Your Dreams May Come True with $MTP_2$...

Caroline Uhler (MIT)

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

Simons Workshop:
Hyperbolic Polynomials and Hyperbolic Programming

May 2, 2019
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 \textbf{multivariate totally positive of order 2 (MTP}\(_2\)) if

\[
p(x)p(y) \leq p(x \land y)p(x \lor y) \quad \text{for all } x, y \in \mathcal{X},
\]

where \( \land \) and \( \lor \) 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

$$p(x)p(y) \leq p(x \land y)p(x \lor y)$$

for all $x, y \in \mathcal{X}$, where $\land$ and $\lor$ 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 $\text{MTP}_2$!

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

<table>
<thead>
<tr>
<th>Victim</th>
<th>Murderer</th>
<th>Sentence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Death</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>Black</td>
<td>11</td>
<td>2309</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>White</td>
<td>0</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Black</td>
<td>48</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>White</td>
<td>72</td>
<td>2074</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 $\text{MTP}_2$ distribution

**Reminder:** A distribution $p$ on $\mathcal{X} \subseteq \mathbb{R}^m$ is $\text{MTP}_2$ if

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

**Theorem (Lebowitz, 1972; Karlin and Rinott, 1980)**

*If $X$ is $\text{MTP}_2$, then*

(i) *any marginal distribution is $\text{MTP}_2$*

(ii) *any conditional distribution is $\text{MTP}_2$*

(iii) $X_A \perp \perp X_B \iff \text{cov}(X_u, X_v) = 0$ for all $u \in A$, $v \in B$
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

\[ K_{uv} \leq 0 \quad \text{for all} \quad u \neq v. \]
Gaussian $MTP_2$ 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:** 2016 Monthly correlations of global stock markets (InvestmentFrontier.com)

$$S = \begin{pmatrix}
1.000 & 0.606 & 0.731 & 0.618 & 0.613 \\
0.606 & 1.000 & 0.550 & 0.661 & 0.598 \\
0.731 & 0.550 & 1.000 & 0.644 & 0.569 \\
0.618 & 0.661 & 0.644 & 1.000 & 0.615 \\
0.613 & 0.598 & 0.569 & 0.615 & 1.000
\end{pmatrix}

Nasdaq
Canada
Europe
UK
Australia

Nasdaq
Canada
Europe
UK
Australia
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

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

**Ex:** 2016 monthly correlations of global stock markets (*InvestmentFrontier.com*)

$$S^{-1} = \begin{pmatrix}
\text{Nasdaq} & \text{Canada} & \text{Europe} & \text{UK} & \text{Australia} \\
2.629 & -0.480 & -1.249 & -0.202 & -0.490 \\
-0.480 & 2.109 & -0.039 & -0.790 & -0.459 \\
-1.249 & -0.039 & 2.491 & -0.675 & -0.213 \\
-0.202 & -0.790 & -0.675 & 2.378 & -0.482 \\
-0.490 & -0.459 & -0.213 & -0.482 & 1.992
\end{pmatrix}$$

Sample distribution is $\text{MTP}_2$! If you sample a correlation matrix uniformly at random the probability of it being $\text{MTP}_2$ is $<10^{-6}$!
Discrete $\text{MTP}_2$ distributions

Reminder: A distribution $p$ on $\mathcal{X} \subseteq \mathbb{R}^m$ is $\text{MTP}_2$ 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 $\text{MTP}_2$ iff

  $$p_{001}p_{110} \leq p_{000}p_{111} \quad p_{010}p_{101} \leq p_{000}p_{111} \quad p_{100}p_{011} \leq p_{000}p_{111}$$

  $$p_{011}p_{101} \leq p_{001}p_{111} \quad p_{011}p_{110} \leq p_{010}p_{111} \quad p_{101}p_{110} \leq p_{100}p_{111}$$

  $$p_{001}p_{010} \leq p_{000}p_{011} \quad p_{001}p_{100} \leq p_{000}p_{101} \quad p_{010}p_{100} \leq p_{000}p_{110}$$

- 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 $\text{MTP}_2$! Although when you sample 3-dim binary distributions only about 2% are $\text{MTP}_2$. 

