Your Dreams May Come True with $\text{MTP}_2$...

Caroline Uhler (MIT)

Joint work with Steffen Lauritzen, Piotr Zwiernik, Elina Robeva, Bernd Sturmfels, Ngoc Tran

Bay Area Optimization Meeting
Stanford

May 19, 2018
A distribution (i.e. density function) \( p \) on \( \mathcal{X} = \prod_{v \in V} \mathcal{X}_v \), with \( \mathcal{X}_v \subseteq \mathbb{R} \) discrete or open subset, is multivariate totally positive of order 2 (MTP\(_2\)) if

\[
p(x)p(y) \leq p(x \wedge y)p(x \vee y)
\]

for all \( x, y \in \mathcal{X}, \)

where \( \wedge \) and \( \vee \) are applied coordinate-wise.
Positive dependence and $\text{MTP}_2$ distributions

- A distribution (i.e. density function) $p$ on $\mathcal{X} = \prod_{v \in V} \mathcal{X}_v$, with $\mathcal{X}_v \subseteq \mathbb{R}$ discrete or open subset, is **multivariate totally positive of order 2** ($\text{MTP}_2$) if
  \[
p(x)p(y) \leq p(x \wedge y)p(x \vee y) \quad \text{for all } x, y \in \mathcal{X},\]
  where $\wedge$ and $\vee$ are applied coordinate-wise.

- A random vector $X$ is **positively associated** if for any non-decreasing functions $\phi, \psi : \mathbb{R}^m \to \mathbb{R}$
  \[
  \text{cov}\{\phi(X), \psi(X)\} \geq 0.
  \]

**Theorem (Fortuin, Kasteleyn, Ginibre inequality, 1971, Karlin & Rinott, 1980)**

$\text{MTP}_2$ implies positive association.
No Yule-Simpson Paradox under MTP$_2$!

The **Yule-Simpson paradox** says that we may have two random variables $X$ and $Y$ positively associated, but $X$ and $Y$ negatively associated conditionally on a third variable $Z$.

Sentences in 4863 murder cases in Florida over the six years 1973-1978:

<table>
<thead>
<tr>
<th>Murderer</th>
<th>Sentence</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence</td>
<td>Death</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>59</td>
<td>2547</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>72</td>
<td>2185</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Victim</th>
<th>Murderer</th>
<th>Sentence</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>Black</td>
<td>11</td>
<td>2309</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Black</td>
<td>0</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>White</td>
<td>48</td>
<td>238</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>White</td>
<td>72</td>
<td>2074</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall greater proportion of white murderers receiving death sentence than black (3.2% vs. 2.3%); this trend is reversed given color of victim.

Data from: Range (1979)
Gaussian-like properties of $MTP_2$ distribution

Theorem (Karlin and Rinott, 1980)

If $X$ is $MTP_2$, then

(i) any marginal distribution is $MTP_2$

(ii) any conditional distribution is $MTP_2$

Theorem (Lebowitz, 1972)

If $X$ is positively associated and $A, B \subseteq V$ are disjoint, then

\[ X_A \perp \perp X_B \iff \text{cov}(X_u, X_v) = 0 \text{ for all } u \in A, v \in B. \]
Theorem (Bølviken 1982, Karlin & Rinott, 1983)

A multivariate Gaussian distribution \( p(x; K) \) is \( \text{MTP}_2 \) if and only if the inverse covariance matrix \( K \) is an \( M \)-matrix, that is
\[
K_{uv} \leq 0 \quad \text{for all } u \neq v.
\]
Gaussian $\text{MTP}_2$ distributions

**Theorem (Bølviken 1982, Karlin & Rinott, 1983)**

A multivariate Gaussian distribution $p(x; K)$ is MTP$_2$ if and only if the inverse covariance matrix $K$ is an $M$-matrix, that is  

$$K_{uv} \leq 0 \quad \text{for all } u \neq v.$$  

**Ex:** Grades of 88 students in 5 math subjects  
(Mardia, Kent and Bibby, 1979)

\[
S = \begin{pmatrix}
305.7680 & 127.2226 & 101.5794 & 106.2727 & 117.4049 \\
127.2226 & 172.8422 & 85.1573 & 94.6729 & 99.0120 \\
101.5794 & 85.1573 & 112.8860 & 112.1134 & 121.8706 \\
106.2727 & 94.6729 & 112.1134 & 220.3804 & 155.5355 \\
117.4049 & 99.0120 & 121.8706 & 155.5355 & 297.7554
\end{pmatrix}
\]
Gaussian $\text{MTP}_2$ distributions

