Your Dreams May Come True with MTP_2...

Caroline Uhler (MIT)

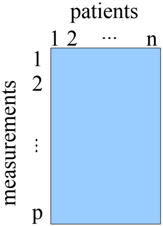
Joint work with Shaun Fallat, Steffen Lauritzen, Kayvan Sadeghi, Nanny Wermuth, and Piotr Zwiernik

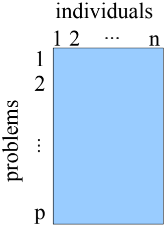
Optimization and Parsimonious Modeling
IMA

January 26, 2016

Problem

Given large data sets for example from **medical tests** or **IQ tests**, determine a sparse graph that describes the dependencies between the variables.





Two common approaches:

- **Machine Learning:** Graphical lasso
- **Applied Statistics:** Chow-Liu and subsequent stepwise selection

Overview

1. MTP_2 distributions
2. Properties of MTP_2 distributions related to sparsity
3. Models that imply MTP_2
4. Maximum likelihood estimation under MTP_2

Positive dependence and MTP_2 distributions

- A distribution (i.e. density function) p on $\mathcal{X}$ is **multivariate totally positive of order** 2 (MTP_2) if

$$p(x)p(y) \quad \leq \quad p(x \wedge y)p(x \vee y) \qquad \text{for all } x, y \in \mathcal{X} \subset \mathbb{R}^m.$$

- A random vector X is **positively associated** if for any non-decreasing functions $\phi, \psi : \mathbb{R}^m \to \mathbb{R}$

$$\mathrm{cov}\{\phi(X), \psi(X)\} \geq 0.$$

Theorem (Fortuin Kasteleyn Ginibre inequality, 1971)

MTP_2 *implies positive association.*

Discrete and Gaussian MTP_2 distribution

Example: Binary vector $X = (X_1, X_2, X_3) \in \{0,1\}^3$ is MTP_2 if and only if

$$\begin{array}{lll}
p_{001}p_{110} \le p_{000}p_{111} & p_{010}p_{101} \le p_{000}p_{111} & p_{100}p_{011} \le p_{000}p_{111} \\
p_{011}p_{101} \le p_{001}p_{111} & p_{011}p_{110} \le p_{010}p_{111} & p_{101}p_{110} \le p_{100}p_{111} \\
p_{001}p_{010} \le p_{000}p_{011} & p_{001}p_{100} \le p_{000}p_{101} & p_{010}p_{100} \le p_{000}p_{110}
\end{array}$$

Theorem (Horn and Johnson, 1991)

*A multivariate Gaussian distribution $p(x;\theta)$ is MTP_2 if and only if the inverse covariance matrix θ is an **M-matrix**, that is*

$$\theta_{ij} \le 0 \qquad \textit{for all } \; i \neq j.$$

Theorem (Karlin and Rinott, 1980)

If $p(x) > 0$ and p is MTP_2 for any pair of coordinates when the others are held constant, then p is MTP_2.

Properties of MTP_2 distribution

Theorem (FLSUWZ, 2015)

If $X = (X_1, \ldots, X_m)$ *is* MTP_2*, then*

(i) *any marginal distribution is* MTP_2

(ii) *any conditional distribution is* MTP_2

(iii) *marginal independence structure:*

$$X_i \perp\!\!\!\perp X_j \iff cov(X_i, X_j) = 0$$

(iv) *conditional independence structure:*

$$X_A \perp\!\!\!\perp X_B \mid X_C \implies X_A \perp\!\!\!\perp X_B \mid X_{C \cup \{k\}}$$

(iv) *composition property:*

$$X_A \perp\!\!\!\perp X_B \mid X_C \ and \ X_A \perp\!\!\!\perp X_D \mid X_C \implies X_A \perp\!\!\!\perp X_{B \cup D} \mid X_C$$

(iv) *singelton transitivity property:*

$$X_i \perp\!\!\!\perp X_j \mid X_C \ and \ X_i \perp\!\!\!\perp X_j \mid X_{C \cup \{k\}} \implies X_i \perp\!\!\!\perp X_k \mid X_C \ or \ X_j \perp\!\!\!\perp X_k \mid X_C$$

Occurrence of MTP_2 distributions

MTP_2 contraints appear to be extremely **restrictive**:

- 3-dim. Gaussian distributions: about 5% are MTP_2
- 4-dim. Gaussian distributions: about 0.09% are MTP_2
- 3-dim. binary distributions: about 2% are MTP_2
- 4-dim. binary distributions: about 0% are MTP_2

Constraints are less restrictive with additional Markov structure!

