Hüsler–Reiss Graphical Models for Multivariate Extremes

Manuel Hentschel

University of Geneva

May 16, 2024

Workshop on graphical models and clustering, Montpellier 2024
Based on

*Statistical Inference for Hüsler-Reiss Graphical Models Through Matrix Completions*
(Hentschel, M., Engelke, S., and Segers, J., arxiv 2210.14292)

*Graphical models for multivariate extremes*
(Engelke, S., Hentschel, M., Lalancette, M., and Röttger, F., arxiv 2402.02187)

*R package: graphicalExtremes*
(Engelke, S., Hitz, A., Gnecco, N., and Hentschel, M.)
https://github.com/sebastian-engelke/graphicalExtremes
Outline

Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
  Inference on known graph structures
  Structure estimation

Application
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
  Inference on known graph structures
  Structure estimation

Application
Definition (Graph)
An undirected graph $G = (V, E)$ consists of a finite set of vertices $V$ and a set of edges

$$E \subseteq \{\{i, j\} \mid i, j \in V, i \neq j\}.$$ 

Definition (Graphical model)
A random vector $X$ indexed by $V$ is a graphical model with respect to graph $G = (V, E)$, if it satisfies the Markov property

$$\{i, j\} \notin E \Rightarrow X_i \perp \perp X_j \mid X_{V \setminus \{i, j\}}.$$
Graphical models – Motivation

- Use of structural (expert) knowledge
- Interpretability
- Sparsity
  - Faster computation
  - Implicit regularization
Gaussian graphical models

- Parametrized in terms of precision matrix $\Theta = \Sigma^{-1}$
- Density

$$p(x) = (2\pi)^{-d/2} |\Theta|^{1/2} \exp\left(-\frac{1}{2} (x - \mu)^\top \Theta (x - \mu)\right)$$

- Conditional independence is encoded as

$$\Theta_{ij} = 0 \iff X_i \perp \!\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}$$

- Popular inference method: “Graphical Lasso”

$$\Theta_{\text{glasso}} = \arg\min_{\Theta \in S^+} -\ell(\Theta, \hat{\Sigma}) + \lambda \|\Theta\|_1$$

$$\Theta = \begin{pmatrix} 5 & 0 & -4 & 0 \\ 0 & 3 & -2 & 0 \\ -4 & -2 & 8 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
  Inference on known graph structures
  Structure estimation

Application
Multivariate generalized Pareto distribution

- Random vector $X$ with standard Pareto univariate margins
- $Z$ is multivariate Pareto if

$$
\mathbb{P}(Z \leq z) = \lim_{u \to \infty} \mathbb{P}(u^{-1}X < z \mid \max_{j=1,\ldots,d} X_j > u)
$$

- supported on

$$
\mathcal{L} = \left\{ x \in [0, \infty)^d \mid \max_{j=1,\ldots,d} x_j \geq 1 \right\}
$$
Exponent measure (density)

- Distribution of $Z$ characterized by exponent measure $\Lambda$ with

$$
\mathbb{P}(Z \leq z) = \frac{\Lambda([0, z] \setminus [0, 1])}{\Lambda(L)}.
$$

- We assume that $\Lambda$ is absolutely continuous:

$$
\Lambda(A) = \int_A \lambda(x) \, dx.
$$

- Then $Z$ has density

$$
f_Z(z) = \frac{\lambda(z)}{\Lambda(L)}.
$$

- Important homogeneity properties:

$$
\lambda(\alpha x) = \alpha^{-(d+1)} \lambda(x), \quad \Lambda(\alpha A) = \alpha^{-1} \Lambda(A).
$$
Background: Graphical models

Background: Multivariate generalized Pareto distributions

**Extremal graphical models**

Hüsler–Reiss graphical models

**Inference for Hüsler–Reiss graphical models**
- Inference on known graph structures
- Structure estimation

Application
Extremal graphical models

- Conditional independence problematic on $\mathcal{L}$.
- For $z_2 < 1$ we must (!) have

$$Z_1 \not\perp \perp Z_3 \mid Z_2 = z_2$$

- Approach by Engelke and Hitz (2020):
  - Assume positive, continuous exponent measure density $\lambda$
  - Let $Z^{(k)} = Z \mid \{Z_k > 1\}$
  - Define

$$Z_A \perp_e Z_B \mid Z_C \iff Z_A^{(k)} \perp Z_B^{(k)} \mid Z_C^{(k)} \quad \forall k \in V$$

