasserman. Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532, 2016.

[WF01] Y. Weiss and W. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, 13, 2001.

[Whi93] D. J. White. Extension of the Frank-Wolfe algorithm to concave nondifferentiable objective functions. *Journal of Optimization Theory and its Applications*, 78(2):283–301, 1993.

[WJ94] M. P. Wand and M. C. Jones. *Kernel Smoothing*. CRC Press, 1994.

[Zwi16] P. Zwiernik. *Semialgebraic Statistics and Latent Tree Models*, volume 146 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2016.

APPENDIX A: PROOF OF THEOREM 1.1

We here provide the proof of our main theorem. First we prove the MTP_2 case, then we outline the necessary steps to adapt the proof to the LLC case. Our proof builds upon [RW17, Theorem 3.1, Corollary 3.5], which states that if the set of functions that one optimizes over is equi upper-semicontinuous (equi-usc), then the MLE exists and is consistent. In general, the class of all log-concave MTP_2 densities on a compact support set is not equi-usc, as noted from the examples in [RW17]. Our theorem strengthens [RW17, Proposition 4.6] by removing the equi-usc assumption and showing that three sample points are sufficient for existence and uniqueness of the log-concave MTP_2 MLE. In particular, we show that one can restrict attention to a strictly smaller subset of log-concave MTP_2 densities, and that a.s. this class is totally bounded when $n \geq 3$ in the MTP_2 case, and $n \geq 2$ in the LLC case. As total boundedness on a compact set is stronger than equi-usc, the conclusion then follows from [RW17]. We start with some lemmas.

LEMMA A.1. *Suppose $n \geq 3$ and $d \geq 2$. Almost surely, X is not contained in a bimonotone hyperplane.*

PROOF. Suppose for contradiction that it is, and let the hyperplane be $H = \{x \in \mathbb{R}^d : ax_i - bx_j = c_{ij}\}$. Let X_1, X_2, X_3 be three points in X that belong to H. Let $\pi_{ij} : \mathbb{R}^d \to \mathbb{R}^2$ be the map $x \mapsto (x_i, x_j)$. Then $\pi_{ij}(X_1), \pi_{ij}(X_2), \pi_{ij}(X_3) \in \pi_{ij}(H)$. However, $\pi_{ij}(X_1), \pi_{ij}(X_2), \pi_{ij}(X_3)$ are distributed as three i.i.d points from a distribution with density $\pi_{ij}(f_0)$, which is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^2$, and hence they a.s. do not lie on a line. □

LEMMA A.2. *Let $L \subset \mathbb{R}^d$ be a linear subspace such that $2 \leq \dim L < d$ and which is not contained in any bimonotone hyperplane. Then, there exist points $u, v \in L$ such that $u \wedge v, u \vee v \in \mathbb{R}^d \setminus L$.*

PROOF. Write $L = \{x : Ax = 0\}$, where $A \in \mathbb{R}^{(n-\dim L)\times d}$. Let B be the reduced-row-Echelon form of A, and let $b \in \mathbb{R}^d$ be the last row of B. Since L is not in any bimonotone hyperplane, either b has three nonzero entries, or it has two nonzero entries which have the same sign. Assume first that

$$b = (\underbrace{0, \ldots, 0}_{i-1}, b_i, \ldots, b_j, \ldots, b_k)$$

has three nonzero entries b_i, b_j, b_k, where $1 \leq i < j < k \leq d$. Let $u \in \mathbb{R}^d$ be such that $u_i = \operatorname{sign}(b_j)b_j$, $u_j = -\operatorname{sign}(b_j)b_i$, and $u_s = 0$ for all $s > i$, $s \neq j$. For the first $i-1$ entries of u, use the reduced-row-Echelon form B so that $Bu = 0$. Similarly, let $v \in \mathbb{R}^d$ be such that $v_j = -\operatorname{sign}(b_j)b_k$, $v_k = \operatorname{sign}(b_j)b_j$, and $v_s = 0$ for all $s \geq 0, s \neq j, k$. For the first $i-1$ entries of v, use the reduced-row-echelon form B so that $Bv = 0$. Now, consider the vector $u \wedge v$. We have that $(u \wedge v)_i = 0, (u \wedge v)_j = (-\operatorname{sign}(b_j)b_i) \wedge (-\operatorname{sign}(b_j)b_k) \neq 0$, and $(u \wedge v)_k = 0$. Therefore, $b \cdot (u \wedge v) = b_j((-\operatorname{sign}(b_j)b_i) \wedge (-\operatorname{sign}(b_j)b_k)) \neq 0$, and thus $B(u \wedge v) \neq 0$. Thus, $(u \wedge v) \notin L$, and therefore also $(u \vee v) \notin L$.