For 3-dim. Gaussian distributions:

- if $1 \perp\!\!\!\perp 2 \mid 3$: 25% are MTP_2,
- if in addition $1 \perp\!\!\!\perp 3 \mid 2$: 50% are MTP_2,
- if $1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3$: 100% are MTP_2.

Example: EPH-gestosis

Dataset collected 40 years ago in a study on "Pregnancy and Child Development" by the German Research Foundation and recently analyzed by *Wermuth and Marchetti (2014)*.

EPH-gestosis: disease syndrome for pregnant women; three symptoms

- edema (high body water retention)
- proteinuria (high amounts of urinary proteins)
- hypertension (elevated blood pressure)

Observed counts:

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

This sample distribution is MTP_2!

Example: Math grades

Data: grades of 88 students in

`Mechanics`, `Vectors`, `Algebra`, `Analysis`, `Statistics`

$$S = \begin{array}{cc} & \begin{array}{ccccc} \text{mechanics} & \text{vectors} & \text{algebra} & \text{analysis} & \text{statistics} \end{array} \\ & \left(\begin{array}{rrrrr} 305.7680 & 127.2226 & 101.5794 & 106.2727 & 117.4049 \\ 127.2226 & 172.8422 & 85.1573 & 94.6729 & 99.0120 \\ 101.5794 & 85.1573 & 112.8860 & 112.1134 & 121.8706 \\ 106.2727 & 94.6729 & 112.1134 & 220.3804 & 155.5355 \\ 117.4049 & 99.0120 & 121.8706 & 155.5355 & 297.7554 \end{array}\right) \begin{array}{l} \text{mechanics} \\ \text{vectors} \\ \text{algebra} \\ \text{analysis} \\ \text{statistics} \end{array} \end{array}$$

$$S^{-1} = 10^{-3} \cdot \begin{array}{cc} & \begin{array}{ccccc} \text{mechanics} & \text{vectors} & \text{algebra} & \text{analysis} & \text{statistics} \end{array} \\ & \left(\begin{array}{rrrrr} 5.2446 & -2.4351 & -2.7395 & 0.0116 & -0.1430 \\ -2.4351 & 10.4268 & -4.7078 & -0.7928 & -0.1660 \\ -2.7395 & -4.7078 & 26.9548 & -7.0486 & -4.7050 \\ 0.0116 & -0.7928 & -7.0486 & 9.8829 & -2.0184 \\ -0.1430 & -0.1660 & -4.7050 & -2.0184 & 6.4501 \end{array}\right) \begin{array}{l} \text{mechanics} \\ \text{vectors} \\ \text{algebra} \\ \text{analysis} \\ \text{statistics} \end{array} \end{array}$$

Although sample distribution is not quite MTP_2, any fitted reasonable Gaussian graphical model is MTP_2

MTP_2 constraints are often implicit

Pairwise interaction model for a graph $G = (V, E)$:

$$p(x) = \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j),$$

where ψ_{ij} positive functions, Z the normalizing constant.

Theorem (FLSUWZ, 2015)

p is MTP_2 if and only if ψ_{ij} are MTP_2 functions.

- **Example:** Ferromagnetism in Ising models

$$\psi_{ij}(x_i, x_j) = \exp(-\theta_{ij} x_i x_j), \quad \theta_{ij} \leq 0$$

Signed MTP_2 distributions

A Gaussian / discrete random vector $X = (X_1, \dots, X_m)$ has a **signed MTP_2 distribution** if and only if:

- **Discrete:** The distribution of X is MTP_2 up to a permutation of the values in each $\mathcal{X}_i$
- **Gaussian:** There exists a diagonal matrix $D \in \{-1, +1\}^m$ such that DX is MTP_2.