**Theorem (Bølviken 1982, Karlin & Rinott, 1983)**

A multivariate Gaussian distribution $p(x; K)$ is $\text{MTP}_2$ if and only if the inverse covariance matrix $K$ is an $M$-matrix, that is $K_{uv} \leq 0$ for all $u \neq v$.

**Ex:** Grades of 88 students in 5 math subjects (Mardia, Kent and Bibby, 1979)

\[
S^{-1} = 10^{-3} \cdot \begin{pmatrix}
5.2446 & -2.4351 & -2.7395 & 0.0116 & -0.1430 \\
-2.4351 & 10.4268 & -4.7078 & -0.7928 & -0.1660 \\
0.0116 & -0.7928 & -7.0486 & 9.8829 & -2.0184 \\
-0.1430 & -0.1660 & -4.7050 & -2.0184 & 6.4501
\end{pmatrix}
\]

Although sample distribution is not quite $\text{MTP}_2$, any fitted reasonable Gaussian graphical model is $\text{MTP}_2$.
Discrete $MTP_2$ distributions

**Reminder:** A distribution $p$ on $\mathcal{X} \subseteq \mathbb{R}^m$ is $MTP_2$ if

$$p(x)p(y) \leq p(x \land y)p(x \lor y), \text{ for all } x, y \in \mathcal{X}.$$ 

Distribution of 3 binary variables $X$, $Y$ and $Z$ is $MTP_2$ iff

- $p_{001}p_{110} \leq p_{000}p_{111}$
- $p_{010}p_{101} \leq p_{000}p_{111}$
- $p_{100}p_{011} \leq p_{000}p_{111}$
- $p_{011}p_{101} \leq p_{001}p_{111}$
- $p_{011}p_{110} \leq p_{010}p_{111}$
- $p_{101}p_{110} \leq p_{100}p_{111}$
- $p_{001}p_{010} \leq p_{000}p_{011}$
- $p_{001}p_{100} \leq p_{000}p_{101}$
- $p_{010}p_{100} \leq p_{000}p_{110}$
Reminder: A distribution \( p \) on \( \mathcal{X} \subseteq \mathbb{R}^m \) is MTP\(_2\) if
\[
p(x)p(y) \leq p(x \land y)p(x \lor y), \quad \text{for all } x, y \in \mathcal{X}.
\]

- Distribution of 3 binary variables \( X, Y \) and \( Z \) is MTP\(_2\) iff
  \[
  \begin{align*}
  p_{001}p_{110} & \leq p_{000}p_{111} & p_{010}p_{101} & \leq p_{000}p_{111} & p_{100}p_{011} & \leq p_{000}p_{111} \\
  p_{011}p_{101} & \leq p_{001}p_{111} & p_{011}p_{110} & \leq p_{010}p_{111} & p_{101}p_{110} & \leq p_{100}p_{111} \\
  p_{001}p_{010} & \leq p_{000}p_{011} & p_{001}p_{100} & \leq p_{000}p_{101} & p_{010}p_{100} & \leq p_{000}p_{110}
  \end{align*}
  \]

- Dataset on **EPH-gestosis** analyzed by *Wermuth & Marchetti (2014)*
  - edema (high body water retention)
  - proteinuria (high amounts of urinary proteins)
  - hypertension (elevated blood pressure)

\[
\begin{bmatrix}
  n_{000} & n_{010} & n_{001} & n_{011} \\
  n_{100} & n_{110} & n_{101} & n_{111}
\end{bmatrix}
= \begin{bmatrix}
  3299 & 107 & 1012 & 58 \\
  78 & 11 & 65 & 19
\end{bmatrix}
\]

- This sample distribution is MTP\(_2\)!
  Although when you sample 3-dim binary distributions only about 2% are MTP\(_2\).
$MTP_2$ constraints are often implicit

$X$ is $MTP_2$ in:
- ferromagnetic Ising models
- Markov chains with $MTP_2$ transitions
- order statistics of i.i.d. variables
- Brownian motion tree models

$|X|$ is $MTP_2$ in:
- Gaussian / binary tree models
- Gaussian / binary latent tree models
  - Binary latent class models
  - Single factor analysis models
Density estimation

Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ from an unknown distribution on $\mathbb{R}^d$ with density $p$, can we estimate $p$?

