Probabilistic graphical models and their application to extreme value statistics

Frank Röttger (University of Twente)

44th Finnish Summer School on Probability and Statistics
May 25-29, 2026

UNIVERSITY
OF TWENTE.

Structure of the Mini-Course

- **Part 1:** Undirected graphical models
- **Part 2:** Directed graphical models
- **Part 3:** Graphical models in extremes

Intended learning goals

- Get an introduction to the vast area of graphical models.
- In particular, learn about graphical models in extremes.

Part 3: Graphical models in extremes

Table of contents

Motivation

Motivation

- In many multivariate applications, the most catastrophic outcomes are due to the concurrence of several rare events.

Motivation

- In many multivariate applications, the most catastrophic outcomes are due to the concurrence of several rare events.
- However, we typically only have very few extreme observations!

Motivation

- In many multivariate applications, the most catastrophic outcomes are due to the concurrence of several rare events.
- However, we typically only have very few extreme observations!

$\Rightarrow$ We are interested in models that allow for notions of conditional independence, graphical models and sparsity in extremes.

Motivation

- In many multivariate applications, the most catastrophic outcomes are due to the concurrence of several rare events.
- However, we typically only have very few extreme observations!

$\Rightarrow$ We are interested in models that allow for notions of conditional independence, graphical models and sparsity in extremes.

- Example: Danube river data of Asadi et al. (2015).

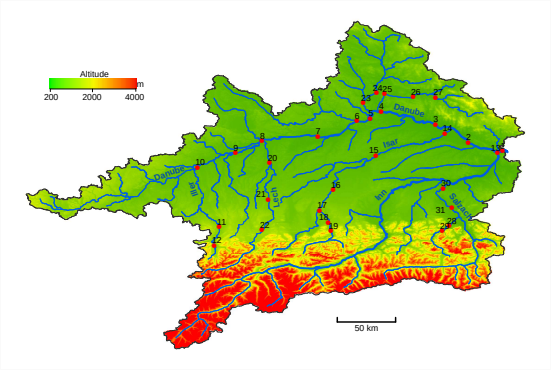


Figure 1: Topographic map of the upper Danube basin.

A closer look at the Danube river data

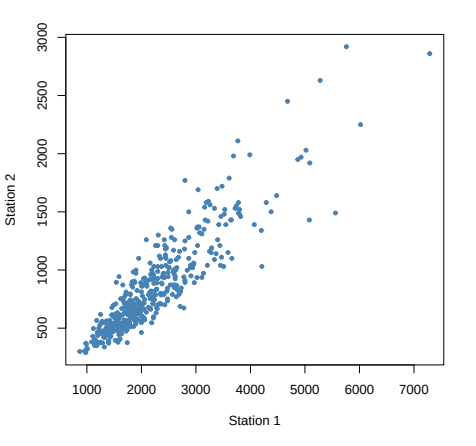


Figure 2: Scatter plot of data for Stations 1 and 2.

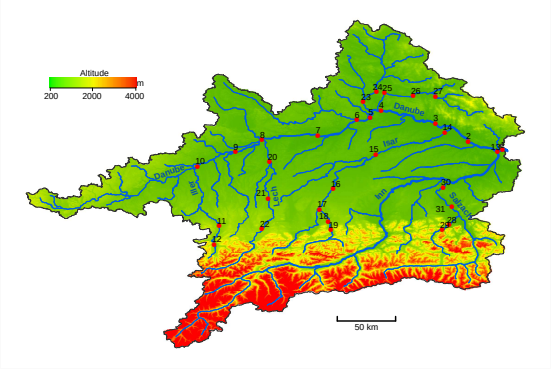


Figure 1: Topographic map of the upper Danube basin.

A closer look at the Danube river data

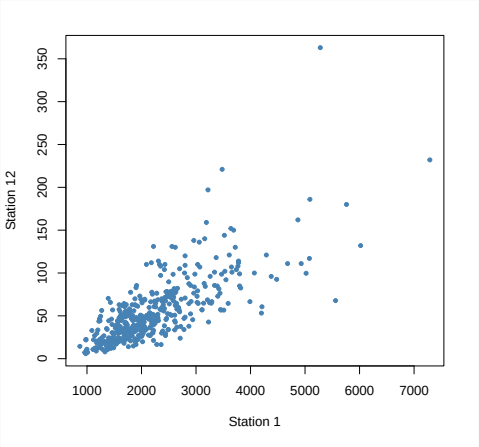


Figure 3: Scatter plot of data for Stations 1 and 12.

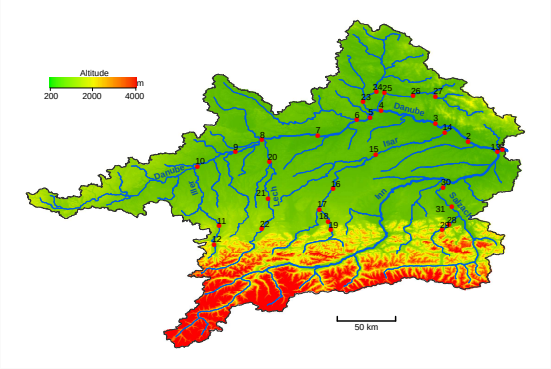


Figure 1: Topographic map of the upper Danube basin.

A closer look at the Danube river data

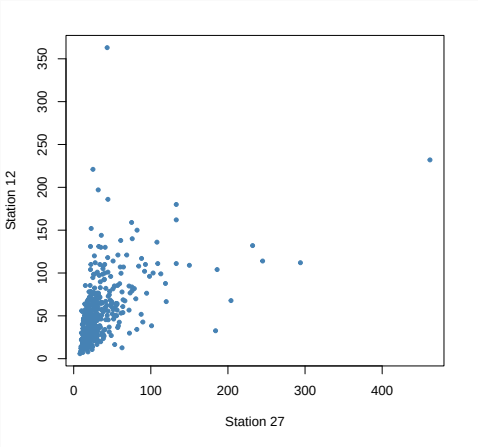


Figure 4: Scatter plot of data for Stations 27 and 12.

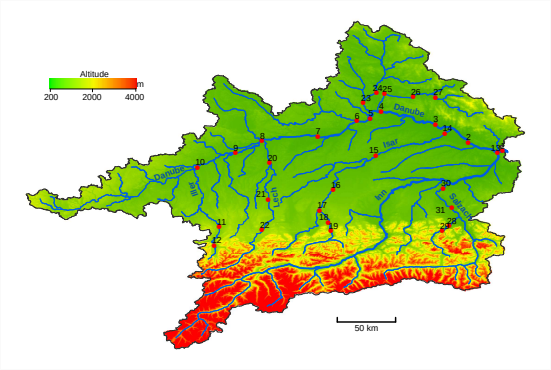


Figure 1: Topographic map of the upper Danube basin.

Marginal rescaling and threshold exceedances

Simplifying dependence modeling: Marginal rescaling

- We only want to model the dependence structure of the data-generating process X^*, therefore we standardize the margins.

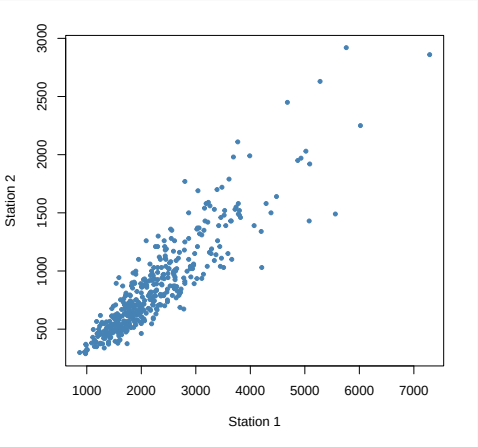


Figure 2: Scatter plot of data for Stations 1 and 2.

Simplifying dependence modeling: Marginal rescaling

- We only want to model the dependence structure of the data-generating process X^*, therefore we standardize the margins.
- If X_i^* has a continuous cdf F_i, then

$$X_i = -\log(1 - F_i(X_i^*))$$

is standard exponential.

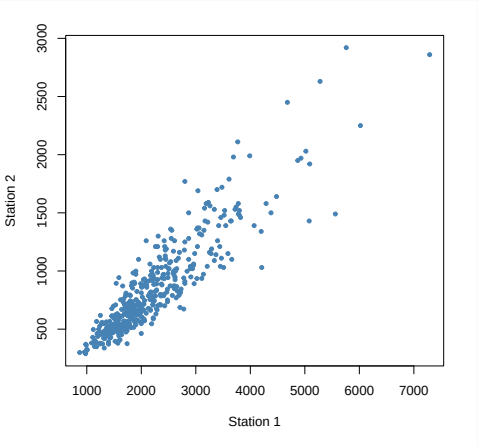


Figure 2: Scatter plot of data for Stations 1 and 2.

Simplifying dependence modeling: Marginal rescaling

- We only want to model the dependence structure of the data-generating process X^*, therefore we standardize the margins.
- If X_i^* has a continuous cdf F_i, then

$$X_i = -\log(1 - F_i(X_i^*))$$

  is standard exponential.
- In practice, we estimate F_i via the empirical distribution function.

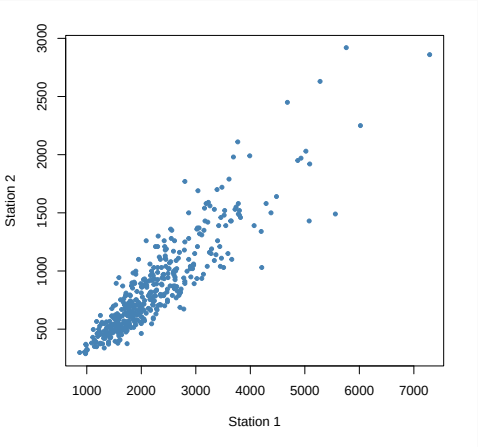


Figure 2: Scatter plot of data for Stations 1 and 2.

Simplifying dependence modeling: Marginal rescaling

- We only want to model the dependence structure of the data-generating process X^*, therefore we standardize the margins.
- If X_i^* has a continuous cdf F_i, then

$$X_i = -\log(1 - F_i(X_i^*))$$

  is standard exponential.
- In practice, we estimate F_i via the empirical distribution function.

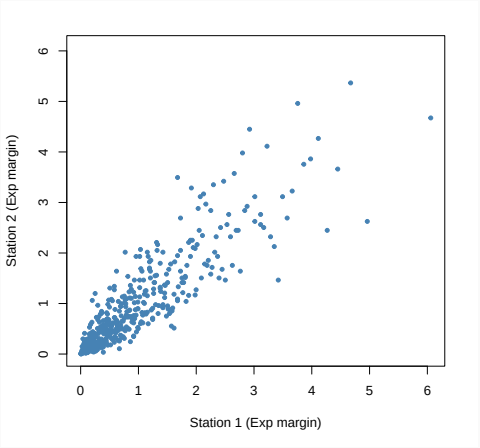


Figure 5: Standardized scatter plot of data for Stations 1 and 2.

When do we consider an observation as extreme?

- In the approach of threshold exceedances, we consider a multivariate observation as large/ extreme if it exceeds a large quantile in at least one margin.

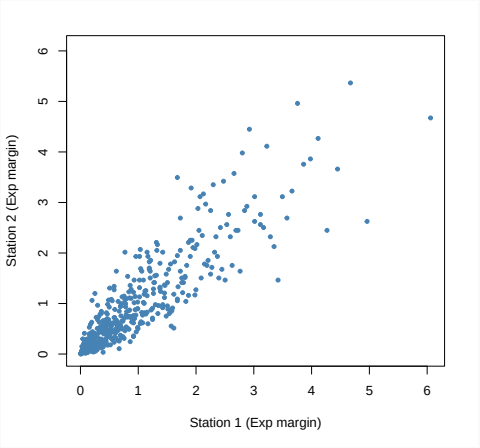


Figure 5: Standardized scatter plot of data for Stations 1 and 2

When do we consider an observation as extreme?

- In the approach of threshold exceedances, we consider a multivariate observation as large/ extreme if it exceeds a large quantile in at least one margin.
- For our standardized data, we can choose a large quantile of the standard exponential distribution.