- Implies factorization of density $\lambda$ (for $A \cup B \cup C = V$):

$$\lambda(z) = \frac{\lambda_{A \cup C}(z_{A \cup C})\lambda_{B \cup C}(z_{B \cup C})}{\lambda_C(z_C)}$$
More general definitions

- The previous definition requires **connected models** with Lebesgue densities

- More general definition in Engelke et al. (2022a)
  - Replaces $\mathcal{L}^{(k)}$ by **rectangular test sets**
  - Works with general exponent measures
  - Allows **disconnected** models
  - Compatible with “$\cdot \perp_{e} \cdot | \cdot$”
  - Can be applied to Lévy processes

- (Active research about directed graphical models)
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
  Inference on known graph structures
  Structure estimation

Application
Choice of univariate margins (!!!)

- So far, we have considered standard Pareto margins
- For connected Hüsler–Reiss models, standard exponential margins are more convenient
- We can always transform between the two!

\[
\begin{align*}
Y_i & \sim \text{Exp}(1) & Z_i & \sim \text{Pareto}(1) \\
Y_i & = \log(Z_i) & Z_i & = \exp(Y_i) \\
f_Y(y) & = f_Z(\exp(y)) \prod_{i=1}^{d} \exp(y_i) & f_Z(z) & = f_Y(\log(z)) \prod_{i=1}^{d} z_i^{-1} \\
\mathcal{L}_Y & = \{y \in [-\infty, \infty)^d \mid \max_i y_i \geq 0\} & \mathcal{L}_Z & = \{z \in [0, \infty)^d \mid \max_i z_i \geq 1\} \\
f_Y(y + \mathbf{1}\beta) & = f_Y(y) \exp(-\beta) & f_Z(\alpha z) & = \alpha^{-(d+1)} f_Z(z)
\end{align*}
\]
Choice of univariate margins (!!)

Support of $Z, Z^{(1)}, Z^{(2)}, ...$:

Support of $Y, Y^{(1)}, Y^{(2)}, ...$:
Variogram matrices

Definition
A square matrix $\Gamma \in \mathbb{R}^{d \times d}$ is conditionally negative definite if it satisfies

$$\Gamma = \Gamma^\top,$$
$$\text{diag}(\Gamma) = 0,$$
$$v^\top \Gamma v < 0 \quad \forall 0 \neq v \perp 1.$$

Also arise as
- variogram matrices (for $\Sigma_{ij} = \text{Cov}(X_i, X_j)$):

$$\Gamma_{ij} = \text{Var}(X_i - X_j) = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}$$

- Euclidean distance matrices:

$$\Gamma_{ij} = |P_i - P_j|^2$$

$$\Gamma = \begin{pmatrix}
0 & 9 & 25 \\
9 & 0 & 16 \\
25 & 16 & 0
\end{pmatrix}$$

$$\Sigma = \begin{pmatrix}
9.78 & 3.78 & -1.56 \\
3.78 & 6.78 & 1.44 \\
-1.56 & 1.44 & 12.11
\end{pmatrix}$$
The Hüsler–Reiss distribution

Definition (Hüsler–Reiss Pareto distribution)

$Y$ is Hüsler–Reiss Pareto distributed with parameter matrix $\Gamma$ if its exponent measure density satisfies

$$
\lambda(y; \Gamma) = \exp\left(-y_k\right) \cdot \varphi_{d-1}(\tilde{y}_{\setminus \{k\}}; \Sigma^{(k)}), \quad y \in \mathbb{R}^d, k \in \{1, \ldots, d\},
$$

where $\varphi_{d-1}$ denotes the $d-1$-dimensional centered normal density with covariance $\Sigma^{(k)}$ and

$$
\Sigma^{(k)}_{ij} = \frac{1}{2} (\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}), \quad i, j \neq k,
$$

$$
\tilde{y}_i = y_i - y_k + \frac{1}{2} \Gamma_{ik}, \quad i \neq k.
$$

Introduced by Hüsler and Reiss (1989) as limit of $\max_{i=1\ldots n} X_i$ where $X_i \sim \mathcal{N}(0, \Sigma_n)$ and

$$
(1 - (\Sigma_n)_{ij}) \log n \to \frac{1}{4} \Gamma_{ij}
$$
A stochastic construction