- **parametric**: assume $p$ lies in some parametric family
  - finite-dimensional optimization problem (estimate parameters)
  - restrictive: real-world distribution might not lie in specified family

- **non-parametric**: assume that $p$ lies in a non-parametric family:
  - infinite-dimensional optimization problem
  - need constraints that are:
    - strong enough so that there is no spiky behavior
    - weak enough so that function class is large
ML Estimation for Gaussian $\text{MTP}_2$ distributions

Let $X_1, \ldots, X_n \sim \mathcal{N}(0, \Sigma)$, $S := \frac{1}{n} \sum_{i=1}^{n} X_iX_i^T$ sample covariance matrix.

**Primal: Max-Likelihood:**

\[
\begin{align*}
\text{maximize} & \quad \log \det(K) - \text{trace}(KS) \\
\text{subject to} & \quad K_{uv} \leq 0, \ \forall \ u \neq v.
\end{align*}
\]

**Dual: Min-Entropy:**

\[
\begin{align*}
\text{minimize} & \quad - \log \det(\Sigma) - p \\
\text{subject to} & \quad \Sigma_{vv} = S_{vv}, \ \Sigma_{uv} \geq S_{uv}.
\end{align*}
\]

- Maximum likelihood estimation under $\text{MTP}_2$ is a convex optimization problem with strong duality
ML Estimation for Gaussian MTP\textsubscript{2} distributions

Let $X_1, \ldots, X_n \sim \mathcal{N}(0, \Sigma)$, $S := \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T$ sample covariance matrix.

**Primal: Max-Likelihood:**

maximize $\log \det(K) - \text{trace}(KS)$

subject to $K_{uv} \leq 0, \ \forall \ u \neq v.$

**Dual: Min-Entropy:**

minimize $- \log \det(\Sigma) - p$

subject to $\Sigma_{vv} = S_{vv}, \ \Sigma_{uv} \geq S_{uv}.$

- Maximum likelihood estimation under MTP\textsubscript{2} is a **convex optimization problem with strong duality**
- the global optimum is characterized by **KKT conditions**
- **Complimentary slackness** implies that the MLE $\hat{K}^{-1} = \hat{\Sigma}$ satisfies $(\hat{\Sigma}_{uv} - S_{uv}) \hat{K}_{uv} = 0 \ \forall u \neq v$
Let $X_1, \ldots, X_n \sim \mathcal{N}(0, \Sigma)$, $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T$ sample covariance matrix.

**Primal: Max-Likelihood:**

<table>
<thead>
<tr>
<th>maximize $K_{\geq 0}$</th>
<th>log det($K$) − trace($KS$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject to $K_{uv} \leq 0$, $\forall u \neq v$.</td>
<td></td>
</tr>
</tbody>
</table>

**Dual: Min-Entropy:**

<table>
<thead>
<tr>
<th>minimize $\Sigma_{\geq 0}$</th>
<th>$- \log det(\Sigma) − p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject to $\Sigma_{vv} = S_{vv}$, $\Sigma_{uv} \geq S_{uv}$.</td>
<td></td>
</tr>
</tbody>
</table>

- Maximum likelihood estimation under $\text{MTP}_2$ is a convex optimization problem with strong duality
- the global optimum is characterized by KKT conditions
- Complimentary slackness implies that the MLE $\hat{K}^{-1} = \hat{\Sigma}$ satisfies $(\hat{\Sigma}_{uv} - S_{uv}) \hat{K}_{uv} = 0$ $\forall u \neq v$
- **Linear algebra:** If $M$ is an M-matrix, then $(M^{-1})_{ij} \geq 0$ for all $i, j$
ML Estimation for Gaussian MTP$_2$ distributions

Let $X_1, \ldots, X_n \sim \mathcal{N}(0, \Sigma)$, $S := \frac{1}{n} \sum_{i=1}^{n} X_iX_i^T$ sample covariance matrix.