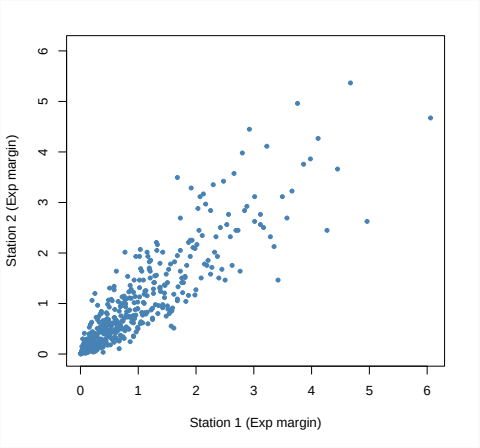


Figure 5: Standardized scatter plot of data for Stations 1 and 2

When do we consider an observation as extreme?

- In the approach of threshold exceedances, we consider a multivariate observation as large/ extreme if it exceeds a large quantile in at least one margin.
- For our standardized data, we can choose a large quantile of the standard exponential distribution.

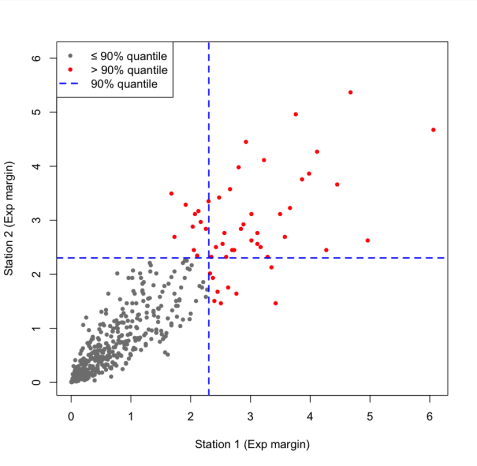


Figure 6: Standardized scatter plot of data for Stations 1 and 2 with highlighted threshold exceedances.

When do we consider an observation as extreme?

- In the approach of threshold exceedances, we consider a multivariate observation as large/ extreme if it exceeds a large quantile in at least one margin.
- For our standardized data, we can choose a large quantile of the standard exponential distribution.
- Finally, we subtract the quantile from each dimension of exceeding observations.

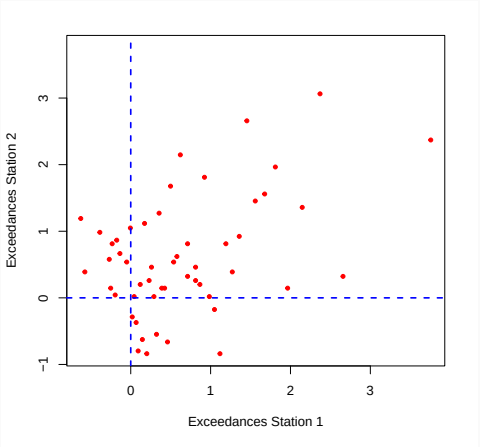


Figure 7: Standardized threshold exceedances

Multivariate Pareto distributions

Multivariate (generalized) Pareto distributions

Let X be a random vector with exponential margins,
i.e. each X_i follows a standard exponential distribution.

Multivariate (generalized) Pareto distributions

Let X be a random vector with exponential margins,
i.e. each X_i follows a standard exponential distribution.

Threshold exceedances

The limit distribution for threshold exceedances

$$\mathbb{P}(Y \leq \mathrm{y}) = \lim_{u \to \infty} \mathbb{P}\left(X - u\mathbf{1} \leq \mathrm{y} \mid X \not\leq u\right)$$

is a multivariate (generalized) Pareto distribution.

Multivariate (generalized) Pareto distributions

Let X be a random vector with exponential margins, i.e. each X_i follows a standard exponential distribution.

Threshold exceedances

The limit distribution for threshold exceedances

$$\mathbb{P}(Y \leq \mathrm{y}) = \lim_{u \to \infty} \mathbb{P}(X - u\mathbf{1} \leq \mathrm{y} \mid X \not\leq u)$$

is a multivariate (generalized) Pareto distribution.

$\Rightarrow$ Y lives on an L-shaped space $\mathcal{L} = \{y \in \mathbb{R}^d : y \not\leq 0\}$.

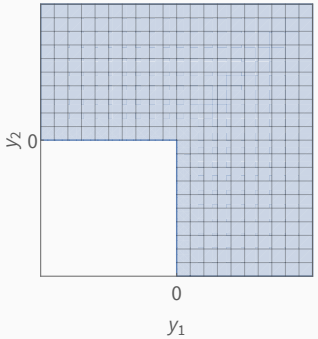


Figure 8: Support of Y for $d = 2$.

The exponent measure

- Equivalent to the existence of the threshold exceedances limit is multivariate regular variation:

$$\lim_{t\to\infty} t\mathbb{P}(X - t\mathbf{1} \in A) = \Lambda(A),$$

for all Borel subsets $A \subset \mathbb{R}^d$ bounded away from $\{-\infty\}^d$, with zero mass on the boundary.

The exponent measure

- Equivalent to the existence of the threshold exceedances limit is multivariate regular variation:

$$\lim_{t\to\infty} t\mathbb{P}(X - t\mathbf{1} \in A) = \Lambda(A),$$

  for all Borel subsets $A \subset \mathbb{R}^d$ bounded away from $\{-\infty\}^d$, with zero mass on the boundary.
- Λ is a Radon measure on $\mathbb{R}^d$ that we call the exponent measure.

The exponent measure

- Equivalent to the existence of the threshold exceedances limit is multivariate regular variation:

$$\lim_{t\to\infty} t\mathbb{P}(X - t\mathbf{1} \in A) = \Lambda(A),$$

  for all Borel subsets $A \subset \mathbb{R}^d$ bounded away from $\{-\infty\}^d$, with zero mass on the boundary.
- Λ is a Radon measure on $\mathbb{R}^d$ that we call the exponent measure.

Homogeneity

Λ is homogeneous:

$$\Lambda(tA) = e^{-t}\Lambda(A)$$

for any Borel set $A \subset \mathbb{R}^d$ and $t \in \mathbb{R}$.

Distribution of Y

Exponent measure density

If Λ is absolutely continuous with respect to the Lebesgue measure, its derivative λ satisfies the same homogeneity as Λ. We call λ the exponent measure density.

Distribution of Y

Exponent measure density

If Λ is absolutely continuous with respect to the Lebesgue measure, its derivative λ satisfies the same homogeneity as Λ. We call λ the exponent measure density.

The exponent measure fully describes the distribution of Y since

$$\mathbb{P}(Y \leq y) = \frac{\Lambda\left([-\infty, y]^c \cap \mathcal{L}\right)}{\Lambda(\mathcal{L})}.$$

Distribution of Y

Exponent measure density

If Λ is absolutely continuous with respect to the Lebesgue measure, its derivative λ satisfies the same homogeneity as Λ. We call λ the exponent measure density.

The exponent measure fully describes the distribution of Y since

$$\mathbb{P}(Y \leq y) = \frac{\Lambda\left([-\infty, y]^c \cap \mathcal{L}\right)}{\Lambda(\mathcal{L})}.$$

Density of Y

The density of Y is then given by $f_Y(y) = \frac{\lambda(y)}{\Lambda(\mathcal{L})}$ for $y \in \mathcal{L}$.

Extremal conditional independence and graphical models

Let Y be multivariate Pareto with support on the L-shaped space $\mathcal{L} = \{y \in \mathbb{R}^d : y \not\leq 0\}$.

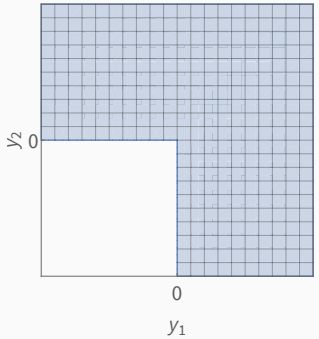


Figure 9: Support of Y for $d = 2$.

Extremal conditional independence

Let Y be multivariate Pareto with support on the L-shaped space $\mathcal{L} = \{y \in \mathbb{R}^d : y \not\leq 0\}$.

- **Objective:** Study (conditional) independence for Y.

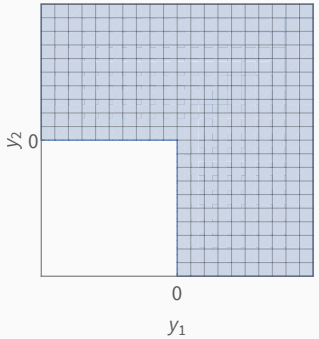


Figure 9: Support of Y for $d = 2$.

Extremal conditional independence

Let Y be multivariate Pareto with support on the L-shaped space $\mathcal{L} = \{y \in \mathbb{R}^d : y \not\leq 0\}$.

- **Objective:** Study (conditional) independence for Y.
- **Problem:** $\mathcal{L}$ is not a product space $\Rightarrow$ no classical notion of (conditional) independence.

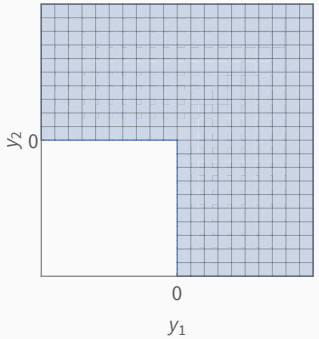


Figure 9: Support of Y for $d = 2$.

Extremal conditional independence

Let Y be multivariate Pareto with support on the L-shaped space $\mathcal{L} = \{y \in \mathbb{R}^d : y \not\leq 0\}$.

- **Objective:** Study (conditional) independence for Y.
- **Problem:** $\mathcal{L}$ is not a product space $\Rightarrow$ no classical notion of (conditional) independence.
- **Idea of Engelke and Hitz (2020):** Control the location of the exceedance $Y_k > 0$.

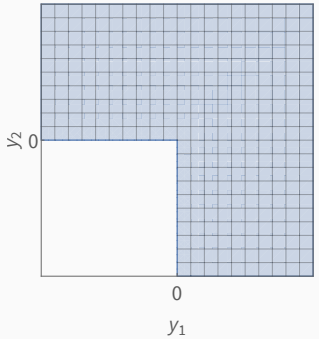


Figure 9: Support of Y for $d = 2$.

Extremal conditional independence

Approach of Engelke and Hitz (2020):

Extremal conditional independence

Approach of Engelke and Hitz (2020):

Let $Y^k = Y \mid \{Y_k > 0\}$, where $k \in V = \{1, \dots, d\}$.

Approach of Engelke and Hitz (2020):

Let $Y^k = Y \mid \{Y_k > 0\}$, where $k \in V = \{1, \ldots, d\}$.

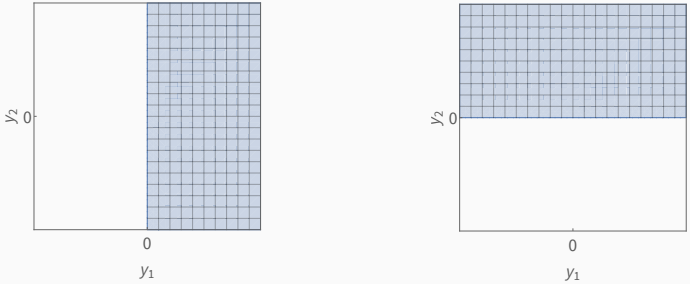


Figure 10: Support of bivariate Y^1 and Y^2

Approach of Engelke and Hitz (2020):

Let $Y^k = Y \mid \{Y_k > 0\}$, where $k \in V = \{1, \dots, d\}$.