- Recall the density
  \[ \lambda(y; \Gamma) = \exp(-y_k) \cdot \varphi_{d-1}(\tilde{y}\setminus\{k\}; \Sigma^{(k)}) \]

- Let \( Y \sim \lambda(y; \Gamma) \) and \( Y^{(k)} = Y | \{Y_k > 0\} \).
- Then:
  \[ Y^{(k)} \overset{d}{=} W^{(k)} + E1, \]

  with \( W_{k}^{(k)} = 0 \), \( \mu_{i}^{(k)} = -\frac{1}{2} \Gamma_{ik} \), and

  \[ E \sim \text{Exp}(1), \]

  \[ W_{\setminus\{k\}}^{(k)} \sim \mathcal{N}(\mu^{(k)}, \Sigma^{(k)}). \]
Hüsler–Reiss graphical models

- (Gaussian) precision matrices

\[
\Theta^{(k)} = (\Sigma^{(k)})^{-1} \in \mathbb{R}^{(d-1)\times(d-1)}
\]
\[
\Theta^{(k)}_{ij} = \begin{cases} 0 \iff W^{(k)}_i \perp \perp W^{(k)}_j | W^{(k)} \setminus \{i,j\} \\
\end{cases}
\]

- Proposition 1 in Engelke and Hitz (2020):

\[
\Theta^{(k)}_{ij} = \begin{cases} 0 \iff Y_i \perp \perp e_{Y_j | Y_{\setminus \{i,j\}}} \\
\end{cases}
\]

- Lemma 1 in Engelke and Hitz (2020):

\[
\Theta^{(k)}_{ij} = \Theta^{(k')}_{ij} \quad \forall i, j \neq k, k'
\]

- Definition 3.1 in Hentschel et al. (2022):

Hüsler–Reiss precision matrix \(\Theta \in \mathbb{R}^{d\times d}\) with

\[
\Theta_{ij} = \Theta^{(k)}_{ij} \quad \text{for some } k \neq i, j
\]
Hüsler–Reiss precision matrix

Proposition

With $P = I - d^{-1}11^\top$, the Hüsler–Reiss precision matrix $\Theta$ can be expressed as

$$
\Theta = \Sigma^+
$$

$$
\Sigma = P(-\frac{1}{2}\Gamma)P
$$

Intuition:

- $\Sigma$ is positive semi-definite:

$$
\nu^\top\Sigma\nu = -\frac{1}{2}(P\nu)^\top\Gamma(P\nu) \geq 0
$$

- with kernel $\mathbb{R}1$:

$$
\Sigma 1 = (\ldots P)1 = 0
$$

- and so is its Moore–Penrose inverse $\Theta$.
Density and stochastic construction

Proposition

The Hüsler–Reiss exponent measure density $\lambda(\cdot; \Gamma)$ can be expressed as

$$
\lambda(y; \Gamma) = c \cdot \exp(-y^\top e_d) \cdot \exp(-\frac{1}{2} y^\top \Theta y + y^\top r_{\Theta}),
$$

with $e_d = d^{-1} 1$ and $r_{\Theta} \perp 1$.

Corresponding stochastic construction:

$$
Y^{(1)} \overset{d}{=} W + E1
$$

$W \sim \mathcal{N}(\mu_{\Theta}, \Theta^+)$

for $Y^{(1)} = Y \mid 1^\top Y > 0$. 

15
Let $S \in \mathbb{R}^{d \times d}$ such that

$$\Gamma_{ij} = S_{ii} + S_{jj} - 2S_{ij}.$$ 

Then $\Theta$ can also be expressed as

$$\Theta = \lim_{t \to \infty} (S + t11^T)^{-1}.$$ 

Furthermore (Wan and Zhou (2023)),

$$(\Sigma + t11^T)^{-1} = \Theta + t^{-1}d^{-2}11^T.$$
Summary: Hüsler–Reiss precision matrix

<table>
<thead>
<tr>
<th>Parameter Matrix</th>
<th>Gaussian</th>
<th>Hüsler–Reiss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision Matrix</td>
<td>$\Sigma^{-1}$</td>
<td>$(P(-\frac{1}{2}\Gamma)P)^+$</td>
</tr>
<tr>
<td>Density</td>
<td>$\propto \exp\left(-\frac{1}{2}|y - \mu|_\Theta^2\right)$</td>
<td>$\propto \exp\left(-\frac{1}{2}|y - \mu\Theta|_\Theta^2\right) \cdot \ldots$</td>
</tr>
<tr>
<td>Graphical structure</td>
<td>$\Theta_{ij} = 0$</td>
<td>$\Theta_{ij} = 0$</td>
</tr>
</tbody>
</table>
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