The second case where b has two nonzero entries which have the same sign is similar. This completes the proof. $\square$

Starting from $C^{(0)} = \operatorname{conv}(X)$, we iteratively construct the convex sets

$$C^{(i+1)} = \operatorname{conv}\left(C^{(i)} \cup (C^{(i)} \wedge C^{(i)}) \cup (C^{(i)} \vee C^{(i)})\right) \quad \text{for } i = 0, 1, 2, \ldots.$$

We need these sets to prove the following lemmas.

Lemma A.3. *If* $\dim C^{(i)} < d$, *then, almost surely,* $\dim C^{(i+1)} > \dim C^{(i)}$. *In particular, almost surely,* $\dim C^{(d-2)} = d$.

Clearly $C^{(i)} \supseteq X^{(i)}$ where $X^{(i)}$ is the i-th set in the iterative construction of the set $\operatorname{MMconv}(X)$ in (2.1). Therefore, we obtain the following result.

Corollary A.4. *The iteration (2.1) terminates after at most* $d-1$ *steps.*

Proof of Lemma A.3. Assume that $\dim C^{(i)} < d$. Consider the affine span of $C^{(i)}$, namely $a + L$, where L is a $\dim(C^{(i)})$-dimensional subspace, and a is a point in the relative interior of $C^{(i)}$ (with respect to its affine span). Then, there exists a ball $B(a, \epsilon)$ of radius ϵ and center a such that $B(a, \epsilon) \cap (a + L) \subseteq C^{(i)}$. By Lemma A.1, $C^{(0)}$ and thus $C^{(i)}$ almost surely are not contained in a bimonotone hyperplane. By Lemma A.2, there are vectors $u, v \in L$ such that $u \wedge v, u \vee v \notin L$. Now, let $c_\epsilon > 0$ be a large enough constant so that the vectors $u_\epsilon = u/c_\epsilon, v_\epsilon = v/c_\epsilon$ have norm less than ϵ. Thus, $a + u_\epsilon, a + v_\epsilon \in B(a, \epsilon) \cap (a + L)$. But note that

$$(a + u_\epsilon) \wedge (a + v_\epsilon) = a + (u \wedge v)/c_\epsilon \notin (a + L) \quad \text{and}$$
$$(a + u_\epsilon) \vee (a + v_\epsilon) = a + (u \vee v)/c_\epsilon \notin (a + L).$$

Thus, affine span$(C^{(i+1)}) \supsetneq$ affine span$(C^{(i)})$, and so $\dim C^{(i+1)} > \dim C^{(i)}$. Finally, as $n \geq 3$, almost surely, $\dim C^{(0)} \geq 2$. Therefore, almost surely, $\dim C^{(d-2)} = d$. □

LEMMA A.5. *For $n \geq 3$, for any $T \in \mathbb{R}$, there exists a constant M_T such that for any MTP_2 log-concave density* $\exp(f)$*, if there is an $x \in$ MMconv(X) such that $f(x) > M_T$, then the log-likelihood* $\ell(\exp(f), X) := \frac{1}{n}\sum_{i=1}^n f(X_i)$ *satisfies* $\ell(\exp(f), X) \leq T$.