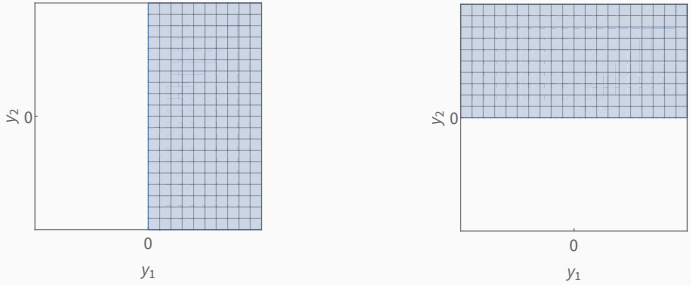


Figure 10: Support of bivariate Y^1 and Y^2

Density of Y^k

For each $k \in V$, Y^k has as probability density the exponent measure density $\lambda(\mathrm{y})$ restricted to its support $\mathcal{L}^k := \{\mathrm{x} \in \mathbb{R}^d : x_k > 0\}$.

Definition

Y_A is extremal conditionally independent of Y_B given Y_C ($Y_A \perp_e Y_B \mid Y_C$) if and only if

$$\forall k \in V : Y_A^k \perp\!\!\!\perp Y_B^k \mid Y_C^k.$$

Definition

Y_A is extremal conditionally independent of Y_B given Y_C ($Y_A \perp_e Y_B \mid Y_C$) if and only if

$$\forall k \in V : Y_A^k \perp\!\!\!\perp Y_B^k \mid Y_C^k.$$

- For any $I \subseteq V$, let $\lambda_I(\mathrm{y}_I) = \int \lambda(\mathrm{y}) d\mathrm{y}_{V \setminus I}$ be the marginal exponent measure density.

Definition

Y_A is extremal conditionally independent of Y_B given Y_C ($Y_A \perp_e Y_B \mid Y_C$) if and only if

$$\forall k \in V : Y_A^k \perp\!\!\!\perp Y_B^k \mid Y_C^k.$$

- For any $I \subseteq V$, let $\lambda_I(\mathrm{y}_I) = \int \lambda(\mathrm{y}) d\mathrm{y}_{V \setminus I}$ be the marginal exponent measure density.

Representation via the exponent measure density

$Y_A \perp_e Y_B \mid Y_C$ if and only if

$$\lambda(\mathrm{y})\lambda_C(\mathrm{y}_C) = \lambda_{A\cup C}(\mathrm{y}_{A\cup C})\lambda_{B\cup C}(\mathrm{y}_{B\cup C})$$

for all $\mathrm{y} \in \mathcal{L}$.

Extremal graphical models

The pairwise Markov property gives rise to extremal graphical models:

Extremal graphical models

The pairwise Markov property gives rise to extremal graphical models:

Undirected extremal graphical model

We call Y an undirected extremal graphical model wrt $G = (V, E)$ when

$$ij \notin E \implies Y_i \perp_e Y_j \mid Y_{V \setminus ij}.$$

The pairwise Markov property gives rise to extremal graphical models:

Undirected extremal graphical model

We call Y an undirected extremal graphical model wrt $G = (V, E)$ when

$$ij \notin E \implies Y_i \perp_e Y_j \mid Y_{V \setminus ij}.$$

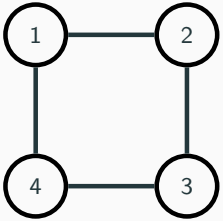


Figure 11: The four-cycle imposes $Y_1 \perp_e Y_3 \mid Y_{24}$ and $Y_2 \perp_e Y_4 \mid Y_{13}$.

The pairwise Markov property gives rise to extremal graphical models:

Undirected extremal graphical model

We call Y an undirected extremal graphical model wrt $G = (V, E)$ when

$$ij \notin E \implies Y_i \perp_e Y_j \mid Y_{V \setminus ij}.$$

Review article:

Engelke, Hentschel, Lalancette, R. (2024). **Graphical models for multivariate extremes.**

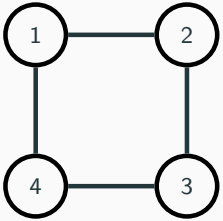


Figure 11: The four-cycle imposes $Y_1 \perp_e Y_3 \mid Y_{24}$ and $Y_2 \perp_e Y_4 \mid Y_{13}$.

An extremal analogue of Gaussians: The Hüsler–Reiss distribution

Definition

A Hüsler–Reiss distribution is a multivariate Pareto with exponent measure density

$$\lambda(\mathrm{y}) = c_\Gamma \exp\left(-\frac{1}{2}\left(\mathrm{y}^T, 1\right)\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1}\begin{pmatrix} \mathrm{y} \\ 1 \end{pmatrix}\right),$$

The Hüsler–Reiss multivariate Pareto distribution

Definition

A Hüsler–Reiss distribution is a multivariate Pareto with exponent measure density

$$\lambda(\mathrm{y}) = c_\Gamma \exp\left(-\frac{1}{2}\left(\mathrm{y}^T, 1\right)\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1}\begin{pmatrix} \mathrm{y} \\ 1 \end{pmatrix}\right),$$

where the parameter Γ is a variogram matrix, i.e.

$$\Gamma_{ij} = \mathrm{Var}(Y_i^k - Y_j^k)$$

for any $k \in V$.

The Hüsler–Reiss multivariate Pareto distribution

Definition

A Hüsler–Reiss distribution is a multivariate Pareto with exponent measure density

$$\lambda(\mathrm{y}) = c_\Gamma \exp\left(-\frac{1}{2}\left(\mathrm{y}^T, 1\right)\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1}\begin{pmatrix} \mathrm{y} \\ 1 \end{pmatrix}\right),$$

where the parameter Γ is a variogram matrix, i.e.

$$\Gamma_{ij} = \mathrm{Var}(Y_i^k - Y_j^k)$$

for any $k \in V$.

Example:

$$\begin{pmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12} & 0 & \Gamma_{23} \\ \Gamma_{13} & \Gamma_{23} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$

The Hüsler–Reiss multivariate Pareto distribution

Definition

A Hüsler–Reiss distribution is a multivariate Pareto with exponent measure density

$$\lambda(\mathrm{y}) = c_\Gamma \exp\left(-\frac{1}{2}\left(\mathrm{y}^T, 1\right)\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1}\begin{pmatrix} \mathrm{y} \\ 1 \end{pmatrix}\right),$$

where the parameter Γ is a variogram matrix, i.e.

$$\Gamma_{ij} = \mathrm{Var}(Y_i^k - Y_j^k)$$

for any $k \in V$.

Any valid Γ is a Euclidean distance matrix.

Example:

$$\begin{pmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12} & 0 & \Gamma_{23} \\ \Gamma_{13} & \Gamma_{23} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$

The Hüsler–Reiss multivariate Pareto distribution

Definition

A Hüsler–Reiss distribution is a multivariate Pareto with exponent measure density

$$\lambda(\mathrm{y}) = c_\Gamma \exp\left(-\frac{1}{2}\left(\mathrm{y}^T, 1\right)\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1}\begin{pmatrix} \mathrm{y} \\ 1 \end{pmatrix}\right),$$

where the parameter Γ is a variogram matrix, i.e.

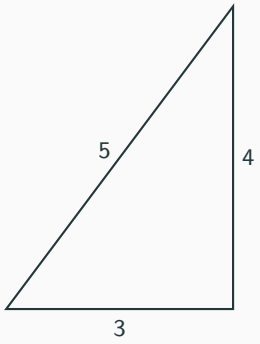
$$\Gamma_{ij} = \mathrm{Var}(Y_i^k - Y_j^k)$$

for any $k \in V$.

Any valid Γ is a Euclidean distance matrix.

Example:

$$\begin{pmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12} & 0 & \Gamma_{23} \\ \Gamma_{13} & \Gamma_{23} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$



Parametric encoding of extremal conditional independence

Theorem

Let Y be a Hüsler–Reiss random vector with variogram Γ and let $A, B, C \subset [d]$ be disjoint. Then

$$Y_A \perp_e Y_B \mid Y_C \iff \operatorname{rank} \begin{pmatrix} -\frac{1}{2}\Gamma_{A\cup C, B\cup C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = |C| + 1.$$

Parametric encoding of extremal conditional independence

Theorem

Let Y be a Hüsler–Reiss random vector with variogram Γ and let $A, B, C \subset [d]$ be disjoint. Then

$$Y_A \perp_e Y_B \mid Y_C \iff \operatorname{rank} \begin{pmatrix} -\frac{1}{2}\Gamma_{A\cup C, B\cup C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = |C| + 1.$$

In particular, for singletons $A = \{i\}$, $B = \{j\}$ this implies

$$Y_i \perp_e Y_j \mid Y_C \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{\{i\}\cup C, \{j\}\cup C} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0,$$

i.e. extremal conditional independence is equivalent to vanishing subdeterminants.

Fiedler–Bapat identity (Devriendt (2022))

Let Γ be a Euclidean distance matrix. Then

$$\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta & p \\ p^T & \sigma^2 \end{pmatrix},$$

Fiedler–Bapat identity (Devriendt (2022))

Let Γ be a Euclidean distance matrix. Then

$$\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta & p \\ p^T & \sigma^2 \end{pmatrix},$$

where Θ is a signed graph Laplacian, i.e. Θ is positive semidefinite, symmetric and $\Theta\mathbf{1} = \mathbf{0}$,

Fiedler–Bapat identity (Devriendt (2022))

Let Γ be a Euclidean distance matrix. Then

$$\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta & p \\ p^T & \sigma^2 \end{pmatrix},$$

where Θ is a signed graph Laplacian, i.e. Θ is positive semidefinite, symmetric and $\Theta\mathbf{1} = \mathbf{0}$, $\sigma^2 = (2\mathbf{1}^T\Gamma^{-1}\mathbf{1})^{-1}$, and $p = 2\sigma^2\Gamma^{-1}\mathbf{1}$.

Fiedler–Bapat identity (Devriendt (2022))

Let Γ be a Euclidean distance matrix. Then

$$\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta & p \\ p^T & \sigma^2 \end{pmatrix},$$

where Θ is a signed graph Laplacian, i.e. Θ is positive semidefinite, symmetric and $\Theta\mathbf{1} = \mathbf{0}$, $\sigma^2 = (2\mathbf{1}^T\Gamma^{-1}\mathbf{1})^{-1}$, and $p = 2\sigma^2\Gamma^{-1}\mathbf{1}$.

This yields that

$$\lambda(\mathrm{y}) = c_\Theta \exp\left(-\frac{1}{2}\left(\mathrm{y}^T, 1\right)\begin{pmatrix} \Theta & p \\ p^T & \sigma^2 \end{pmatrix}\begin{pmatrix} \mathrm{y} \\ 1 \end{pmatrix}\right)$$

Fiedler–Bapat identity (Devriendt (2022))

Let Γ be a Euclidean distance matrix. Then

$$\begin{pmatrix} -\frac{1}{2}\Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta & p \\ p^T & \sigma^2 \end{pmatrix},$$

where Θ is a signed graph Laplacian, i.e. Θ is positive semidefinite, symmetric and $\Theta\mathbf{1} = \mathbf{0}$, $\sigma^2 = (2\mathbf{1}^T\Gamma^{-1}\mathbf{1})^{-1}$, and $p = 2\sigma^2\Gamma^{-1}\mathbf{1}$.