**Inference for Hüsler–Reiss graphical models**

- Inference on known graph structures
- Structure estimation

Application
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
  Inference on known graph structures
  Structure estimation

Application
Inference for Gaussian graphical models (on a known graph)

- Interpretation of entries in matrices $\Sigma$ and $\Theta$:
  - $\Sigma_{ij}$: Covariance between $Y_i$ and $Y_j$ $\Rightarrow$ estimate using $\hat{\Sigma}_{ij}$
  - $\Theta_{ij}$: Conditional independence indicated by zeros $\Rightarrow$ known from graph structure

- Resulting matrix completion problems:

\[
\Sigma = \begin{pmatrix}
5 & 4 & 4 & ? & ? \\
4 & 8 & 4 & ? & ? \\
4 & 4 & 8 & 6 & 6 \\
? & ? & 6 & 10 & 7 \\
? & ? & 6 & 7 & 13 \\
\end{pmatrix}
\]

\[
\Theta = \begin{pmatrix}
? & ? & ? & 0 & 0 \\
? & ? & ? & 0 & 0 \\
0 & 0 & ? & ? & ? \\
0 & 0 & ? & ? & ? \\
\end{pmatrix}
\]

- Solution in Speed and Kiiveri (1986):
  - Explicit construction for decomposable graphs
  - Convergent algorithm for general graphs
Inference for Hüsler–Reiss graphical models (on a known graph)

- Interpretation of entries in matrices $\Gamma$ and $\Theta$:
  \begin{align*}
  \Gamma_{ij}: & \text{ Marginal distribution of } Y_{\{ij\}} \Rightarrow \text{ estimate how?} \\
  \Theta_{ij}: & \text{ Conditional independence indicated by zeros } \Rightarrow \text{ known from graph structure}
  \end{align*}

- Resulting matrix completion problem:

  \[ \begin{pmatrix}
  0 & 1 & 6 & ? & ? \\
  1 & 0 & 3 & ? & ? \\
  6 & 3 & 0 & 3 & 4 \\
  ? & ? & 3 & 0 & 3 \\
  ? & ? & 4 & 3 & 0 \\
  \end{pmatrix} \quad \begin{pmatrix}
  ? & ? & ? & 0 & 0 \\
  ? & ? & ? & 0 & 0 \\
  0 & 0 & ? & ? & ? \\
  0 & 0 & ? & ? & ? \\
  \end{pmatrix} \]

- Solutions?
Empirical variogram

- Engelke and Volgushev (2020) introduce the empirical variogram $\hat{\Gamma}^{(m)}$

$$\hat{\Gamma}^{(m)}_{ij} = \text{Var} \left( \log \left( 1 - \tilde{F}_i(X_{ti}) \right) - \log \left( 1 - \tilde{F}_j(X_{tj}) \right) : \tilde{F}_m(X_{tm}) \geq 1 - k/n \right),$$

where $\tilde{F}$ is the empirical distribution function.

- can be made independent of $m \in V$ by considering

$$\hat{\Gamma}_{ij} = \frac{1}{n} \sum_{m \in V} \hat{\Gamma}^{(m)}_{ij},$$

- which is a consistent estimator of $\Gamma$ under some conditions.
**Graphs**

**Connected graph:** a graph in which there is a path between every pair of vertices.

**Decomposable graph:** a graph in which all cycles of length \( \geq 4 \) have a chord.

**Block graph:** a decomposable graph in which all separators are of size 1.

**Tree (graph):** a connected graph without cycles.
Matrix completion on block graphs and trees

- Let $G = (V, E)$ be a connected block graph (or a tree)
- Let $\hat{\Gamma}$ be a partial variogram, specified on the edges of $G$
- Then we have the additivity property

$$
\Gamma_{st} = \sum_{(i,j) \in \text{ph}(s,t)} \Gamma_{ij},
$$

where $\text{ph}(s, t)$ denotes the shortest path between $s$ and $t$ and the resulting $\Gamma$ has graphical structure $G$.