The following models are signed MTP_2:

- Gaussian / binary pairwise interaction models on trees
- Binary latent class models (*Allman, Rhodes, Sturmfels & Zwiernik, 2013*)
- Gaussian / binary latent tree models
    - Single factor analysis models

ML Estimation for Gaussian graphical models

Primal: Max-Likelihood:

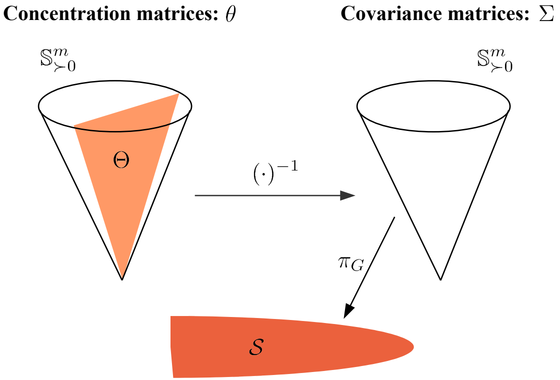
$$\underset{\theta \succeq 0}{\text{maximize}} \quad \log\det(\theta) - \mathrm{trace}(\theta S)$$

$$\text{subject to} \quad \theta_{uv} = 0, \ \ \forall\, uv \notin E, u \neq v.$$

Dual: Min-Entropy: $G = (V, E)$

$$\underset{\Sigma \succeq 0}{\text{minimize}} \quad -\log\det(\Sigma) - m$$

$$\text{subject to} \quad \Sigma_{uv} = S_{uv}, \ \ \forall\, uv \in E, \text{ and } u = v.$$



ML Estimation for Gaussian MTP_2 distributions

Primal: Max-Likelihood:

$$\underset{\theta \succeq 0}{\text{maximize}} \quad \log\det(\theta) - \mathrm{trace}(\theta S)$$
$$\text{subject to} \quad \theta_{uv} \leq 0, \;\; \forall\, u \neq v.$$

Dual: Min-Entropy:

$$\underset{\Sigma \succeq 0}{\text{minimize}} \quad -\log\det(\Sigma) - m$$
$$\text{subject to} \quad \Sigma_{vv} = S_{vv}, \; \Sigma_{uv} \geq S_{uv}.$$

Theorem

The MLE based on S exists if and only if there exists $\Sigma \succ 0$ with $\Sigma \geq S$. It is then equal to the unique element $\hat{\theta} = \hat{\Sigma}^{-1} \succ 0$ that satisfies the following system of equations and inequalities

(a) *Primal feasibility:* $\hat{\theta}_{uv} \leq 0 \;\; \forall u \neq v,$

(b) *Dual feasibility:* $\hat{\Sigma}_{vv} - S_{vv} = 0 \;\; \forall v, \quad \hat{\Sigma}_{uv} - S_{uv} \geq 0 \;\; \forall u \neq v$

(c) *Complimentary slackness:* $(\hat{\Sigma}_{uv} - S_{uv})\,\hat{\theta}_{uv} = 0 \;\; \forall u \neq v.$

Note: We get sparsity for free!!

ML Estimation for Gaussian MTP_2 distributions

Theorem (Slawski and Hein, 2015)

The MLE in a Gaussian MTP_2 model exists with probability 1 when $n \geq 2$.

Theorem (LUZ, 2016)

Let S be a sample correlation matrix and $\hat{\theta}$ the MLE of the concentration matrix in the Gaussian MTP_2 model. Let $G_{MST}(S)$ be the maximal spanning tree of S and $G(\hat{\theta})$ the concentration graph. Then

$$G_{MST}(S) \subset G(\hat{\theta}).$$

Algorithm: *Input:* Sample correlation matrix S
Output: Graph under Gaussian signed MTP_2 model