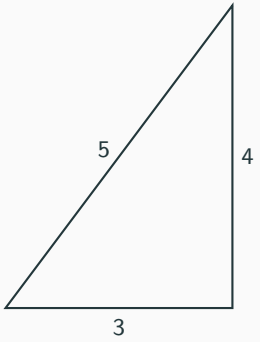
This yields that

$$\begin{aligned} \lambda(\mathrm{y}) &= c_\Theta \exp\left(-\frac{1}{2}\begin{pmatrix}\mathrm{y}^T, 1\end{pmatrix}\begin{pmatrix}\Theta & p \\ p^T & \sigma^2\end{pmatrix}\begin{pmatrix}\mathrm{y} \\ 1\end{pmatrix}\right) \\ &= c_\Theta \exp\left(-\frac{1}{2}\mathrm{y}^T\Theta\mathrm{y} - p^T\mathrm{y} - \frac{1}{2}\sigma^2\right). \end{aligned}$$

The Fiedler–Bapat identity and graph Laplacians: Example

Example:

$$\begin{pmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12} & 0 & \Gamma_{23} \\ \Gamma_{13} & \Gamma_{23} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$



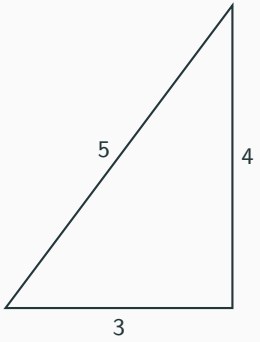
Example

For the example, this yields

$$\begin{pmatrix} 0 & -\frac{9}{2} & -\frac{25}{2} & 1 \\ -\frac{9}{2} & 0 & -\frac{16}{2} & 1 \\ -\frac{25}{2} & -\frac{16}{2} & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{9} & -\frac{1}{9} & 0 & \frac{1}{2} \\ -\frac{1}{9} & \frac{25}{144} & -\frac{1}{16} & 0 \\ 0 & -\frac{1}{16} & \frac{1}{16} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{25}{4} \end{pmatrix}.$$

Example:

$$\begin{pmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12} & 0 & \Gamma_{23} \\ \Gamma_{13} & \Gamma_{23} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$



Example

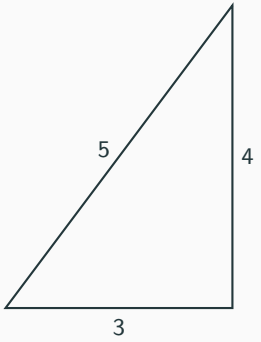
For the example, this yields

$$\begin{pmatrix} 0 & -\frac{9}{2} & -\frac{25}{2} & 1 \\ -\frac{9}{2} & 0 & -\frac{16}{2} & 1 \\ -\frac{25}{2} & -\frac{16}{2} & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{9} & -\frac{1}{9} & 0 & \frac{1}{2} \\ -\frac{1}{9} & \frac{25}{144} & -\frac{1}{16} & 0 \\ 0 & -\frac{1}{16} & \frac{1}{16} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{25}{4} \end{pmatrix}.$$

Figure 12: Path graph

Example:

$$\begin{pmatrix} 0 & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12} & 0 & \Gamma_{23} \\ \Gamma_{13} & \Gamma_{23} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 9 & 25 \\ 9 & 0 & 16 \\ 25 & 16 & 0 \end{pmatrix}$$



Hüsler–Reiss graphical model

- The Fiedler–Bapat identity yields that

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{[d]\setminus\{i\},[d]\setminus\{j\}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0.$$

Hüsler–Reiss graphical model

- The Fiedler–Bapat identity yields that

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{[d]\setminus\{i\},[d]\setminus\{j\}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0.$$

- Using the adjugate matrix expression of the inverse, this simplifies to

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \Theta_{ij} = 0.$$

Hüsler–Reiss graphical model

- The Fiedler–Bapat identity yields that

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{[d]\setminus\{i\},[d]\setminus\{j\}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0.$$

- Using the adjugate matrix expression of the inverse, this simplifies to

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \Theta_{ij} = 0.$$

Hüsler–Reiss graphical model

An undirected Hüsler–Reiss graphical model wrt $G = (V, E)$ imposes

$$ij \notin E \implies \Theta_{ij} = 0.$$

Hüsler–Reiss graphical model

- The Fiedler–Bapat identity yields that

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{[d]\setminus\{i\},[d]\setminus\{j\}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0.$$

- Using the adjugate matrix expression of the inverse, this simplifies to

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \Theta_{ij} = 0.$$

Hüsler–Reiss graphical model

An undirected Hüsler–Reiss graphical model wrt $G = (V, E)$ imposes

$$ij \notin E \implies \Theta_{ij} = 0.$$

- We call Θ the Hüsler–Reiss precision matrix.

Hüsler–Reiss graphical model

- The Fiedler–Bapat identity yields that

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{[d]\setminus\{i\},[d]\setminus\{j\}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0.$$

- Using the adjugate matrix expression of the inverse, this simplifies to

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \Theta_{ij} = 0.$$

Hüsler–Reiss graphical model

An undirected Hüsler–Reiss graphical model wrt $G = (V, E)$ imposes

$$ij \notin E \quad \Longrightarrow \quad \Theta_{ij} = 0.$$

- We call Θ the Hüsler–Reiss precision matrix.

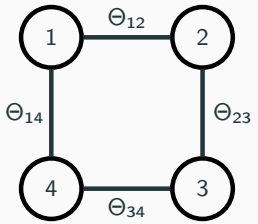


Figure 13: Weighted graph

Hüsler–Reiss graphical model

- The Fiedler–Bapat identity yields that

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \det \begin{pmatrix} -\frac{1}{2}\Gamma_{[d]\setminus\{i\},[d]\setminus\{j\}} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} = 0.$$

- Using the adjugate matrix expression of the inverse, this simplifies to

$$Y_i \perp_e Y_j \mid Y_{[d]\setminus\{i,j\}} \iff \Theta_{ij} = 0.$$

Hüsler–Reiss graphical model

An undirected Hüsler–Reiss graphical model wrt $G = (V, E)$ imposes

$$ij \notin E \quad \Longrightarrow \quad \Theta_{ij} = 0.$$

- We call Θ the Hüsler–Reiss precision matrix.

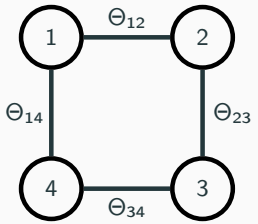


Figure 13: Weighted graph

$$\Theta = \begin{pmatrix} \Theta_{11} & \Theta_{12} & 0 & \Theta_{14} \\ \Theta_{12} & \Theta_{22} & \Theta_{23} & 0 \\ 0 & \Theta_{23} & \Theta_{33} & \Theta_{34} \\ \Theta_{14} & 0 & \Theta_{34} & \Theta_{44} \end{pmatrix}$$

The Hüsler–Reiss elliptope

$\sigma^2 = (2\mathbf{1}^T\Gamma^{-1}\mathbf{1})^{-1}$ is the squared radius of the circumsphere of the simplex corresponding to the Euclidean distance matrix.

$\sigma^2 = (2\mathbf{1}^T\Gamma^{-1}\mathbf{1})^{-1}$ is the squared radius of the circumsphere of the simplex corresponding to the Euclidean distance matrix. If we bound it by 1, we obtain a convex, bounded object:

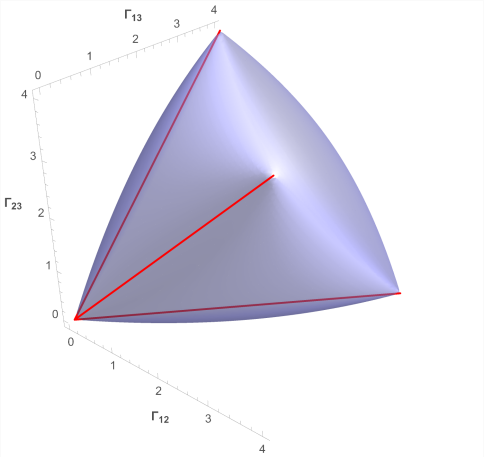


Figure 14: Set of all Euclidean distance matrices with $\sigma^2 \leq 1$.

Conditional independence for Hüsler–Reiss via modular functions

Relative entropy

Let P and Q be two probability measures with probability densities p and q, defined over the same space $\mathcal{X} \subseteq \mathbb{R}^d$ such that P is absolutely continuous with respect to Q. Then, the relative entropy of P with respect to Q is defined as

$$H(P|Q) := \int_{\mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx.$$

Relative entropy

Let P and Q be two probability measures with probability densities p and q, defined over the same space $\mathcal{X} \subseteq \mathbb{R}^d$ such that P is absolutely continuous with respect to Q. Then, the relative entropy of P with respect to Q is defined as

$$H(P|Q) := \int_{\mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx.$$

Multiinformation

The multiinformation m_P of a subset $I \subseteq [d]$ is defined as:

$$m_P(I) := H(P_I | \prod_{i \in I} P_i).$$

Conditional independence via modular functions

Theorem (Studený (2005))

Let X be a random vector with probability measure P and finite multiinformation. Then for any disjoint $A, B, C \subset [d]$ the multiinformation is supermodular

$$m_P(ABC) + m_P(C) \geq m_P(AC) + m_P(BC),$$

with equality (= modular) if and only if the subsets satisfy the conditional independence relation:

$$m_P(ABC) + m_P(C) - m_P(AC) - m_P(BC) = 0 \iff X_A \perp\!\!\!\perp X_B \mid X_C.$$

Conditional independence via modular functions

Theorem (Studený (2005))

Let X be a random vector with probability measure P and finite multiinformation. Then for any disjoint $A, B, C \subset [d]$ the multiinformation is supermodular

$$m_P(ABC) + m_P(C) \geq m_P(AC) + m_P(BC),$$

with equality (= modular) if and only if the subsets satisfy the conditional independence relation:

$$m_P(ABC) + m_P(C) - m_P(AC) - m_P(BC) = 0 \iff X_A \perp\!\!\!\perp X_B \mid X_C.$$

Example: Gaussian

Let X be a d-variate Gaussian distribution with correlation matrix R, the multiinformation function has the explicit expression

$$m_P(I) = -\frac{1}{2}\log(\det(R_{I,I})),$$

for any $I \subseteq [d]$. Thus, for any disjoint $A, B, C \subseteq [d]$ we obtain that $X_A \perp\!\!\!\perp X_B \mid X_C$ is equivalent to

$$-\frac{1}{2}\log(\det(R_{ABC,ABC})) - \frac{1}{2}\log(\det(R_{C,C})) + \frac{1}{2}\log(\det(R_{AC,AC})) + \frac{1}{2}\log(\det(R_{BC,BC})) = 0.$$

Conditional independence for Hüsler–Reiss via modular functions

Definition

Let Y be a Hüsler–Reiss vector with variogram Γ. For $\emptyset \neq I \subseteq [d]$, we define $m^{HR}(I)$ as

$$m^{HR}(I) := -\frac{1}{2} \log \det \begin{pmatrix} -\Gamma_{I,I}/2 & \mathbf{1} \\ -\mathbf{1}^\top & 0 \end{pmatrix},$$

and we let $m^{HR}(\emptyset) := 0$.

Conditional independence for Hüsler–Reiss via modular functions

Definition

Let Y be a Hüsler–Reiss vector with variogram Γ. For $\emptyset \neq I \subseteq [d]$, we define $m^{HR}(I)$ as

$$m^{HR}(I) := -\frac{1}{2} \log \det \begin{pmatrix} -\Gamma_{I,I}/2 & \mathbf{1} \\ -\mathbf{1}^\top & 0 \end{pmatrix},$$

and we let $m^{HR}(\emptyset) := 0$.