(Engelke and Hitz (2020); Asenova et al. (2021); Engelke and Volgushev (2020); Asenova and Segers (2021))
Matrix completion on decomposable graphs

- Let $G = (V, E)$ be a connected, decomposable graph
- Let $\hat{\Gamma}$ be a partial variogram, specified on the edges of $G$

**Proposition**

There exists a unique completion $\Gamma$ such that

$$
\Gamma_{ij} = \hat{\Gamma}_{ij}, \forall (i, j) \in \overline{E}, \\
\Theta_{ij} = 0, \forall (i, j) \notin \overline{E}.
$$

- This completion can be computed explicitly.
- The mapping $\hat{\Gamma} \mapsto \Gamma$ is continuous.
Matrix completion on general connected graphs

- Let $G = (V, E)$ be a connected graph
- Let $\hat{\Gamma}$ be a partial variogram, specified on the edges of $G$

**Proposition**

If there exists any valid completion of $\hat{\Gamma}$ (say $\tilde{\Gamma}$), then there exists a unique $\Gamma$ such that

$$\Gamma_{ij} = \hat{\Gamma}_{ij}, \forall (i,j) \in \overline{E},$$

$$\Theta_{ij} = 0, \forall (i,j) \notin \overline{E}.$$  

- This completion can be computed as the limit of a convergent sequence of matrices, starting with $\tilde{\Gamma}$.
- The mapping $\hat{\Gamma} \mapsto \Gamma$ is continuous.
- Hüsler–Reiss graphical model on known graph \( G = (V, E) \)
- Parametrized by unknown variogram \( \Gamma \)
- \( \hat{\Gamma}_n \) sequence of consistent estimators for \( \Gamma_{ij}, (ij) \in E \), satisfying \( \text{diag} \hat{\Gamma}_n \equiv 0 \)

**Proposition**

*For \( \hat{\Gamma} \) as above,*

- *with probability going to 1, there exists a graphical completion \( \hat{\Gamma}_n^G \in \mathcal{D} \) of \( \hat{\Gamma}_n \).*
- *This completion is consistent for all \( (i, j) \in V \times V \), i.e.,

\[
P\left( \max_{(i,j)\in V\times V} \left| \hat{\Gamma}^G_{ij} - \Gamma_{ij} \right| < \varepsilon \right) \to 1, \quad n \to \infty.\]
Estimation strategies on sparse graphs

- **Setting:**
  - Known, connected graph $G = (V, E)$ with clique set $C$
  - Unknown variogram matrix $\Gamma$
  - Samples from a Hüsler–Reiss graphical model parametrized by $G$, $\Gamma$
  - Consistent estimator $\hat{\Gamma}_{S \times S}$ for $S \subseteq V$ available

- **To be estimated:**
  - Unknown variogram matrix $\Gamma$

Estimation strategies:

```
"Full":
\begin{pmatrix}
0 & 1 & 6 & 9 & 10 \\
1 & 0 & 3 & 6 & 7 \\
6 & 3 & 0 & 3 & 4 \\
9 & 6 & 3 & 0 & 3 \\
10 & 7 & 4 & 3 & 0
\end{pmatrix}
```

```
"Graphical":
\begin{pmatrix}
0 & 1 & 6 & ? & ? \\
1 & 0 & 3 & ? & ? \\
? & ? & 3 & 0 & 3 \\
? & ? & 4 & 3 & 0
\end{pmatrix}
```

```
"Clique-wise":
\begin{pmatrix}
0 & 1 & 6 & ? & ? \\
1 & 0 & 3 & ? & ? \\
6 & 3 & 0 & 3 & 4 \\
? & ? & 3 & 0 & 3 \\
? & ? & 4 & 3 & 0
\end{pmatrix}
```
Simulation study – Setting