PROOF OF LEMMA A.5. Let $\exp(f)$ be a log-concave MTP_2 density supported on MMconv(X). Let

$$m = \min_{i\in\{1,\dots,n\}} f(X_i), \quad M = \max_{x\in\mathbb{R}^d} f(x), \quad \text{and} \quad Z = \mathrm{argmax}_{x\in\mathbb{R}^d} f(x).$$

By concavity of f, we have

$$f(x) \geq m, \text{ for all } x \in C^{(0)},$$

since $C^{(0)} = \mathrm{conv}(X)$. We are going to show by induction that

$$f(x) \geq 2^i m - (2^i - 1)M, \text{ for all } x \in C^{(i)}.$$

We already have the base of the induction when $i = 0$. Now, assume that it is true for some $i \geq 0$, and let $x \in C^{(i+1)}$ Then, there exist $a, b \in C^{(i)}$ and $x' \in C^{(i+1)}$ such that x and x' are the minimum and maximum of a and b. Thus, since $\exp(f)$ is MTP_2, we have that

$$f(x) \geq f(a)+f(b)-f(x') \geq 2(2^i m-(2^i-1)M)-M = 2^{i+1}m-(2^{i+1}-1)M,$$

which completes the induction.

Let $m' = 2^{d-2}m - (2^{d-2} - 1)M$. Our function f is bounded below by m' on $C^{(d-2)}$. Observe that for M sufficiently large, we must have that $M - (2^{d-2}m - (2^{d-2} - 1)M) > 1$ since $\exp(f)$ is a density and its integral over $C^{(d-2)}$ is at most 1. Now, note that for any $x \in C^{(d-2)}$, we have

$$\begin{aligned} f\left(Z + \frac{1}{M-m'}(x-Z)\right) &\geq \frac{1}{M-m'}f(x) + \frac{M-m'-1}{M-m'}f(Z) \\ &\geq \frac{m'}{M-m'} + \frac{(M-m'-1)M}{M-m'} = M-1, \end{aligned}$$

where we used concavity of f in the first inequality. Hence, denoting Lebesgue measure on $\mathbb{R}^d$ by μ, we have that

$$\mu\left(\{x : f(x) \geq M - m'\}\right) \geq \mu\left(\left\{Z + \frac{1}{M-m'}C^{(d-2)}\right\}\right) = \frac{\mu(C^{(d-2)})}{(M-m')^d}.$$

Thus,

$$\int f(x)dx \geq e^{M-1} \frac{\mu(C^{(d-2)})}{(M-m')^d},$$

and, therefore, for $\exp(f)$ to be a density, we need $m' \leq \frac{1}{2} e^{(M-1)/d} \mu(C^{(d-2)})^{1/d}$ for large M. Equivalently, since $m' = 2^{d-2}m - (2^{d-2}-1)M$, we need that

$$m \leq \frac{2^{d-2}-1}{2^{d-2}} M - \frac{1}{2^{d-1}} e^{(M-1)/d} \mu(C^{(d-2)})^{1/d}.$$

But then the log-likelihood function $\ell(\exp(f), X)$ satisfies

$$\begin{aligned}\ell(\exp(f), X) &\leq \frac{n-1}{n} M + \frac{1}{n}\left(\frac{2^{d-2}-1}{2^{d-2}} M - \frac{1}{2^{d-1}} e^{(M-1)/d} \mu(C^{(d-2)})^{1/d}\right) \\ &= \frac{1}{n}\left(n - 1 + \frac{2^{d-2}-1}{2^{d-2}}\right) M - \frac{1}{n 2^{d-1}} e^{(M-1)/d} \mu(C^{(d-2)})^{1/d}.\end{aligned}$$

Note that as $M \to \infty$, this expression converges to $-\infty$. Therefore, for every $T \in \mathbb{R}$, there exists M_T such that whenever $\max f(x) \geq M_T$, then, $\ell(\exp(f), X) \leq T$, which completes the proof of Lemma A.5. □

Proof of Theorem 1.1 for the MTP$_2$ case. Let T be the likelihood of the uniform density on MMconv(X), and M_T be the corresponding constant in Lemma A.5. Let

$$\begin{aligned}\mathcal{F} = \{f &: \mathbb{R}^d \to [0, \infty), f \text{ is log-concave MTP}_2, \\ &\int f = 1, \text{support}(f) = \text{MMconv}(X_1, \ldots, X_n), \sup f \leq M_T\}.\end{aligned}$$

By Lemma A.5 and Proposition 2.7, any log-concave MTP$_2$ density not in $\mathcal{F}$ has likelihood strictly less than that of the uniform density on MMconv(X). Thus, it is sufficient to solve (1.2) over the set of densities in $\mathcal{F}$. Fix $n \geq 3$. By Lemma A.3, MMconv$(X_1, \ldots, X_n)$ is a.s. full-dimensional. Since $\mathcal{F}$ is a set of uniformly bounded densities on a compact set, $\mathcal{F}$ is equi-usc. Therefore, existence and consistency of the MLE follows from [RW17, Theorem 3.1, Corollary 3.5] with $\pi \equiv 0$, $\epsilon = 0$, and $F = \mathcal{F}$. □

To prove the LLC case, the analogue of Lemma A.1 is the following.