Theorem

Let Y be a Hüsler–Reiss vector. Then for any nonempty disjoint $A, B, C \subset [d]$, the function m^{HR} is supermodular

$$m^{HR}(ABC) + m^{HR}(C) \geq m^{HR}(AC) + m^{HR}(BC),$$

with equality (= modular) if and only if the subsets satisfy the extremal conditional independence relation

$$m^{HR}(ABC) + m^{HR}(C) = m^{HR}(AC) + m^{HR}(BC) \iff Y_A \perp_e Y_B \mid Y_C.$$

Modularity of σ^2

The Fiedler–Bapat identity of the submatrix $\Gamma_{I,I}$ gives

$$\begin{pmatrix} -\Gamma_{I,I}/2 & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta(I) & p(I) \\ p(I)^T & \sigma^2(I) \end{pmatrix}.$$

Modularity of σ^2

The Fiedler–Bapat identity of the submatrix $\Gamma_{I,I}$ gives

$$\begin{pmatrix} -\Gamma_{I,I}/2 & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Theta(I) & p(I) \\ p(I)^T & \sigma^2(I) \end{pmatrix}.$$

Theorem

Let Y be a Hüsler–Reiss vector that satisfies $\Theta_{ij} \leq 0$ for all $i \neq j$ and $p \geq 0$. Then for any nonempty disjoint $A, B, C \subset [d]$, the function σ^2 is submodular

$$\sigma^2(ABC) + \sigma^2(C) \leq \sigma^2(AC) + \sigma^2(BC),$$

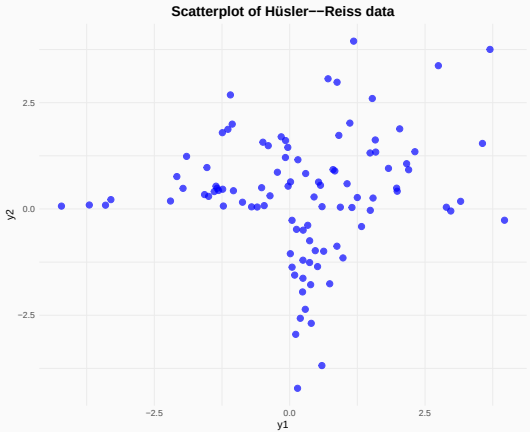
with equality (= modular) if and only if the subsets satisfy the extremal conditional independence relation:

$$\sigma^2(ABC) + \sigma^2(C) = \sigma^2(AC) + \sigma^2(BC) \iff Y_A \perp_e Y_B \mid Y_C.$$

Parameter learning

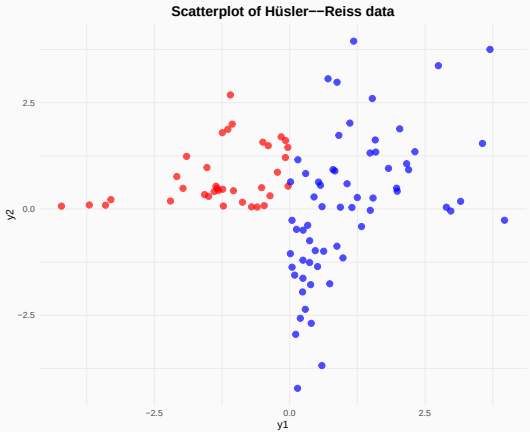
The empirical variogram

- Assume i.i.d. observations $y_1, \ldots, y_n$ from a Hüsler–Reiss distribution.



The empirical variogram

- Assume i.i.d. observations $y_1, \ldots, y_n$ from a Hüsler–Reiss distribution.
- Let $\mathcal{I}_k := \{i \in \{1, \ldots, n\} : y_{ik} > 0\}$.



The empirical variogram

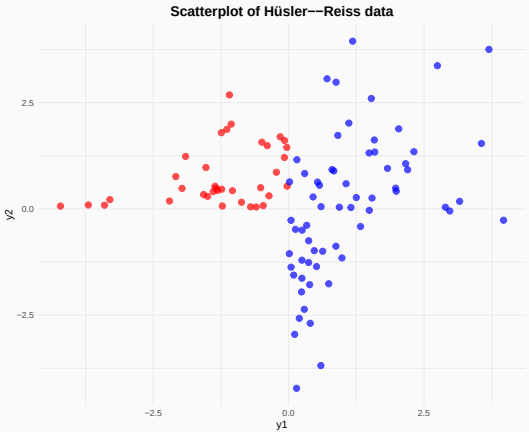
- Assume i.i.d. observations $\mathrm{y}_1, \ldots, \mathrm{y}_n$ from a Hüsler–Reiss distribution.
- Let $\mathcal{I}_k := \{i \in \{1, \ldots, n\} : y_{ik} > 0\}$.

kth sample variogram

Let $\overline{\Gamma}^{(k)}$ be the kth sample variogram with

$$\overline{\Gamma}^{(k)}_{uv} = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \left((y_{iu} - \bar{y}_u) - (y_{iv} - \bar{y}_v) \right)^2 ,$$

where $\bar{\mathrm{y}} = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \mathrm{y}_i$.



The empirical variogram

- Assume i.i.d. observations $\mathrm{y}_1, \ldots, \mathrm{y}_n$ from a Hüsler–Reiss distribution.
- Let $\mathcal{I}_k := \{i \in \{1, \ldots, n\} : y_{ik} > 0\}$.

kth sample variogram

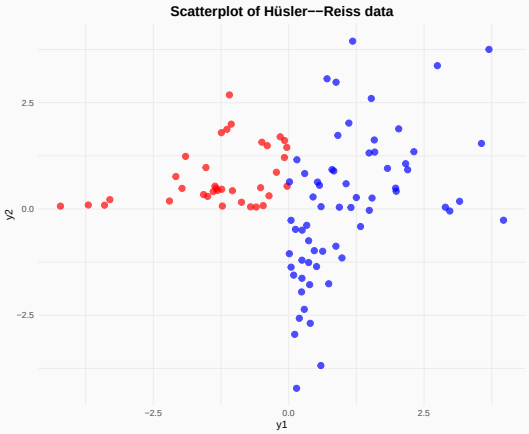
Let $\overline{\Gamma}^{(k)}$ be the kth sample variogram with

$$\overline{\Gamma}^{(k)}_{uv} = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} ((y_{iu} - \bar{y}_u) - (y_{iv} - \bar{y}_v))^2 ,$$

where $\bar{\mathrm{y}} = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \mathrm{y}_i$.

- The empirical variogram is then defined as

$$\overline{\Gamma} = \frac{1}{d} \sum_{k=1}^{d} \overline{\Gamma}^{(k)}.$$



The empirical variogram

- Assume i.i.d. observations $\mathrm{y}_1, \ldots, \mathrm{y}_n$ from a Hüsler–Reiss distribution.
- Let $\mathcal{I}_k := \{i \in \{1, \ldots, n\} : y_{ik} > 0\}$.

kth sample variogram

Let $\overline{\Gamma}^{(k)}$ be the kth sample variogram with

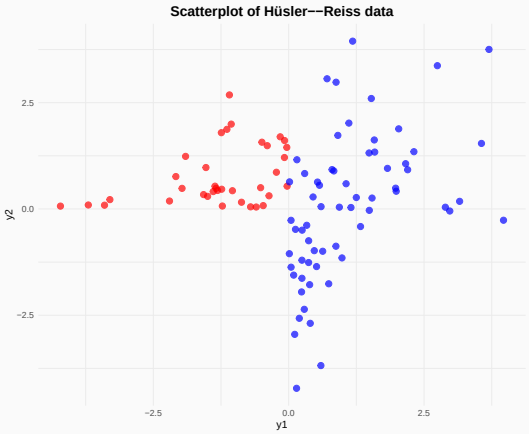
$$\overline{\Gamma}^{(k)}_{uv} = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \left((y_{iu} - \bar{y}_u) - (y_{iv} - \bar{y}_v) \right)^2 ,$$

where $\bar{\mathrm{y}} = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \mathrm{y}_i$.

- The empirical variogram is then defined as

$$\overline{\Gamma} = \frac{1}{d} \sum_{k=1}^{d} \overline{\Gamma}^{(k)}.$$

- The empirical variogram is a consistent estimator of Γ.



Parameter estimation in Hüsler–Reiss graphical models

- **Idea:** Use maximum likelihood estimation to estimate the parameter Γ in Hüsler–Reiss graphical models.

Parameter estimation in Hüsler–Reiss graphical models

- **Idea:** Use maximum likelihood estimation to estimate the parameter Γ in Hüsler–Reiss graphical models.
- **Problem:** Form and support of Hüsler–Reiss log-likelihood.

Parameter estimation in Hüsler–Reiss graphical models

- **Idea:** Use maximum likelihood estimation to estimate the parameter Γ in Hüsler–Reiss graphical models.
- **Problem:** Form and support of Hüsler–Reiss log-likelihood.
- **Alternative approach:** Use a surrogate (Gaussian) log-likelihood.

Laplacian-constrained Gaussian graphical model

Let $\mathcal{H}^{d-1} := \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 0\}$ and $\mathrm{Det}(\cdot)$ the pseudo-determinant.

Let $\mathcal{H}^{d-1} := \{\mathrm{x} \in \mathbb{R}^d : \sum_{i=1}^{d} x_i = 0\}$ and $\mathrm{Det}(\cdot)$ the pseudo-determinant.

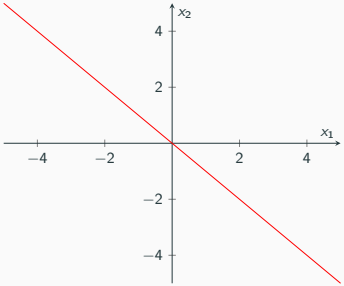


Figure 15: The hyperplane $\mathcal{H}^1$

Let $\mathcal{H}^{d-1} := \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 0\}$ and $\mathrm{Det}(\cdot)$ the pseudo-determinant.

Definition

Let Θ be a p.s.d. signed graph Laplacian and $\mu \in \mathcal{H}^{d-1}$. We call a (degenerate) Gaussian random vector W with density

$$f_W(y) \propto \sqrt{\mathrm{Det}(\Theta)} \exp\left(-\frac{1}{2}(y-\mu)^T \Theta (y-\mu)\right)$$

on $\mathcal{H}^{d-1}$ a Laplacian-constrained Gaussian graphical model (LCGGM).

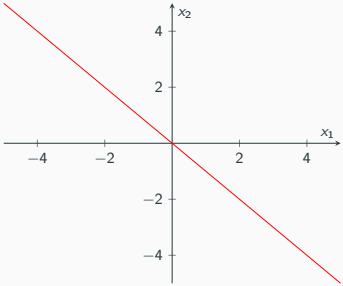


Figure 15: The hyperplane $\mathcal{H}^1$

Let $\mathcal{H}^{d-1} := \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 0\}$ and $\mathrm{Det}(\cdot)$ the pseudo-determinant.

Definition

Let Θ be a p.s.d. signed graph Laplacian and $\mu \in \mathcal{H}^{d-1}$. We call a (degenerate) Gaussian random vector W with density

$$f_W(y) \propto \sqrt{\mathrm{Det}(\Theta)} \exp\left(-\frac{1}{2}(y-\mu)^T \Theta (y-\mu)\right)$$

on $\mathcal{H}^{d-1}$ a Laplacian-constrained Gaussian graphical model (LCGGM).

An LCGGM is an exponential family, its sufficient statistic is a variogram matrix.

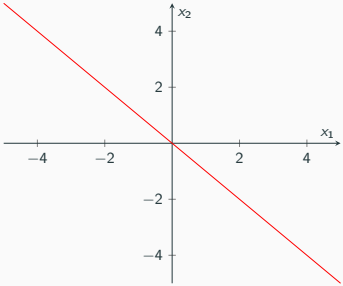


Figure 15: The hyperplane $\mathcal{H}^1$

Decomposition of Hüsler–Reiss exponent measure density

Decomposition of Hüsler–Reiss

For a Hüsler–Reiss vector Y with parameter Γ it is

$$\lambda(y) \propto \exp\left(-\frac{1}{d} y^T 1\right) \cdot f_W(y),$$

where W is an LCGGM with $\mu = \left(I - \frac{1}{d}\mathbf{1}\mathbf{1}^T\right)\left(-\frac{1}{2}\Gamma\right)\mathbf{1}$ and precision matrix Θ.

Decomposition of Hüsler–Reiss exponent measure density

Decomposition of Hüsler–Reiss

For a Hüsler–Reiss vector Y with parameter Γ it is

$$\lambda(y) \propto \exp\left(-\frac{1}{d} y^T 1\right) \cdot f_W(y),$$

where W is an LCGGM with $\mu = \left(I - \frac{1}{d}\mathbf{1}\mathbf{1}^T\right)\left(-\frac{1}{2}\Gamma\right)\mathbf{1}$ and precision matrix Θ.

- The Hüsler–Reiss density has a "radial-angular" representation. Only the angular part depends on Θ.

Decomposition of Hüsler–Reiss

For a Hüsler–Reiss vector Y with parameter Γ it is

$$\lambda(y) \propto \exp\left(-\frac{1}{d} y^T 1\right) \cdot f_W(y),$$

where W is an LCGGM with $\mu = \left(I - \frac{1}{d}\mathbf{1}\mathbf{1}^T\right)\left(-\frac{1}{2}\Gamma\right)\mathbf{1}$ and precision matrix Θ.