- Graph $G = (V, E)$ with $|V| = 10$, $|E| = 18$.
- Sample size $n = 200$
- Parameter matrix $\Theta$:

$$
\begin{pmatrix}
24.15 & -3.95 & 2.45 & -10.61 & -1.42 & -4.85 & -4.51 & -1.25 & 0 & 0 \\
-3.95 & 77.85 & 0 & 0 & 0 & 0 & 0 & 0 & 7.13 & -81.03 \\
2.45 & 0 & 22.58 & -15.78 & -8.16 & -1.09 & 0 & 0 & 0 & 0 \\
-10.61 & 0 & -15.78 & 27.10 & -2.18 & 1.69 & 4.32 & -4.53 & 0 & 0 \\
-1.42 & 0 & -8.16 & -2.18 & 11.76 & 0 & 0 & 0 & 0 & 0 \\
-4.85 & 0 & -1.09 & 1.69 & 0 & 4.25 & 0 & 0 & 0 & 0 \\
-4.51 & 0 & 0 & 4.32 & 0 & 0 & 3.97 & -3.77 & 0 & 0 \\
-1.25 & 0 & 0 & -4.53 & 0 & 0 & -3.77 & 9.55 & 0 & 0 \\
0 & 7.13 & 0 & 0 & 0 & 0 & 0 & 0 & 40.63 & -47.76 \\
0 & -81.03 & 0 & 0 & 0 & 0 & 0 & 0 & -47.76 & 128.79
\end{pmatrix}
$$
Simulation Study – Results

*n = 200, d = 10*

### Method
- Full variogram
- Clique-wise variogram
- Full MLE
- Graphical MLE
- Clique-wise MLE

### Table

<table>
<thead>
<tr>
<th>Method</th>
<th>Time</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full variogram</td>
<td>1.30E-03</td>
<td>4.87E-03</td>
</tr>
<tr>
<td>Clique-wise variogram</td>
<td>9.40E-02</td>
<td>4.51E-03</td>
</tr>
<tr>
<td>Clique-wise MLE</td>
<td>4.97E+00</td>
<td>2.07E-03</td>
</tr>
<tr>
<td>Graphical MLE</td>
<td>3.57E+02</td>
<td>2.19E-03</td>
</tr>
<tr>
<td>Full MLE</td>
<td>1.24E+03</td>
<td>2.09E-03</td>
</tr>
</tbody>
</table>
Colored models

- Number of edges in the graph/cliques might still be very large
- Desirable to further reduce the dimensionality of the problem
- Röttger et al. (2023) suggest colored graphical models:
  - Edge coloring: $\lambda : E \to \{1, \ldots, r\}$
  - RCON: $\Theta_{ij} = \Theta_{kl}$ if $\lambda((ij)) = \lambda((kl))$
  - RVAR: $\Gamma_{ij} = \Gamma_{kl}$ if $\lambda((ij)) = \lambda((kl))$
- Inference:
  - Determine $\lambda$ through a clustering step
  - Determine $r$ through cross-validation

\[
\Theta = \begin{pmatrix}
5 & -1 & -2 & -2 \\
-1 & 2 & -1 & 0 \\
-2 & -1 & 4 & -1 \\
-2 & 0 & -1 & 3
\end{pmatrix}
\]

\[
\Gamma = \begin{pmatrix}
0 & 2 & 4 & 4 \\
2 & 0 & 2 & 3 \\
4 & 2 & 0 & 2 \\
4 & 3 & 2 & 0
\end{pmatrix}
\]
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

**Inference for Hüsler–Reiss graphical models**

- Inference on known graph structures
  - Structure estimation

Application
Tree graphs

- Recall the additivity property (for trees/block graphs):

\[ \Gamma_{st} = \sum_{(i,j) \in \text{ph}(s,t)} \Gamma_{ij}, \]

- Similar result for extremal correlation \( \chi \)

\[ \chi_{ij} \leq \chi_{st} \quad \forall (ij) \in \text{ph}(s,t) \]

- Minimum spanning tree (EMST):

\[ T_{mst} = \text{argmin}_{(V,E) \in \mathcal{T}} \sum_{(ij) \in E} \rho_{ij}, \]

with \( \rho_{ij} = \Gamma_{ij} \) or \( \rho_{ij} = -\log \chi_{ij} \)

- Also works for non-Hüsler–Reiss tree models (Engelke and Volgushev (2020))

\[ \Gamma = \begin{pmatrix} 0 & 0.75 & 0.25 & 1.25 \\ 0.75 & 0 & 0.5 & 1.5 \\ 0.25 & 0.5 & 0 & 1 \\ 1.25 & 1.5 & 1 & 0 \end{pmatrix} \]

\[ \chi = \begin{pmatrix} 1 & 0.67 & 0.80 & 0.58 \\ 0.67 & 1 & 0.72 & 0.54 \\ 0.80 & 0.72 & 1 & 0.62 \\ 0.58 & 0.54 & 0.62 & 1 \end{pmatrix} \]
“Ideal extremal graphical lasso”

- Normal distribution:
  \[
  f(x) \propto \sqrt{\det \Theta} \exp((x - \mu)^\top \Theta (x - \mu))
  \]
  \[(ij) \notin E \Rightarrow \Theta_{ij} = 0\]

- Gaussian graphical lasso:
  \[\Theta_{\text{glasso}} = \arg\min_{\Theta \in S^+} -\ell(\Theta, \hat{\Sigma}) + \lambda \|\Theta\|_1\]

- Efficiently estimates structure and parameters at once!