Caroline Uhler (MIT) | MTP$_2$ distributions | Berkeley, May 2019
MTP$_2$ constraints are often implicit

$|X|$ is MTP$_2$ in:

- Gaussian / binary tree models
- Gaussian / binary latent tree models
  - Binary latent class models
  - Single factor analysis models
An exponential family is a parametric model with density

$$p_\theta(x) = \exp\left(\langle \theta, T(x) \rangle - A(\theta) \right),$$

sample space \( \mathcal{X} \) with measure \( \nu \), sufficient statistics \( T : \mathcal{X} \to \mathbb{R}^d \), and space of canonical parameters: \( C = \{ \theta \in \mathbb{R}^d : A(\theta) < +\infty \} \)

- Gaussian distribution: \( A(\theta) = -\alpha \log \det(\theta) \), \( C = \mathbb{S}^p_{>0} \)
- Hyperbolic exponential family: \( A(\theta) = -\alpha \log(f(\theta)) \), \( f \) hyperbolic with hyperbolicity cone \( C \)
Hyperbolic MTP\(_2\) exponential families

- An exponential family is a parametric model with density
  \[
p_{\theta}(x) = \exp\left(\langle \theta, T(x) \rangle - A(\theta) \right),
\]
sample space \(\mathcal{X}\) with measure \(\nu\), sufficient statistics \(T : \mathcal{X} \to \mathbb{R}^d\), and space of canonical parameters: \(C = \{\theta \in \mathbb{R}^d : A(\theta) < +\infty\}\)

- Gaussian distribution: \(A(\theta) = -\alpha \log \det(\theta), \ C = \mathbb{S}^p_{>0}\)

- Hyperbolic exponential family: \(A(\theta) = -\alpha \log(f(\theta)), f\) hyperbolic with hyperbolicity cone \(C\)

**Theorem (Lauritzen, Uhler & Zwiernik, 2019)**

The space of canonical parameters for any MTP\(_2\) exponential family is given by \(C \cap K\), where \(K \subset \mathbb{R}\) is a closed convex cone whose dual is generated by
\[
\{T(x \land y) + T(x \lor y) - T(x) - T(y) : x, y \in \mathcal{X}\ \text{differing in 2 entries}\}.
\]
Density estimation

Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^m$ from an unknown distribution on $\mathbb{R}^m$ 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
ML Estimation for Gaussian $\text{MTP}_2$ distributions

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

**Primal: Max-Likelihood:**

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

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

**Dual: Min-Entropy:**

$\minimize \quad -\log \det(\Sigma) - m$

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

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

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

**Primal: Max-Likelihood:**

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

**Dual: Min-Entropy:**

<table>
<thead>
<tr>
<th>minimize $\Sigma \succeq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$- \log \det(\Sigma) - m$</td>
</tr>
<tr>
<td>subject to $\Sigma_{vv} = S_{vv}$, $\Sigma_{uv} \geq S_{uv}$.</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
- Complementary slackness implies that the MLE $\hat{K}^{-1} = \hat{\Sigma}$ satisfies $(\hat{\Sigma}_{uv} - S_{uv}) \hat{K}_{uv} = 0$, $\forall u \neq v$
ML Estimation for Gaussian MTP2 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:**

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

**Dual: Min-Entropy:**

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

- Maximum likelihood estimation under MTP2 is a convex optimization problem with strong duality.
- the global optimum is characterized by KKT conditions.
- Complementary 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 \( i, j \).
ML Estimation for Gaussian MTP$_2$ distributions

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

**Primal: Max-Likelihood:**

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

**Dual: Min-Entropy:**

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

- Maximum likelihood estimation under MTP$_2$ is a convex optimization problem with strong duality.
- The global optimum is characterized by KKT conditions.
- Complementary 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 $i, j$.
- **Graphical model:** $\hat{G}$ (support of $\hat{K}$) is in general **sparse**!!!
Ultrametric matrices and inverse M-matrices

- \( U \) is **ultrametric**: \( U_{ii} \geq U_{ij} = U_{ji} \geq \min(U_{ik}, U_{jk}) \geq 0 \) for all \( i, j, k \).

**Theorem (Dellacherie, Martinez and San Martin, 2014)**

Let \( U \) be an ultrametric matrix with strictly positive entries on the diagonal. Then \( U \) is non-singular if and only if no two rows are equal. Moreover, if \( U \) is non-singular, then \( U^{-1} \) is an M-matrix.

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

The MLE in a Gaussian MTP\(_2\) model exists with probability 1 when \( n \geq 2 \).
Ultrametric matrices and inverse M-matrices

- *U* is **ultrametric**: \( U_{ii} \geq U_{ij} = U_{ji} \geq \min(U_{ik}, U_{jk}) \geq 0 \) for all \( i, j, k \).