- The Hüsler–Reiss density has a "radial-angular" representation. Only the angular part depends on Θ.

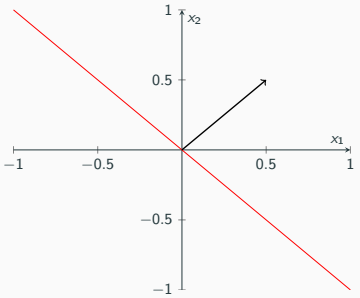


Figure 16: "Radial-angular" decomposition

Decomposition of Hüsler–Reiss

For a Hüsler–Reiss vector Y with parameter Γ it is

$$\lambda(y) \propto \exp\left(-\frac{1}{d} y^T 1\right) \cdot f_W(y),$$

where W is an LCGGM with $\mu = \left(I - \frac{1}{d}\mathbf{1}\mathbf{1}^T\right)\left(-\frac{1}{2}\Gamma\right)\mathbf{1}$ and precision matrix Θ.

- The Hüsler–Reiss density has a "radial-angular" representation. Only the angular part depends on Θ.
- **Problem:** Both μ and Θ depend on Γ.

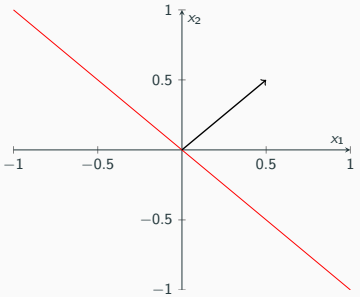


Figure 16: "Radial-angular" decomposition

Decomposition of Hüsler–Reiss

For a Hüsler–Reiss vector Y with parameter Γ it is

$$\lambda(y) \propto \exp\left(-\frac{1}{d} y^T 1\right) \cdot f_W(y),$$

where W is an LCGGM with $\mu = \left(I - \frac{1}{d}\mathbf{1}\mathbf{1}^T\right)\left(-\frac{1}{2}\Gamma\right)\mathbf{1}$ and precision matrix Θ.

- The Hüsler–Reiss density has a "radial-angular" representation. Only the angular part depends on Θ.
- **Problem:** Both μ and Θ depend on Γ.
- **Approach:** Set μ to zero and use empirical variogram as sufficient statistic.

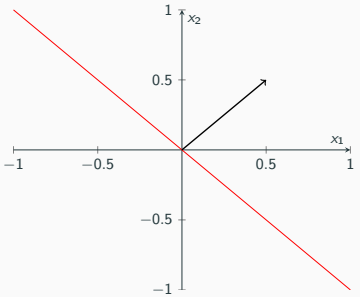


Figure 16: "Radial-angular" decomposition

LCGGM log-likelihood

The LCGGM log-likelihood is

$$\ell(\Theta; \overline{\Gamma}) = \log \mathrm{Det}(\Theta) + \frac{1}{2} \mathrm{tr}(\overline{\Gamma}\Theta),$$

where $\overline{\Gamma}$ is a sample variogram matrix.

LCGGM log-likelihood

The LCGGM log-likelihood is

$$\ell(\Theta; \overline{\Gamma}) = \log \mathrm{Det}(\Theta) + \frac{1}{2} \mathrm{tr}(\overline{\Gamma}\Theta),$$

where $\overline{\Gamma}$ is a sample variogram matrix.

Maximum likelihood estimator (MLE)

The MLE is

$$\widehat{\Theta} = \operatorname{argmax} \ell(\Theta; \overline{\Gamma}), \qquad \text{s.t. } \Theta \text{ is p.s.d.}$$

LCGGM log-likelihood

The LCGGM log-likelihood is

$$\ell(\Theta; \overline{\Gamma}) = \log \mathrm{Det}(\Theta) + \frac{1}{2} \mathrm{tr}(\overline{\Gamma}\Theta),$$

where $\overline{\Gamma}$ is a sample variogram matrix.

Maximum likelihood estimator (MLE)

The MLE is

$$\widehat{\Theta} = \operatorname{argmax} \ell(\Theta; \overline{\Gamma}), \qquad \text{s.t. } \Theta \text{ is p.s.d.}$$

The scores $\nabla_\Theta \ell(\Theta; \overline{\Gamma}) = \Gamma - \overline{\Gamma}$ yield that the unconstrained MLE is $\widehat{\Gamma} = \overline{\Gamma}$.

Surrogate MLE under graphical constraints

Let ℓ_{W} be the log-likelihood of the LCGGM W and let $\overline{\Gamma}$ be an empirical variogram. We call

$$\widehat{\Theta} = \operatorname*{argmax}_{\Theta} \ell(\Theta; \overline{\Gamma}) \qquad \text{s.t. } \Theta_{ij} = 0, \ \ \forall ij \in E.$$

the surrogate MLE for the Hüsler–Reiss graphical model with respect to $G = (V, E)$.

Surrogate MLE under graphical constraints

Let ℓ_{W} be the log-likelihood of the LCGGM W and let $\overline{\Gamma}$ be an empirical variogram. We call

$$\widehat{\Theta} = \operatorname*{argmax}_{\Theta} \ell(\Theta; \overline{\Gamma}) \qquad \text{s.t. } \Theta_{ij} = 0, \ \ \forall ij \in E.$$

the surrogate MLE for the Hüsler–Reiss graphical model with respect to $G = (V, E)$.

This is a matrix completion problem:

$$\begin{aligned} \Gamma_{ij} &= \overline{\Gamma}_{ij}, \qquad \forall (i,j) \in E, \\ \Theta_{ij} &= 0, \qquad \forall (i,j) \notin E, \end{aligned}$$

Example

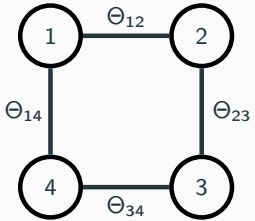


Figure 17: Example

Example

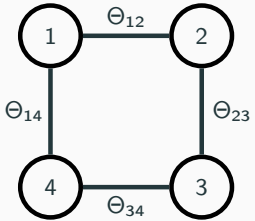


Figure 17: Example

$$\widehat{\Gamma} = \begin{pmatrix} 0 & \overline{\Gamma}_{12} & ? & \overline{\Gamma}_{14} \\ \overline{\Gamma}_{12} & 0 & \overline{\Gamma}_{23} & ? \\ ? & \overline{\Gamma}_{23} & 0 & \overline{\Gamma}_{34} \\ \overline{\Gamma}_{14} & ? & \overline{\Gamma}_{34} & 0 \end{pmatrix}, \qquad \widehat{\Theta} = \begin{pmatrix} * & ? & 0 & ? \\ ? & * & ? & 0 \\ 0 & ? & * & ? \\ ? & 0 & ? & * \end{pmatrix}.$$

Example

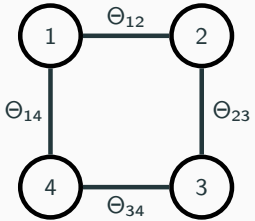


Figure 17: Example

$$\widehat{\Gamma} = \begin{pmatrix} 0 & \overline{\Gamma}_{12} & ? & \overline{\Gamma}_{14} \\ \overline{\Gamma}_{12} & 0 & \overline{\Gamma}_{23} & ? \\ ? & \overline{\Gamma}_{23} & 0 & \overline{\Gamma}_{34} \\ \overline{\Gamma}_{14} & ? & \overline{\Gamma}_{34} & 0 \end{pmatrix}, \qquad \widehat{\Theta} = \begin{pmatrix} * & ? & 0 & ? \\ ? & * & ? & 0 \\ 0 & ? & * & ? \\ ? & 0 & ? & * \end{pmatrix}.$$

This graph is not decomposable/ chordal, and we need to obtain the solution numerically. However, if the graph is decomposable, the matrix completion problem has a rational solution!

Rational solution for the Laplacian matrix completion problem

- For some $d \times d$-matrix M and some index set $A \subset V$ with $\#V = d$, let $\left[M_{A,A}\right]^d$ be the $d \times d$-matrix that equals $M_{i,j}$ for $i, j \in A$, and zero elsewhere.

Rational solution for the Laplacian matrix completion problem

- For some $d \times d$-matrix M and some index set $A \subset V$ with $\#V = d$, let $\left[M_{A,A}\right]^d$ be the $d \times d$-matrix that equals $M_{i,j}$ for $i, j \in A$, and zero elsewhere.
- Let $G = (V, E)$ be a decomposable connected graph with $\#V = d$, set of maximal cliques $\mathcal{C}$ and set of separating cliques $\mathcal{S}$.

Rational solution for the Laplacian matrix completion problem

- For some $d \times d$-matrix M and some index set $A \subset V$ with $\#V = d$, let $\left[M_{A,A}\right]^d$ be the $d \times d$-matrix that equals $M_{i,j}$ for $i, j \in A$, and zero elsewhere.
- Let $G = (V, E)$ be a decomposable connected graph with $\#V = d$, set of maximal cliques $\mathcal{C}$ and set of separating cliques $\mathcal{S}$.

Solution for decomposable graphs

Then, the Laplacian matrix completion problem is solved by

$$\widehat{\Theta} = \sum_{C \in \mathcal{C}} \left[\theta(\overline{\Gamma}_{C,C})\right]^d - \sum_{S \in \mathcal{S}} \nu(S) \left[\theta(\overline{\Gamma}_{S,S})\right]^d,$$

where $\theta(\Gamma_{A,A})$ is the signed graph Laplacian corresponding to $\Gamma_{A,A}$ and $\nu(S)$ is the number of times the separating clique S appears in a perfect sequence.

Example

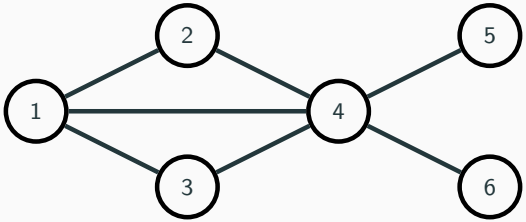


Figure 18: A decomposable graph

Example

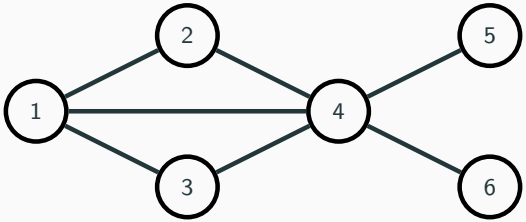


Figure 18: A decomposable graph

$$\widehat{\Theta} = \begin{pmatrix} \theta_{11}(\overline{\Gamma}_{C_1}) & \theta_{12}(\overline{\Gamma}_{C_1}) & 0 & \theta_{14}(\overline{\Gamma}_{C_1}) & 0 & 0 \\ \theta_{12}(\overline{\Gamma}_{C_1}) & \theta_{22}(\overline{\Gamma}_{C_1}) & 0 & \theta_{24}(\overline{\Gamma}_{C_1}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{14}(\overline{\Gamma}_{C_1}) & \theta_{24}(\overline{\Gamma}_{C_1}) & 0 & \theta_{44}(\overline{\Gamma}_{C_1}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \theta_{11}(\overline{\Gamma}_{C_2}) & 0 & \theta_{13}(\overline{\Gamma}_{C_2}) & \theta_{14}(\overline{\Gamma}_{C_2}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{13}(\overline{\Gamma}_{C_2}) & 0 & \theta_{33}(\overline{\Gamma}_{C_2}) & \theta_{34}(\overline{\Gamma}_{C_2}) & 0 & 0 \\ \theta_{14}(\overline{\Gamma}_{C_2}) & 0 & \theta_{34}(\overline{\Gamma}_{C_2}) & \theta_{44}(\overline{\Gamma}_{C_2}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{44}(\overline{\Gamma}_{C_3}) & \theta_{45}(\overline{\Gamma}_{C_3}) & 0 \\ 0 & 0 & 0 & \theta_{45}(\overline{\Gamma}_{C_3}) & \theta_{55}(\overline{\Gamma}_{C_3}) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{44}(\overline{\Gamma}_{C_4}) & 0 & \theta_{46}(\overline{\Gamma}_{C_4}) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{46}(\overline{\Gamma}_{C_4}) & 0 & \theta_{66}(\overline{\Gamma}_{C_4}) \end{pmatrix} - \begin{pmatrix} \theta_{11}(\overline{\Gamma}_{S_1}) & 0 & 0 & \theta_{14}(\overline{\Gamma}_{S_1}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{14}(\overline{\Gamma}_{S_1}) & 0 & 0 & \theta_{44}(\overline{\Gamma}_{S_1}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Structure learning for Hüsler–Reiss and extremal MTP_2

Structure learning

- In structure learning we are interested in learning the sparsity pattern of Θ from data.