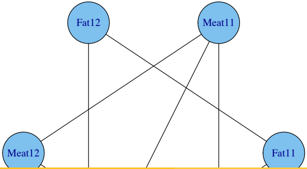
- Let $D \in \{-1, 1\}^p$ diagonal s.t. Chow-Liu tree of DSD is positive
- Compute MLE $\hat{\Sigma}$ based on DSD under Gaussian MTP_2 model
- Output $G(\hat{\Sigma}^{-1})$

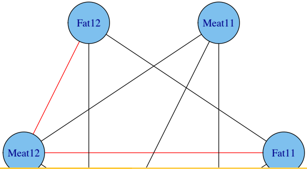
Example: Carcass

344 measurements of the thickness of meat and fat layers at different locations of a slaughter pig

$$
S^{-1} = \begin{array}{c} \\ \end{array}
\begin{array}{cccccc|c}
\text{Fat11} & \text{Meat11} & \text{Fat12} & \text{Meat12} & \text{Fat13} & \text{Meat13} & \\
0.40 & \color{red}{0.04} & -0.23 & -0.07 & -0.19 & \color{red}{0.05} & \text{Fat11} \\
\color{red}{0.04} & 0.15 & -0.01 & -0.06 & -0.05 & -0.06 & \text{Meat11} \\
-0.23 & -0.01 & 0.51 & \color{red}{0.07} & -0.23 & -0.05 & \text{Fat12} \\
-0.07 & -0.06 & \color{red}{0.07} & 0.14 & -0.00 & -0.09 & \text{Meat12} \\
-0.19 & -0.05 & -0.23 & -0.00 & 0.54 & \color{red}{0.03} & \text{Fat13} \\
\color{red}{0.05} & -0.06 & -0.05 & -0.09 & \color{red}{0.03} & 0.16 & \text{Meat13}
\end{array}
$$

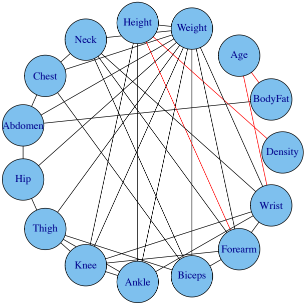
$$
S = \begin{array}{cccccc|c}
\text{Fat11} & \text{Meat11} & \text{Fat12} & \text{Meat12} & \text{Fat13} & \text{Meat13} & \\
11.34 & 0.74 & 8.42 & 2.06 & 7.66 & \color{red}{-0.76} & \text{Fat11} \\
0.74 & 32.97 & 0.67 & 35.94 & 2.01 & 31.97 & \text{Meat11} \\
8.42 & 0.67 & 8.91 & 0.31 & 6.84 & \color{red}{-0.60} & \text{Fat12} \\
2.06 & 35.94 & 0.31 & 51.79 & 2.18 & 41.47 & \text{Meat12} \\
7.66 & 2.01 & 6.84 & 2.18 & 7.62 & 0.38 & \text{Fat13} \\
\color{red}{-0.76} & 31.97 & \color{red}{-0.60} & 41.47 & 0.38 & 41.44 & \text{Meat13}
\end{array}
$$



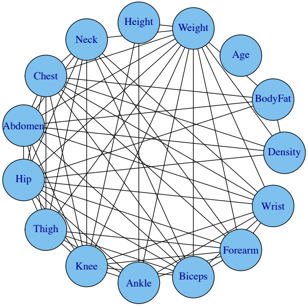


Example: BodyFat

241 observations on 15 variables: age, weight, height, percentage of body fat, body density, and the circumferences of various body parts.



Under MTP_2 constraint



Using glasso

Conclusions and future work

- MTP_2 constraints reflect real processes and models
  - ferromagnetism
  - latent class models with positive associations
  - latent Gaussian/binary tree models
- they lead to some beautiful theory (exponential families, convexity, combinatorics, semialgebraic geometry)
- they are useful in high-dimensional settings

References

- Fallat, Lauritzen, Sadeghi, Uhler, Wermuth, and Zwiernik: Total positivity in Markov structures (arXiv:1510.01290)
- Lauritzen, Uhler, and Zwiernik: Totally positive exponential families and graphical models (on the arXiv shortly)

Thank you!