**Theorem (Dellacherie, Martinez and San Martin, 2014)**

*Let \( U \) be an ultrametric matrix with strictly positive entries on the diagonal. Then \( U \) is non-singular if and only if no two rows are equal. Moreover, if \( U \) is non-singular, then \( U^{-1} \) is an M-matrix.*

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

*The MLE in a Gaussian MTP\(_2\) model exists with probability 1 when \( n \geq 2 \).*

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

Given i.i.d. samples $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^m$ from an unknown distribution on $\mathbb{R}^m$ 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
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 $X = \{x_1, ..., x_n\} \subset \mathbb{R}^m$,

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

s.t. \quad p \text{ is an MTP}_2 \text{ density.}$$
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 $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^m$,

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

s.t. \quad p \text{ is an MTP}_2 \text{ density.}

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

- Log-concavity is a natural assumption: it ensures the 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.

Figure 1. The 'tent-like' structure of the graph of the logarithm of the maximum likelihood estimator for bivariate data.

Theorem (Cule, Samworth and Stewart, 2008)
When \(n \geq m + 1\), a log-concave MLE \(\hat{p}\) exists and is unique with probability 1. Moreover, \(\log(\hat{p})\) is a tent-function supported on the convex hull of the data.

Finite-dimensional optimization problem!
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)**

When $n \geq m + 1$, a log-concave MLE $\hat{p}$ exists and is unique with probability 1. Moreover, $\log(\hat{p})$ is a **tent-function** supported on the convex hull of the data.

**Finite-dimensional optimization problem!**
Questions:

- When does the MLE under log-concavity and MTP$_2$ / LLC exist? Is it unique?
- What is the shape of the MLE under log-concavity and MTP$_2$ / LLC?
  - Is the MLE always exp(tent function)?
- Can we compute the MLE?
A function $f : \mathbb{R}^m \to \mathbb{R}$ is MTP$_2$ if
\[ f(x)f(y) \leq f(x \land y)f(x \lor y) \quad \text{for all } x, y \in \mathbb{R}^m. \]

A function $f : \mathbb{R}^m \to \mathbb{R}$ is log-$L^\|$-concave (LLC) if
\[ f(x)f(y) \leq f((x + \alpha \mathbf{1}) \land y)f(x \lor (y - \alpha \mathbf{1})) \quad \forall \alpha \geq 0 \text{ and } x, y \in \mathbb{R}^m. \]
Log-$L^\triangleright$-concave (LLC) functions

- A function $f : \mathbb{R}^m \to \mathbb{R}$ is MTP$_2$ if
  \[ f(x)f(y) \leq f(x \land y) f(x \lor y) \quad \text{for all } x, y \in \mathbb{R}^m. \]

- A function $f : \mathbb{R}^m \to \mathbb{R}$ is log-$L^\triangleright$-concave (LLC) if
  \[ f(x)f(y) \leq f((x + \alpha 1) \land y) f(x \lor (y - \alpha 1)) \quad \forall \alpha \geq 0 \text{ and } x, y \in \mathbb{R}^m. \]

**Theorem (Murota, 2008)**

A function $f : \mathbb{Z}^m \to \mathbb{R}$ is LLC if and only if it is log-concave, i.e.,
\[ f(x)f(y) \leq f\left(\left\lfloor \frac{x + y}{2} \right\rfloor\right) f\left(\left\lceil \frac{x + y}{2} \right\rceil\right) \quad \text{for all } x, y \in \mathbb{Z}^m. \]

**Ex.:** A Gaussian distribution with covariance matrix $\Sigma$ is LLC if and only if $K = \Sigma^{-1}$ is a **diagonally dominant M-matrix**, i.e.,
\[ K_{ij} \leq 0 \text{ for all } i \neq j \quad \text{and} \quad \sum_{j=1}^{m} K_{ij} \geq 0 \text{ for all } i = 1, \ldots m. \]
**Existence and uniqueness of the MLE**

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

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 $\mathbb{R}^m$. 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.