- Hüsler–Reiss distribution:
  \[
  \lambda(y; \Theta) \propto c_\Theta \sqrt{\det \Theta} \exp((y - \mu_\Theta)^\top \Theta (y - \mu_\Theta))
  \]
  \[(ij) \notin E \Rightarrow \Theta_{ij} = 0\]
  \[Y^{(1)} \overset{d}{=} W + E\mathbf{1}\]

- “Extremal graphical lasso”?
  \[\Theta_{\text{eglasso}} = \arg\min_{\Theta \in S_1^+} -\ell(\Theta, \hat{\Gamma}) + \lambda \|\Theta\|_{1,\text{off}}\]

- Does not work! Ying et al. (2020, 2021)
Majority voting

- Engelke et al. (2022b) consider the matrices

\[ \Theta^{(1)}, \ldots, \Theta^{(d)} \in \mathbb{R}^{(d-1) \times (d-1)} \]

- \( \Theta_{ij}^{(k)} = 0 \iff \Theta_{ij} = 0 \)
- \( \Theta^{(k)} \) is a full rank precision matrix

- Apply graphical lasso for each \( k \), yielding sparse

\[ \widehat{\Theta}^{(1)}, \ldots, \widehat{\Theta}^{(d)} \]

- Obtain \( \widehat{G} = (V, \widehat{E}) \) by a majority vote:

\[ \#\{k \neq i, j \mid \widehat{\Theta}_{ij}^{(k)} \neq 0\} > \frac{d - 2}{2} \implies (ij) \in \widehat{E} \]

- Some sparsistency results, but no parameter estimate!
Score matching

- Alternative to MLE that minimizes the Fisher information distance

\[ J(f) = \int_{\mathbb{R}^d} f_0(x) \| \nabla_x \log f(x) - \nabla_x \log f_0(x) \|^2 dx \]

- Computationally efficient without computing the normalization constant
- Can be $\ell^1$-penalized to obtain sparse graphs
- Lederer and Oesting (2023) apply this approach to Hüsler–Reiss models
  - Optimize over a superset of Hüsler–Reiss parametrizations
    ⇒ might yield invalid $\Theta$ matrices
  - Provide concentration guarantees
Parameter shift

- Wan and Zhou (2023) consider for some $c \in \mathbb{R}$ the full rank matrices

$$\Theta^* = \Theta + c\mathbf{1}\mathbf{1}^\top$$

$$\Sigma^* = (\Theta^*)^{-1} = \Sigma + c^{-1}d^{-2}\mathbf{1}\mathbf{1}^\top$$

- Estimate $\hat{\Sigma}$ from data and set $\hat{\Sigma}^* = \hat{\Sigma} + c^{-1}d^{-2}\mathbf{1}\mathbf{1}^\top$

- Apply graphical lasso with modified penalty term

$$\hat{\Theta}^* = \arg\min_{\Theta^* \in S^+} -\ell(\Theta^*, \hat{\Sigma}^*) + \lambda \|\Theta^* - c\mathbf{1}\mathbf{1}^\top\|_{1,\text{off}}$$

- Obtain $\hat{\Theta} = \hat{\Theta}^* - c\mathbf{1}\mathbf{1}^\top$

- No guarantee that $\hat{\Theta}$ is a valid precision matrix

- But still consistent
- Röttger et al. (2021) introduce extremal multivariate total positivity (EMTP$_2$)
- Requires $Y^{(k)}$ to be MTP$_2$, i.e. their densities satisfy

$$f(x \lor y)f(x \land y) \geq f(x)f(y)$$

- For Hüsler–Reiss models encoded as

$$\Theta_{ij} \leq 0 \quad \forall i \neq j$$

- Constraint typically yields sparse estimates $\hat{\Theta}_+$
- Some asymptotic structural consistency results
Background: Graphical models