Lemma A.6. *If $n, d \geq 2$ then, almost surely, $P(X)$ is full-dimensional.*

Proof. Suppose for contradiction that it is not. Since $P(X)$ is $L^\natural$-convex, it must be contained in an $L^\natural$-hyperplane. Such a hyperplane has the form

$x_i - x_j = c_{ij}$ or $x_i = c_i$. Define $\pi_{ij} : \mathbb{R}^d \to \mathbb{R}, x \mapsto x_i - x_j \in \mathbb{R}$. Then $\pi_{ij}(P(X))$ is a point, and hence $\pi_{ij}(X)$ is also a point. But $\pi_{ij}(X)$ are n i.i.d points from a distribution with density $\pi_{ij}(f_0)$ which is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}$, since $\mathrm{supp}(f_0)$ has full dimension. Therefore, a.s. $\pi_{ij}(X)$ is not a singleton for $n \geq 2$, a contradiction. □

Proof of Theorem 1.1 for the LLC case. For sets $C, C' \subset \mathbb{R}^d$, let $C \oplus C' := \{(x - \alpha\mathbf{1}) \wedge y, (x + \alpha\mathbf{1}) \wedge y : \alpha \in \mathbb{R}\, x \in C, y \in C'\}$. Replace the sets $C^{(i)}$ by $D^{(0)} = \mathrm{conv}(X)$, and

$$D^{(i+1)} = \mathrm{conv}(D^{(i)} \cup (D^{(i)} \oplus D^{(i)})).$$

The exact analogue of Lemma A.5 for LLC has an identical proof, with the sets $D^{(i)}$ replacing the sets $C^{(i)}$. Similarly, the proof of Theorem 1.1 for the LLC case is the same as the MTP_2 case, with

$$\mathcal{F} = \{f : \mathbb{R}^d \to [0, \infty) : f \text{ is log-concave LLC}, \int f = 1, \mathrm{support}(f) = P(X), \sup f \leq M_T\}.$$

This completes the proof of Theorem 1.1. □

APPENDIX B: PROOFS OF RESULTS IN SECTION 2

Proof of Proposition 2.6. By definition, $X \subseteq P(X)$, so $P(X) \neq \emptyset$. Since the coordinate of each point in $P(X)$ has a lower and upper bound, $P(X)$ is compact, and thus it is a polytope. Let $P^\sharp(X) \subset \mathbb{R}^{d+1}$ denote the embedding of $P(X)$ into $\mathbb{R}^{d+1}$ via the map $\sharp : \mathbb{R}^d \to \mathbb{R}^{d+1}, x \mapsto \{x + r\mathbf{1}, r \in \mathbb{R}\}$. By [Mur03, §7], a set $P \subset \mathbb{R}^d$ is $L^\natural$-convex if and only if its preimage $P^\sharp \subset \mathbb{R}^{d+1}$ has the form

$$P^\sharp = \{y \in \mathbb{R}^{d+1} : y_i - y_j \leq c_{ij}, i, j \in [n]\}.$$

Taking $c_{ij} = \max_{x \in X}(x_i - x_j)$, we see that $P^\sharp(X)$ has this form, therefore $P(X)$ is $L^\natural$-convex. To see that it is the smallest $L^\natural$-convex set containing X, let P be another $L^\natural$-convex set with pre-image $P^\sharp \subset \mathbb{R}^{d+1}$. For each pair $i, j \in [n], i \neq j$, $\pi_{ij}(P^\sharp)$ must also be L-convex, so it has the following form for some constants $a_{ij} \leq b_{ij}$:

$$\pi_{ij}(P^\sharp) = \{y \in \mathbb{R}^2 : a_{ij} \leq y_1 - y_2 \leq b_{ij}\}.$$