Structure learning

- In structure learning we are interested in learning the sparsity pattern of Θ from data.
- Various graphical lasso approaches have been proposed, where an ℓ_1-penalty is added to a log-likelihood. $\rightarrow$ Engelke et al. (2025), Wan and Zhou (2023)

Structure learning

- In structure learning we are interested in learning the sparsity pattern of Θ from data.
- Various graphical lasso approaches have been proposed, where an ℓ_1-penalty is added to a log-likelihood. $\rightarrow$ Engelke et al. (2025), Wan and Zhou (2023)
- Another option is penalized score matching. $\rightarrow$ Lederer and Oesting (2023)

Structure learning

- In structure learning we are interested in learning the sparsity pattern of Θ from data.
- Various graphical lasso approaches have been proposed, where an ℓ_1-penalty is added to a log-likelihood. $\rightarrow$ Engelke et al. (2025), Wan and Zhou (2023)
- Another option is penalized score matching. $\rightarrow$ Lederer and Oesting (2023)
- If positive dependence can be assumed, the extremal MTP_2 method is a combined structural and parametric estimation procedure with favorable properties. $\rightarrow$ R. et al. (2023)

Extremal MTP_2

Definition

A random vector with density f is multivariate totally positive of order 2 (MTP_2) if for all $x, y \in \mathbb{R}^d$:

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y).$$

Extremal MTP_2

Definition

A random vector with density f is multivariate totally positive of order 2 (MTP_2) if for all $x, y \in \mathbb{R}^d$:

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y).$$

- Clearly, the definition of MTP_2 is only useful on product spaces.

Extremal MTP_2

Definition

A random vector with density f is multivariate totally positive of order 2 (MTP_2) if for all $x, y \in \mathbb{R}^d$:

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y).$$

- Clearly, the definition of MTP_2 is only useful on product spaces.
- Therefore, we introduce an extremal version in a similar way to conditional independence:

Extremal MTP_2

Definition

A random vector with density f is multivariate totally positive of order 2 (MTP_2) if for all $\mathrm{x}, \mathrm{y} \in \mathbb{R}^d$:

$$f(\mathrm{x})f(\mathrm{y}) \leq f(\mathrm{x} \wedge \mathrm{y})f(\mathrm{x} \vee \mathrm{y}).$$

- Clearly, the definition of MTP_2 is only useful on product spaces.
- Therefore, we introduce an extremal version in a similar way to conditional independence:

Definition

Y is extremal MTP_2 (EMTP_2) when $\forall k \in V$ it holds that Y^k is MTP_2.

EMTP_2 for Hüsler–Reiss

Theorem

A Hüsler–Reiss distribution with parameter matrix Γ and precision matrix Θ is EMTP_2 if and only if $\Theta_{ij} \leq 0$ for all $i \neq j$.

EMTP_2 for Hüsler–Reiss

Theorem

A Hüsler–Reiss distribution with parameter matrix Γ and precision matrix Θ is EMTP_2 if and only if $\Theta_{ij} \leq 0$ for all $i \neq j$.

Assume a Hüsler–Reiss distribution and an empirical variogram matrix $\overline{\Gamma}$.

EMTP_2 for Hüsler–Reiss

Theorem

A Hüsler–Reiss distribution with parameter matrix Γ and precision matrix Θ is EMTP_2 if and only if $\Theta_{ij} \leq 0$ for all $i \neq j$.

Assume a Hüsler–Reiss distribution and an empirical variogram matrix $\overline{\Gamma}$.

Surrogate MLE

We propose a surrogate maximum likelihood estimator (MLE) under EMTP_2:

Theorem

A Hüsler–Reiss distribution with parameter matrix Γ and precision matrix Θ is EMTP$_2$ if and only if $\Theta_{ij} \leq 0$ for all $i \neq j$.

Assume a Hüsler–Reiss distribution and an empirical variogram matrix $\overline{\Gamma}$.

Surrogate MLE

We propose a surrogate maximum likelihood estimator (MLE) under EMTP$_2$:

$$\widehat{\Theta} = \operatorname{argmax} \log \operatorname{Det}(\Theta) + \frac{1}{2}\langle \Theta, \overline{\Gamma} \rangle, \text{ s.t. } \Theta_{ij} \leq 0 \quad \forall i \neq j.$$

where Det denotes the pseudo-determinant.

Properties of the surrogate MLE under EMTP_2

- $\widehat{\Theta}$ is usually sparse through the KKT conditions

Properties of the surrogate MLE under EMTP$_2$

- $\widehat{\Theta}$ is usually sparse through the KKT conditions $\Rightarrow$ Sparsity without regularization!

Properties of the surrogate MLE under EMTP$_2$

- $\widehat{\Theta}$ is usually sparse through the KKT conditions $\Rightarrow$ Sparsity without regularization!
- Ying et al. (2021) showed that the estimator exists when $\overline{\Gamma} > 0$.

Properties of the surrogate MLE under EMTP$_2$

- $\widehat{\Theta}$ is usually sparse through the KKT conditions $\Rightarrow$ Sparsity without regularization!
- Ying et al. (2021) showed that the estimator exists when $\overline{\Gamma} > 0$.
- We implement a block descent algorithm in `R`.

Properties of the surrogate MLE under EMTP_2

- $\widehat{\Theta}$ is usually sparse through the KKT conditions $\Rightarrow$ Sparsity without regularization!
- Ying et al. (2021) showed that the estimator exists when $\overline{\Gamma} > 0$.
- We implement a block descent algorithm in `R`.

Theorem

For a Hüsler–Reiss distribution that is EMTP_2, the estimator $\widehat{\Theta}$ converges in probability to Θ.

Properties of the surrogate MLE under EMTP_2

- $\widehat{\Theta}$ is usually sparse through the KKT conditions $\Rightarrow$ Sparsity without regularization!
- Ying et al. (2021) showed that the estimator exists when $\overline{\Gamma} > 0$.
- We implement a block descent algorithm in `R`.

Theorem

For a Hüsler–Reiss distribution that is EMTP_2, the estimator $\widehat{\Theta}$ converges in probability to Θ.

Theorem

The estimated EMTP_2 graph is asymptotically a super-graph of the true underlying graph.

Application

Data example from Asadi et al. (2015, AoAS): Danube basin flow data measured at 31 stations.

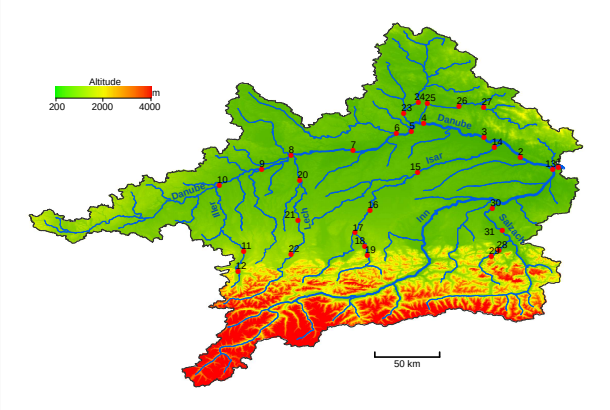


Figure 19: Topographic map.

Data example from Asadi et al. (2015, AoAS): Danube basin flow data measured at 31 stations.

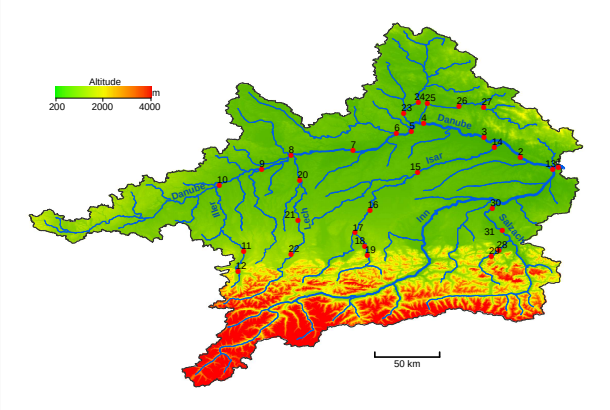


Figure 19: Topographic map.

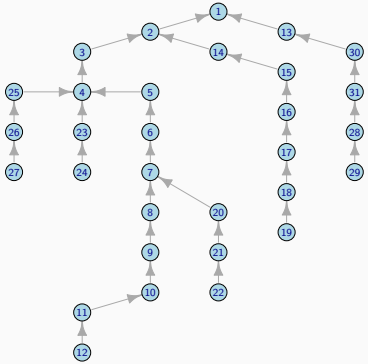
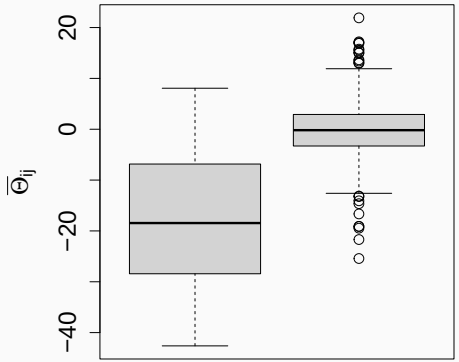


Figure 20: Flow graph.

Empirical estimates $\overline{\Theta}_{ij}$ for edges in the flow graph (left) and the non-diagonal remaining entries (right).



$EMTP_2$ graph

We obtain a sparse estimate for the $EMTP_2$ graph:

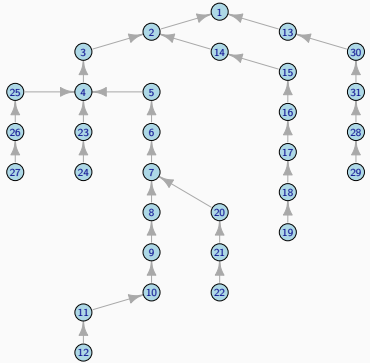


Figure 7: Flow graph.

EMTP_2 graph

We obtain a sparse estimate for the EMTP_2 graph:

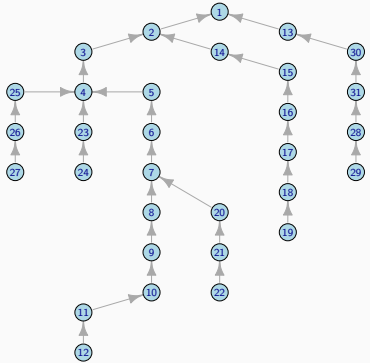


Figure 7: Flow graph.

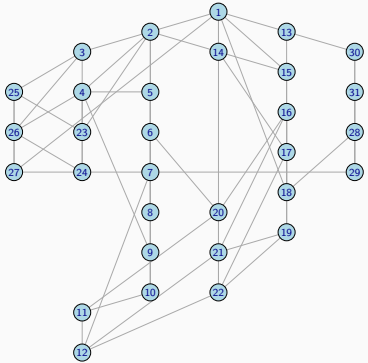


Figure 8: EMTP_2 graph.

AIC table

The EMTP_2 estimator performs better than other graphical methods and is competitive with the specialized model of Asadi et al. (2015, AoAS)!

	twice neg logLH	nb par	AIC
empirical variogram	270.58	465	1200.58
flow graph	1465.22	30	1525.22
MST	1390.04	30	1450.04
best block graph MST	1263.55	42	1347.55
Asadi et al.	1107.78	6	1119.78
EMTP_2 estimator	1034.42	67	1168.42

Table 1: Results for the different models fitted to the Danube river data set; columns show twice the negative log-likelihood, the number of model parameters and the AIC values, respectively.