Background: Multivariate generalized Pareto distributions

Extremal graphical models

Hüsler–Reiss graphical models

Inference for Hüsler–Reiss graphical models
  - Inference on known graph structures
  - Structure estimation

Application
The data set

- Domestic flight data from the U.S. Bureau of Transportation Statistics\(^1\)
- Airports and airlines with \(\geq 1\%\) marketshare
- Years 2005-2020
- Filtered:
  - Airports with at least 1000 flights per year
  - in the contiguous United States
- Total positive delays aggregated per day
- Resulting in \(n = 5601\) observations of \(d = 118\) airports

The data set – Illustration

- Size proportional to daily flights
- Color indicates shape of univariate generalized Pareto MLE:
  - Cyan $< 0$
  - Black $= 0$
  - Orange $> 0$

→ Normalized marginals from here
Clustering

- Ensuring (strong) extremal dependence within the analyzed set of airports
- Using $k$-medoids clustering
  - Empirical extremal correlation $\hat{\chi}(p = 0.85)$ as (dis)similarity measure
  - $k = 4, \ldots, 7$ yields “reasonable” results
- Clusters show strong geographic proximity (even though no explicit geographic information is used!)

![Histogram of $\hat{\chi}(p)$ values with count distribution over different $\hat{\chi}(p)$ ranges.](chart.png)
Clustering – The “Texas cluster”

- Centered around Houston, Dallas
- 29 Airports
- 83 Pairs with \((\geq\) monthly) connections
- Size proportional to daily flights
Estimation

- Train/test split: before/after 2010
- Structure learning:
  - “Flight graph”
  - Full graph
  - Random graph
  - Minimum spanning tree (“EMST”)
  - Majority voting (“EGLearn”)
  - Parameter shift
  → Use matrix completion with empirical variogram at threshold $p = 0.95$ for parameters
- Joint structure and parameter learning:
  - EMTP$_2$
  - Score matching
- Colored graphical model on the EMTP$_2$ graph
Results – Graph structures

(Flight graph, full variogram, random graph)

(EMST, EGLearn, EMTP$_2$)

(Parameter shift, colored, score matching)
Results – Edge counts and validation likelihoods

<table>
<thead>
<tr>
<th></th>
<th>Edges</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flight Graph</td>
<td>129</td>
<td>-9.23</td>
</tr>
<tr>
<td>Full Variogram</td>
<td>406</td>
<td>-293.68</td>
</tr>
<tr>
<td>Random Graph</td>
<td>100</td>
<td>-713.79</td>
</tr>
<tr>
<td>EMST</td>
<td>28</td>
<td>-6435.20</td>
</tr>
<tr>
<td>EGLearn</td>
<td>101</td>
<td>1123.45</td>
</tr>
<tr>
<td>EMTP₂</td>
<td>126</td>
<td>1212.97</td>
</tr>
<tr>
<td>Parameter Shift</td>
<td>142</td>
<td>1192.02</td>
</tr>
<tr>
<td>Colored Graph</td>
<td>126</td>
<td>1341.17</td>
</tr>
<tr>
<td>Score Matching</td>
<td>173</td>
<td>1144.97</td>
</tr>
</tbody>
</table>

- Hüsler–Reiss likelihoods with standardized margins, computed on the validation set
- Random graph was set to have 100 edges
- EMST must have $d - 1 = 28$ edges
Results – Implied $\chi$

(Flight graph, full variogram, random graph)

(EMST, EGLearn, EMTP$_2$)

(Parameter shift, colored, score matching)
Results – Hyperparameters

Validation log-likelihoods for different hyperparameters:

Convergence of entries in $\Theta$ to zero for penalized methods:

(EGLearn, parameter shift, score matching, [colored])
Discussion points/open questions

- Stochastic interpretation of extremal conditional independence?
- Relation to pre-asymptotic graphical models, domain of attraction?
  - Do pre-asymptotic graphical models converge to extremal graphical models?
- Focus on Hüsler–Reiss distribution
  - Comparable to focus on Gaussian graphical models?
  - Corresponding limit theorem?
  - Do results translate to general covariance/variogram statements?
  - “Good enough approximation” of other distributions?
- Disconnected graphical models
  - “Asymptotic independence” very different from normal independence!
  - Do methods for connected models transfer to disconnected ones?


Thank you!