Since $X \subseteq P^\sharp$, it follows that $b_{ij} \geq \max_{x \in X}(x_i - x_j)$, and $a_{ij} \leq \min_{x \in X}(x_i - x_j)$. So $\pi_{ij}(P^\sharp) \supseteq \pi_{ij}(P^\sharp(X))$. Since $P^\sharp$ is convex, it follows that $\pi_{ij}(P^\sharp) \supseteq$

$\bigcap_{i \neq j} \pi_{ij}^{-1}(\text{conv}(\pi_{ij}(P^\sharp(X)))) = P^\sharp(X)$. This concludes the proof of the first statement. For the second statement, suppose that $\tilde{X}$ exists. Since $\tilde{X}$ is $L^\natural$-closed, it satisfies (2.2) for some constant $r > 0$. Since X is finite, this implies $X - v \in r \cdot \mathbb{Z}^d$ for any $v \in X$. Thus there is a unique minimal constant r^* such that $X - v \in r^* \cdot \mathbb{Z}^d$. Define $X' = \frac{1}{r^*}(X - v)$. Then X' is a discrete $L^\natural$-convex set in the sense of [Mur03], so $X' = P(X') \cap \mathbb{Z}^d$. Rearranging gives the RHS of (2.4). Conversely, suppose that there exists some $v \in X$ and $r > 0$ such that $X - v \subset r \cdot \mathbb{Z}^d$. Since X is finite, there is a smallest r with this property, denote it r^*. If $v' \in X$, then $v' - v \in r \cdot \mathbb{Z}^d$, so $X - v' \subset r \cdot \mathbb{Z}^d$, so r^* is independent of the choice of v. Finally, define X' as the RHS of (2.4). By [Mur03], X' is $L^\natural$-closed. By construction, $X' \supseteq X$. To see that it is the smallest set, suppose that $Y \subset \mathbb{R}^d$ is another $L^\natural$-closed set containing X. Since Y is finite, there exists a minimal constant $r(Y) > 0$ such that $X - v \subset r(Y) \cdot \mathbb{Z}^d$. By minimality of r^*, one have $r(Y) \geq r^*$. So $Y \supseteq X'$, with $Y = X'$ if and only if $r(Y) = r^*$. This concludes the proof. □

Proof of Proposition 2.7. Let f^* be the log-concave MTP_2 MLE with support $S \subset \mathbb{R}^d$. Let ψ_n be the objective function. Since f^* is log-concave and MTP_2, S is a min-max closed convex set. Since $\psi_n(f^*) > -\infty$, $X \subseteq S$. As $\text{MMconv}(X)$ is the smallest min-max closed convex set containing X, $\text{MMconv}(X) \subseteq S$. Now, let $S' = S \backslash \text{MMconv}(X)$, and f' be the density f^* restricted to $\text{MMconv}(X)$. Since f^* is MTP_2, f' is also log-concave MTP_2. Furthermore, $\psi_n(f') \geq \psi_n(f^*)$, and equality occurs if and only if S' has measure 0. As f^* is the maximizer, we must have equality, so $\text{MMconv}(X) = S$ a.s, as required. For the LLC case, $P(\tilde{X})$ is the smallest $L^\natural$-convex set that contains X, so the same proof applies. □

APPENDIX C: PROOF OF THEOREM 3.4

To prove Theorem 3.4 for $d = 2$, we use the following simple lemma.

Lemma C.1. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be concave and piecewise linear on a polyhedral complex Δ, and let $u, v \in \text{supp}(f)$. If $u \vee v$ and $u \wedge v$ lie in the same cell of Δ then f is supermodular on the four points, i.e.*

$$f(u) + f(v) \leq f(u \wedge v) + f(u \vee v). \tag{C.1}$$

Proof. Let m be the common midpoint of the segments between u and v and between $u \vee v$ and $u \wedge v$. The"same face" hypothesis implies that the right hand side of (C.1) is equal to $2f(m)$. Since f is concave, the left hand side of (C.1) is bounded above by $2f(m)$. □