- This result is in contrast with existence of the MLE under log-concavity, where $n \geq m + 1$ samples are needed for existence.
- Proof uses convergence properties for log-concave distributions, and does not shed light on the shape of the MLE.
Under MTP$_2$ we need the density to be nonzero at additional points:

→ "Min-max convex hull" of $X$
Support of the MLE

Under MTP$_2$ we need the density to be nonzero at additional points:

$$\implies \text{"Min-max convex hull" of } X$$

- **MM}(X) := smallest min-max closed set $S$ containing $X$, i.e. $x, y \in S \Rightarrow x \land y, x \lor y \in S$
- **MMconv}(X) := smallest min-max closed & convex set containing $X$

Is it always true that MMconv(X) = conv(MM(X))?
Support of the MLE

Under MTP$_2$ we need the density to be nonzero at additional points:

\[ \Rightarrow \text{"Min-max convex hull" of } X \]

- **MM($X$)** := smallest min-max closed set $S$ containing $X$, i.e.
  \[ x, y \in S \Rightarrow x \wedge y, x \vee y \in S \]

- **MMconv($X$)** := smallest min-max closed & convex set containing $X$

Is it always true that MMconv($X$) = conv(MM($X$))? 

**Lemma**

*If $X \subseteq \mathbb{R}^2$ or $X \subseteq \{0, 1\}^m$, then MMconv($X$) = conv(MM($X$)).*
Ex: Consider \( X = \{(0, 0, 0), (6, 0, 0), (6, 4, 0), (8, 4, 2)\} \subseteq \mathbb{R}^3 \).

- \( \text{MM}(X) = X \)
- But \( \text{conv}(\text{MM}(X)) \) is not min-max closed!

\[
(6, 4, 3/2) = \max\{(6, 4, 0), (6, 3, 3/2)\} \not\in \text{conv}(\text{MM}(X)).
\]
**Ex:** Consider \( X = \{(0, 0, 0), (6, 0, 0), (6, 4, 0), (8, 4, 2)\} \subseteq \mathbb{R}^3 \).

- \( \text{MM}(X) = X \)
- But \( \text{conv}(\text{MM}(X)) \) is not min-max closed!

\[
(6, 4, 3/2) = \max\{(6, 4, 0), (6, 3, 3/2)\} \not\in \text{conv}(\text{MM}(X)).
\]

**Theorem (The 2-D Projections Theorem)**

\[
\text{Let } \pi_{ij} : \mathbb{R}^m \to \mathbb{R}, \ x \mapsto (x_i, x_j). \ \text{Then for any finite subset } X \subseteq \mathbb{R}^m,
\]

\[
\text{MMconv}(X) = \bigcap_{1 \leq i < j \leq m} \pi_{ij}^{-1}\left(\text{conv}(\pi_{ij}(\text{MM}(X)))\right).
\]
Theorem (Robeva, Sturmfels, Tran and Uhler, 2018)

Let $X \subset \mathbb{R}^m$ be a finite set of points. The exponential of a tent function $h_{X,y}$ is $\text{MTP}_2$ if and only if all of the walls of the subdivision $h$ induces are bimonotone.

A linear inequality $a \cdot x + b \leq 0$ is bimonotone if it has the form $a_i x_i + a_j x_j + b \leq 0$, where $a_i a_j \leq 0$. 

\[ (0, 0) \quad (0, 1) \quad (1, 0) \quad (1, 1) \]
Shape of the MLE

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

If \( X \subseteq \mathbb{R}^2 \) or \( X \subseteq \{0, 1\}^m \) (\( X \subseteq \mathbb{Q}^m \)), then the MTP\(_2\) (LLC) MLE is of the form \( \exp(\text{tent function}) \) and the set of MTP\(_2\) (LLC) tent pole heights define a convex polytope.

\[ \implies \text{We can use the conditional gradient method to compute the MLE} \]
Conclusions

- We conjecture that the \( \text{MTP}_2 \)-MLE is always the exponential of a tent function (we provide conjectured tent pole locations).
- LLC estimate provides an \( \text{MTP}_2 \) estimate (might not be the MLE).
- Total positivity constraints are often implicit and reflect real processes.
  - ferromagnetism
  - latent tree models
- Total positivity 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.
- \( \text{MTP}_2 \) / LLC is well-suited for high-dimensional applications.
MTP2 distributions not only have broad applications for data analysis, but also lead to interesting new problems in combinatorics, geometry & algebra.

- Robeva, Sturmfels, Tran & Uhler: Maximum likelihood estimation for totally positive log-concave densities (arXiv:1806.10120)
- Lauritzen, Uhler, & Zwiernik: Total positivity in structured binary distributions (to appear on the arXiv today!)