**Primal: Max-Likelihood:**

<table>
<thead>
<tr>
<th>maximize $K_{\geq 0}$</th>
<th>$\log \det(K) - \text{trace}(KS)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject to $K_{uv} \leq 0, \forall u \neq v.$</td>
<td></td>
</tr>
</tbody>
</table>

**Dual: Min-Entropy:**

<table>
<thead>
<tr>
<th>minimize $\Sigma \geq 0$</th>
<th>$- \log \det(\Sigma) - p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject to $\Sigma_{vv} = S_{vv}, \Sigma_{uv} \geq S_{uv}.$</td>
<td></td>
</tr>
</tbody>
</table>

- Maximum likelihood estimation under MTP$_2$ is a convex optimization problem with strong duality
- the global optimum is characterized by KKT conditions
- Complimentary slackness implies that the MLE $\hat{K}^{-1} = \hat{\Sigma}$ satisfies $(\hat{\Sigma}_{uv} - S_{uv}) \hat{K}_{uv} = 0 \quad \forall u \neq v$
- **Linear algebra:** If $M$ is an M-matrix, then $(M^{-1})_{ij} \geq 0$ for all $i, j$
- **Graphical model:** $\hat{G}$ (support of $\hat{K}$) is in general sparse!!!
Estimation in the high-dimensional setting $(p > n)$

**Theorem (Slawski and Hein, 2015)**

The MLE in a Gaussian $\text{MTP}_2$ model exists with probability 1 when the sample size $n \geq 2$ (independent of the number of variables $m$).

New proof: Construct primal & dual feasible point by single-linkage clustering

\[
S = \begin{pmatrix}
1 & 0.7 & 0.6 \\
0.7 & 1 & 0.5 \\
0.6 & 0.5 & 1 \\
0.2 & 0.1 & -0.3 \\
0.1 & -0.5 & 0.1
\end{pmatrix} \quad \rightarrow \quad \begin{pmatrix}
1 & 0.7 & 0.6 & 0.2 & 0.2 \\
0.7 & 1 & 0.6 & 0.2 & 0.2 \\
0.6 & 0.6 & 1 & 0.2 & 0.2 \\
0.2 & 0.2 & 0.2 & 1 & 0.4 \\
0.2 & 0.2 & 0.2 & 0.4 & 1
\end{pmatrix}
\]
Properties of the maximum likelihood graph \( \hat{G} \)

Let \( R \) be a correlation matrix.

- **MWSF\((R)\):** maximum weight spanning forest of \( R \)
- **EC\((R)\):** excess correlation graph; identifies correlations that are not explained by MWSF\((R)\), i.e.

\[(i, j) \in EC(R) \iff R_{ij} \geq \prod_{uv \in \overline{ij}} R_{uv},\]

where \( \overline{ij} \) is unique path between \( i \) and \( j \) in MWSF\((R)\).

**Theorem (Lauritzen, Uhler and Zwiernik, 2018)**

\[ MWSF(R) \subseteq \hat{G} \subseteq EC(R) \quad (\text{under regularity conditions}) \]

\( \Rightarrow \) can be used in coordinate descent algorithms
Personality traits

240 individuals were asked to rate themselves on the scale 1-9 with respect to 32 different personality traits.

Correlation matrix of personality traits from the data set described in Malle and Horowitz (1995).
Correlation matrix of the data set on personality traits resulting from switching the signs of the 16 (negative) traits that constitute the first block of variables in the previous figure.
Graphical model resulting from estimation under MTP₂.

- thick red edges: $\text{MWSF}(R)$
- blue edges: $\hat{G} \setminus \text{MWSF}(R)$
- thin grey edges: $\text{EC}(R) \setminus \hat{G}$
Density estimation

Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ from an unknown distribution on $\mathbb{R}^d$ with density $p$, can we estimate $p$?