Proof of Theorem 3.4. When $X = \{0,1\}^d$, if y is a vector of tight MTP_2 heights, then $h_{X,y}$ equals the Lovász extension [Mur03, §7], which is $L^\natural$-concave, and thus it is also MTP_2. The case $X \subseteq \prod_{i=1}^d \{a_i, b_i\}$ reduces to the discrete cube by an invertible map in each coordinate that preserves the bimonotone property. Now we will prove the case when $X \subset \mathbb{R}^2$ and min-max closed. Note that a polygon in $\mathbb{R}^2$ is not bimonotone if and only if it has an edge with a negative slope. Let y be a set of tight MTP_2 heights on X and let $h_{X,y}$ be the corresponding tent function. Suppose for contradiction that $h_{X,y}$ is not bimonotone. By Theorem 5.1, there exists an edge of $h_{X,y}$ with negative slope. Let $x, x' \in X$ be the two vertices of this edge. Since $x - x'$ has negative slope, x and x' are not comparable. Since X is min-max closed, $x \wedge x', x \vee x' \in X$. As y is MTP_2,

$$y(x') + y(x) \leq y(x \wedge x') + y(x \vee x'). \tag{C.2}$$

Since $[x, x']$ is an edge of the regular subdivision of $h_{X,y}$, and since $h_{X,y}$ is concave,

$$h_{X,y}(x') + h_{X,y}(x) > h_{X,y}(x \wedge x') + h_{X,y}(x \vee x'). \tag{C.3}$$

Since y is concave, $h_{X,y}$ equals y at all points of X. Therefore, (C.2) and (C.3) cannot simultaneously hold. This is the desired contradiction. □

APPENDIX D: PROOFS FOR SECTION 4

Lemma D.1. *For a finite set $X \subset \mathbb{R}^d$, let*

$$\begin{aligned} \min(X) &= (\min_{x\in X} x_i : i = 1, \ldots, d) \in \mathbb{R}^d, \\ \textit{and} \quad \max(X) &= (\max_{x\in X} x_i : i = 1, \ldots, d) \in \mathbb{R}^d. \end{aligned}$$

For any set $X \subset \mathbb{R}^2$, its min-max convex hull equals

$$\mathrm{MMconv}(X) = \mathrm{conv}(X \cup \{\min(X), \max(X)\}).$$

Proof of Lemma D.1. First, note that $\mathrm{conv}(X \cup \{\min(X), \max(X)\}) \subseteq \mathrm{MMconv}(X)$. Now, we will show that the sides of the polygon $\mathrm{conv}(X \cup \{\min(X), \max(X)\})$ are bimonotone. By Corollary 2.3, this implies the claim.

First consider the sides of the polygon located above the diagonal between $\min(X)$ and $\max(X)$. Traveling from $\min(X)$ to $\max(X)$, the sides are either vertical or have nonnegative slope. This is because they have first coordinate bigger than or equal to that of $\min(X)$, and $\mathrm{conv}(X \cup \{\min(X), \max(X)\})$ is convex. Similarly, the sides below the diagonal between $\min(X)$ and $\max(X)$ are either vertical or have nonnegative slope. Therefore, all of the sides of $\mathrm{conv}(X \cup \{\min(X), \max(X)\})$ are bimonotone. □

Lemma D.2. *Algorithm 7 outputs the set of inequalities that defines the set of tight $L^\natural$-concave heights for any $L^\natural$-tidy set $X \subset \mathbb{Z}^d$.*

Proof. Let $\sharp$ be the operation that identifies $\mathbb{R}^d$ with the subspace of $\mathbb{R}^{d+1}$ defined by $\sum_{i=0}^{d} x_i = 0$. By definition, $\mathcal{S}$ contains all inequalities of the form $y(x) + y(x') - y(x \vee x') - y(x \wedge x') \geq 0$ for pairs $x, x' \in \tilde{X}^\sharp$. By [Mur03, Proposition 7.5], it is sufficient to include these inequalities for all pairs $x, x' \in \tilde{X}^\sharp$ such that $\max_i |x_i - x'_i| = 1$. To avoid listing the same inequality multiple times, Algorithm 7 goes through each $x \in \tilde{X}^\sharp$ listing all inequalities where x is the minimal point amongst the four points involved. □

APPENDIX E: PROOF OF THEOREM 5.1

Proof of Theorem 5.1. The only-if direction is simpler, and we prove it first. Let P be a cell of Δ. Suppose that P is not min-max closed. Then we can find $u, v \in P$ such that $u \wedge v \notin P$ or $u \vee v \notin P$. Since P is a convex polyhedron, we have $\frac{1}{2}(u+v) \in P$, and it follows that

$$f(u) + f(v) \ = \ 2 \cdot f(\frac{1}{2}(u+v)) \ > \ f(u \wedge v) + f(u \vee v).$$

The strict inequality follows because f is concave, but not linear on the segment between $u \wedge v$ and $u \vee v$. This contradicts the assumption that f is supermodular. Hence the polyhedron P is min-max closed and by Corollary 2.3, P is defined by bimonotone linear inequalities.

We now prove the if-direction. Suppose that each cell of Δ is defined by bimonotone linear inequalities. Let $u, v \in |\Delta|$. Our goal is to show that $f(u) + f(v) \leq f(u \wedge v) + f(u \vee v)$. Let L be the two-dimensional plane in $\mathbb{R}^d$ that contains $x, y, x \wedge y$ and $x \vee y$. After relabeling we can assume that $u_i \geq v_i$ for $i = 1, \ldots, k$, and $u_i < v_i$ for $i = k+1, \ldots, d$. The plane equals

$$\begin{aligned} L \quad = \quad u \wedge v \quad + \quad \operatorname{span}\bigl\{&(0, \ldots, 0, u_{k+1} - v_{k+1}, \ldots, u_d - v_d)^T, \\ &(v_1 - u_1, \ldots, v_k - u_k, 0, \ldots, 0)^T\bigr\}. \end{aligned}$$

From this representation we see that L is defined by $d-2$ linear equations that are bimonotone, namely $k-1$ in the first k coordinates, and $d-k-1$ in the last $d-k$ coordinates. We identify L with $\mathbb{R}^2$ via the following order-preserving affine-linear isomorphism $\mathbb{R}^2 \to L$:

$$\begin{aligned} (\alpha, \beta) \quad \mapsto \quad u \wedge v \ + \ &\alpha(0, \ldots, 0, u_{k+1} - v_{k+1}, \ldots, u_d - v_d)^T \\ + \ &\beta(v_1 - u_1, \ldots, v_k - u_k, 0, \ldots, 0)^T. \end{aligned}$$

This isomorphism preserves the property of a linear function to be bimonotone. Indeed, suppose ℓ is bimonotone on $\mathbb{R}^d$ and does not vanish on L.

Let $\ell(z) = c + a_i z_i + a_j z_j$, where $a_i a_j \leq 0$. The restriction of ℓ to L is the following affine-linear form in the coordinates α, β:

$$\begin{aligned} &\ell'(\alpha,\beta) = \\ &= c \,+\, a_i \cdot ((u \wedge v)_i + \alpha(0 \vee (u_i - v_i))) \,+\, a_j \cdot ((u \wedge v)_j + \beta(0 \vee (v_j - u_j))) \\ &= (c + a_i(u \wedge v)_i + a_j(u \wedge v)_j) + (a_i(0 \vee (u_i - v_i))) \cdot \alpha + (a_j(0 \vee (v_j - u_j))) \cdot \beta. \end{aligned}$$

This function is bimonotone because $(a_i(0\vee(u_i-v_i)))\cdot(a_j(0\vee(v_j-u_j))) \leq 0$.

The restriction of f to $L \simeq \mathbb{R}^2$ is a piecewise-linear concave function f_L in (α, β), whose support is the restriction $\Delta|_L$ of Δ to L. We apply the reasoning above to each linear function ℓ that vanishes on a codimension one cell of Δ. This shows that $f|_L$ induces a bimonotone subdivision on its support $|\Delta| \cap L$. Our claim follows if $f|_L$ is supermodular. This reduces the proof of Theorem 5.1 to the special case $d = 2$, where we know it already. □

E. Robeva
Department of Mathematics
Massachusetts Institute of Technology
E-mail: erobeva@mit.edu

C. Uhler
Laboratory for Information and Decision Systems,
and Institute for Data, Systems and Society
Massachusetts Institute of Technology
E-mail: cuhler@mit.edu

B. Sturmfels
Department of Mathematics
University of California at Berkeley and
MPI for Mathematics in the Sciences, Leipzig
E-mail: bernd@mis.mpg.de

N. M. Tran
Department of Mathematics
University of Texas at Austin
E-mail: ntran@math.utexas.edu