- **parametric**: assume $p$ lies in some parametric family
  - finite-dimensional optimization problem (estimate parameters)
  - restrictive: real-world distribution might not lie in specified family

- **non-parametric**: assume that $p$ lies in a non-parametric family:
  - infinite-dimensional optimization problem
  - need constraints that are:
    - strong enough so that there is no spiky behavior
    - weak enough so that function class is large

Caroline Uhler (MIT)
Shape-constrained density estimation

- monotonically decreasing densities: [Grenander 1956, Rao 1969]
- log-concave densities: [Cule, Samworth, and Stewart 2010]
- generalized additive models with shape constraints: [Chen and Samworth 2016]
Shape-constrained density estimation

- monotonically decreasing densities: [Grenander 1956, Rao 1969]
- **log-concave** densities: [Cule, Samworth, and Stewart 2010]
- generalized additive models with shape constraints: [Chen and Samworth 2016]

**Maximum likelihood estimation under MTP$_2$:** Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$,

$$
\text{maximize}_p \sum_{i=1}^{n} \log(p(x_i))
$$

s.t. \ $p$ is an MTP$_2$ density.
Shape-constrained density estimation

- monotonically decreasing densities: \([\text{Grenander } 1956, \text{ Rao } 1969]\)
- convex densities: \([\text{Anevski } 1994, \text{ Groeneboom, Jongbloed, and Wellner } 2001]\)
- \textbf{log-concave} densities: \([\text{Cule, Samworth, and Stewart } 2010]\)
- generalized additive models with shape constraints: \([\text{Chen and Samworth } 2016]\)

\textbf{Maximum liklihood estimation under MTP}_2: \text{ Given i.i.d. samples } X = \{x_1, ..., x_n\} \subset \mathbb{R}^d,

\[
\text{maximize}_p \sum_{i=1}^{n} \log(p(x_i)) \quad \text{s.t.} \quad p \text{ is an MTP}_2 \text{ density.}
\]

\[p \text{ log-concave.}\]
Log-concave density estimation

- Log-concavity is a natural assumption: ensures density is continuous and includes many distributions: Gaussian, Uniform\((a, b)\), Gamma\((k, \theta)\) for \(k \geq 1\), Beta\((a, b)\) for \(a, b \geq 1\), etc.
Log-concave density estimation

- Log-concavity is a natural assumption: ensures density is continuous and includes many distributions: Gaussian, Uniform($a, b$), Gamma($k, \theta$) for $k \geq 1$, Beta($a, b$) for $a, b \geq 1$, etc.

**Theorem (Cule, Samworth and Stewart, 2008)**

With probability 1, a log-concave MLE $\hat{\rho}$ exists and is unique for $n \geq d$. Moreover, $\log(\hat{\rho})$ is a tent-function supported on the convex hull of the data $P(X) = \text{conv}(x_1, \ldots, x_n)$ with tent poles on $X$.

Finite-dimensional optimization problem!

---

Figure 1. The ‘tent-like’ structure of the graph of the logarithm of the maximum likelihood estimator for bivariate data. (2009) have studied its theoretical properties. Rufibach (2007) compared different algorithms for computing the univariate estimator, including the iterative convex minorant algorithm (Groeneboom and Wellner, 1992; Jongbloed, 1998), and three others. Dümbgen, Hüsler and Rufibach (2007) also present an Active Set algorithm, which has similarities with the vertex direction and vertex reduction algorithms described in Groeneboom, Jongbloed and Wellner (2008). Walther (2010) provides a nice recent review article on inference and modelling with log-concave densities. Other recent related work includes Seregin and Wellner (2009), Schuhmacher, Hüsler and Dümbgen (2010), Schuhmacher and Dümbgen (2010) and Koenker and Mizera (2010). For univariate data, it is also well-known that there exist maximum likelihood estimators of a non-increasing density supported on $[0, 1]$ (Grenander, 1956) and of a convex, decreasing density (Groeneboom, Jongbloed and Wellner, 2001).

Figure 1 gives a diagram illustrating the structure of the maximum likelihood estimator on the logarithmic scale. This structure is most easily visualised for two-dimensional data, where one can imagine associating a ‘tent pole’ with each observation, extending vertically out of the plane. For certain tent pole heights, the graph of the logarithm of the maximum likelihood estimator can be thought of as the roof of a taut tent stretched over the tent poles. The fact that the logarithm of the maximum likelihood estimator is of this ‘tent function’ form constitutes part of the proof of its existence and uniqueness.

In Sections 3.1 and 3.2, we discuss the computational problem of how to adjust the $n$ tent pole heights so that the corresponding tent functions converge to the logarithm of the maximum likelihood estimator. One reason that this computational problem is so challenging in more than one dimension is the fact that it is difficult to describe the set of tent pole heights that correspond to concave functions. The key observation, discussed in Section 3.1, is that it is possible to minimise a modified objective function that is convex (though non-differentiable). This allows us to apply the powerful non-differentiable convex optimisation methodology of the subgradient method (Shor, 1985) and a variant called Shor’s $\rho$-algorithm, which has been implemented by Kappel and Kuntsevich (2000).

As an illustration of the estimates obtained, Figure 2 presents plots of the maximum likelihood estimator, and its logarithm, for 1000 observations from a standard bivariate normal distribution.
Existence, uniqueness, and shape of the MLE

Theorem (Robeva, Sturmfels, Tran and Uhler, 2018)

The maximum likelihood estimator under log-concavity and MTP$_2$ exists and is unique for $n \geq 3$ samples with probability 1.

Proof uses convergence properties for log-concave distributions, and does not shed light on the shape of the MLE.
Existence, uniqueness, and shape of the MLE

Theorem (Robeva, Sturmfels, Tran and Uhler, 2018)

The maximum likelihood estimator under log-concavity and $MTP_2$ exists and is unique for $n \geq 3$ samples with probability 1.

Proof uses convergence properties for log-concave distributions, and does not shed light on the shape of the MLE.

Theorem (Robeva, Sturmfels, Tran and Uhler, 2018)

If $X \subseteq \mathbb{R}^2$ or $X \subseteq \{0,1\}^d$ (min-max closed), then

- MLE $p^*$ is the exponential of a tent function, i.e., $p^* = \exp(h_{X,y^*})$
- The set of heights for which $\exp(h_{X,y})$ is $MTP_2$ is a convex polytope.
Shape of the MLE

$X_1, \ldots, X_{50} \sim \mathcal{N}(0, I)$; compute MLE using conditional gradient method.
Let $X = \{(0, 0, 0), (6, 0, 0), (6, 4, 0), (6, 4, \frac{3}{2}), (8, 4, 2)\}$. The log-concave MLE is not MTP$_2$. The MLE is a tent function on $X \cup \{(6, 3, \frac{3}{2}), (7.5, 4, \frac{3}{2})\}$ with the following subdivision:
Let $X = \{(0, 0, 0), (6, 0, 0), (6, 4, 0), (6, 4, \frac{3}{2}), (8, 4, 2)\}$. The log-concave MLE is not MTP$_2$. The MLE is a tent function on $X \cup \{(6, 3, \frac{3}{2}), (7.5, 4, \frac{3}{2})\}$ with the following subdivision:
Shape of the MLE in the general case?

- Let $X = \{(0, 0, 0), (6, 0, 0), (6, 4, 0), (6, 4, \frac{3}{2}), (8, 4, 2)\}$. The log-concave MLE is not MTP$_2$. The MLE is a tent function on $X \cup \{(6, 3, \frac{3}{2}), (7.5, 4, \frac{3}{2})\}$ with the following subdivision:

- **Conjecture:** The MTP$_2$ log-concave MLE is the exponential of a tent function.

  **But where to place the tent poles?**
Conclusions

- $\text{MTP}_2$ constraints are often implicit and reflect real processes
  - ferromagnetism
  - latent tree models

- Explicit $\text{MTP}_2$ constraints are useful in high-dimensional setting

- Alternative to glasso that does not require a tuning parameter

- $\text{MTP}_2$ represents interesting shape constraint for non-parametric density estimation: broad enough class to be of interest in applications, constrained enough to obtain good density estimates with few samples
References

- Robeva, Sturmfels, Tran and Uhler: Maximum likelihood estimation for totally positive log-concave densities; on the arXiv within a month

Thank you!