Colored Hüsler–Reiss graphical models

Motivation

- Graphical models are often sparse, but can still have many parameters.

Motivation

- Graphical models are often sparse, but can still have many parameters.
- Parameter reduction that maintains graphical structure possible via colored graphical models.

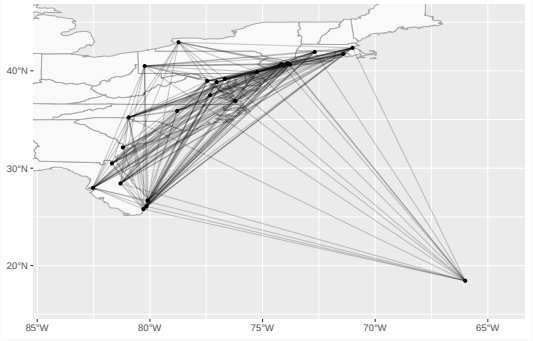


Figure 23: Flight connections in the eastern United States.

- Graphical models are often sparse, but can still have many parameters.
- Parameter reduction that maintains graphical structure possible via colored graphical models.
- **Basic idea:** Edges with the same color impose parameter symmetries.

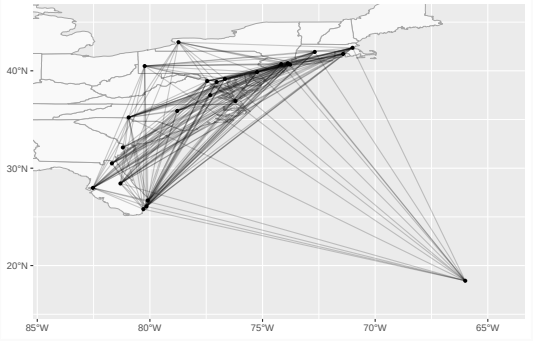


Figure 23: Flight connections in the eastern United States.

- Graphical models are often sparse, but can still have many parameters.
- Parameter reduction that maintains graphical structure possible via colored graphical models.
- **Basic idea:** Edges with the same color impose parameter symmetries.
- Algebraically, these define affine constraints on the parameter space of the model.

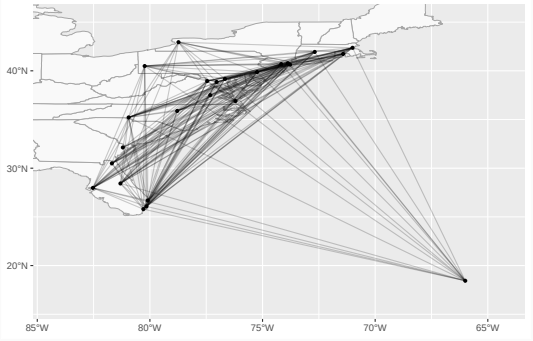


Figure 23: Flight connections in the eastern United States.

Hüsler–Reiss RCON models

Let $G = (V, E)$ be a graph and let $\lambda : E \to [r]$ be a coloring function.

Hüsler–Reiss RCON models

Let $G = (V, E)$ be a graph and let $\lambda : E \to [r]$ be a coloring function.

Definition

A Hüsler–Reiss RCON graphical model with precision matrix Θ satisfies

1. $\Theta_{uv} = 0$ for $uv \notin E$,
2. $\Theta_{st} = \Theta_{uv}$ for $\lambda(st) = \lambda(uv)$ for $st, uv \in E$.

Let $G = (V, E)$ be a graph and let $\lambda : E \to [r]$ be a coloring function.

Definition

A Hüsler–Reiss RCON graphical model with precision matrix Θ satisfies

1. $\Theta_{uv} = 0$ for $uv \notin E$,
2. $\Theta_{st} = \Theta_{uv}$ for $\lambda(st) = \lambda(uv)$ for $st, uv \in E$.

Example: With each edge color we associate a parameter ω_i:

$$\Theta = \begin{pmatrix} * & \omega_1 & \omega_2 & 0 & 0 \\ \omega_1 & * & \omega_2 & 0 & 0 \\ \omega_2 & \omega_2 & * & \omega_3 & \omega_3 \\ 0 & 0 & \omega_3 & * & \omega_1 \\ 0 & 0 & \omega_3 & \omega_1 & * \end{pmatrix}.$$

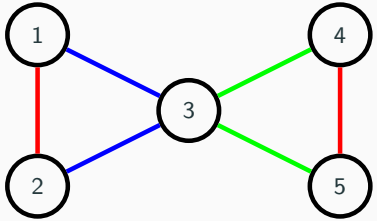


Figure 24: Colored graph with five nodes and six edges

Hüsler–Reiss RVAR models

Definition

A Hüsler–Reiss RVAR graphical model with precision matrix Θ and variogram matrix Γ satisfies

1. $\Theta_{uv} = 0$ when $uv \notin E$,
2. $\Gamma_{st} = \Gamma_{uv}$ when $\lambda(st) = \lambda(uv)$ for $st, uv \in E$.

Definition

A Hüsler–Reiss RVAR graphical model with precision matrix Θ and variogram matrix Γ satisfies

1. $\Theta_{uv} = 0$ when $uv \notin E$,
2. $\Gamma_{st} = \Gamma_{uv}$ when $\lambda(st) = \lambda(uv)$ for $st, uv \in E$.

Example: With each edge color we associate a parameter ν_i:

$$\Gamma = \begin{pmatrix} 0 & \nu_1 & \nu_2 & * & * \\ \nu_1 & 0 & \nu_2 & * & * \\ \nu_2 & \nu_2 & 0 & \nu_3 & \nu_3 \\ * & * & \nu_3 & 0 & \nu_1 \\ * & * & \nu_3 & \nu_1 & 0 \end{pmatrix}, \qquad \Theta = \begin{pmatrix} * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{pmatrix}.$$

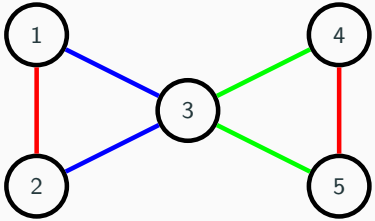


Figure 25: Colored graph with five nodes and six edges

Definition

A Hüsler–Reiss RVAR graphical model with precision matrix Θ and variogram matrix Γ satisfies

1. $\Theta_{uv} = 0$ when $uv \notin E$,
2. $\Gamma_{st} = \Gamma_{uv}$ when $\lambda(st) = \lambda(uv)$ for $st, uv \in E$.

Example: With each edge color we associate a parameter ν_i:

$$\Gamma = \begin{pmatrix} 0 & \nu_1 & \nu_2 & * & * \\ \nu_1 & 0 & \nu_2 & * & * \\ \nu_2 & \nu_2 & 0 & \nu_3 & \nu_3 \\ * & * & \nu_3 & 0 & \nu_1 \\ * & * & \nu_3 & \nu_1 & 0 \end{pmatrix}, \qquad \Theta = \begin{pmatrix} * & * & * & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{pmatrix}.$$

Note that this is a mixed parameterization!

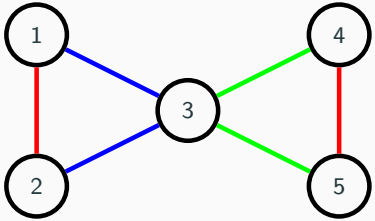


Figure 25: Colored graph with five nodes and six edges

Application

Flights data

- We demonstrate the application of the RVAR model on the flight delays data set of Hentschel et al. (2022), which is available in the `R` package `graphicalExtremes`.

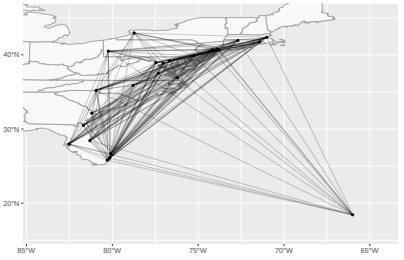


Figure 26: Flight connections for the eastern cluster.

Flights data

- We demonstrate the application of the RVAR model on the flight delays data set of Hentschel et al. (2022), which is available in the `R` package `graphicalExtremes`.
- The data set consists of total daily delays in minutes for airports in the United States.

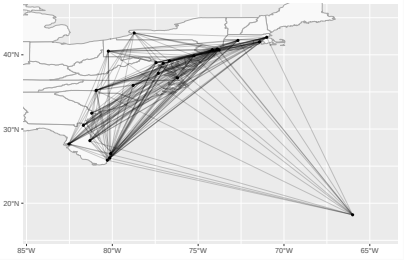


Figure 26: Flight connections for the eastern cluster.

Flights data

- We demonstrate the application of the RVAR model on the flight delays data set of Hentschel et al. (2022), which is available in the `R` package `graphicalExtremes`.
- The data set consists of total daily delays in minutes for airports in the United States.
- We select all daily observations for a cluster of airports in the eastern US with at least 2000 incoming and outgoing annual flights between 2005 and 2020.

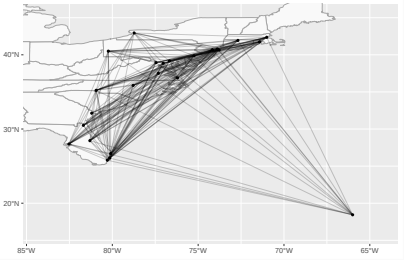


Figure 26: Flight connections for the eastern cluster.

Flights data

- We demonstrate the application of the RVAR model on the flight delays data set of Hentschel et al. (2022), which is available in the `R` package `graphicalExtremes`.
- The data set consists of total daily delays in minutes for airports in the United States.
- We select all daily observations for a cluster of airports in the eastern US with at least 2000 incoming and outgoing annual flights between 2005 and 2020.
- We remove any day with missing observations, which results in $n = 5347$ daily observations.

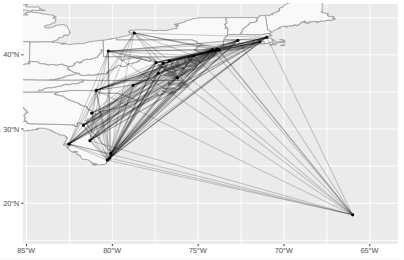


Figure 26: Flight connections for the eastern cluster.

Flights data

- We demonstrate the application of the RVAR model on the flight delays data set of Hentschel et al. (2022), which is available in the `R` package `graphicalExtremes`.
- The data set consists of total daily delays in minutes for airports in the United States.
- We select all daily observations for a cluster of airports in the eastern US with at least 2000 incoming and outgoing annual flights between 2005 and 2020.
- We remove any day with missing observations, which results in $n = 5347$ daily observations.
- The data is split into training and validation data sets.

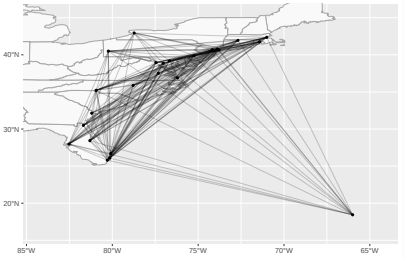


Figure 26: Flight connections for the eastern cluster.

RVAR estimation for the eastern cluster

- We calculate the empirical variogram $\overline{\Gamma}$.

RVAR estimation for the eastern cluster

- We calculate the empirical variogram $\overline{\Gamma}$.
- We estimate the graphical structure with a structure learning method (EMTP_2). This gives a graph with 24 nodes and 101 edges.

RVAR estimation for the eastern cluster

- We calculate the empirical variogram $\overline{\Gamma}$.
- We estimate the graphical structure with a structure learning method (EMTP_2). This gives a graph with 24 nodes and 101 edges.
- For each number of colors between 1 and 25 colors, we cluster the edges into edge color classes depending on $\overline{\Gamma}$.

RVAR estimation for the eastern cluster

- We calculate the empirical variogram $\overline{\Gamma}$.
- We estimate the graphical structure with a structure learning method (EMTP_2). This gives a graph with 24 nodes and 101 edges.
- For each number of colors between 1 and 25 colors, we cluster the edges into edge color classes depending on $\overline{\Gamma}$.
- We calculate the estimator and the Hüsler–Reiss log-likelihood on the test data.

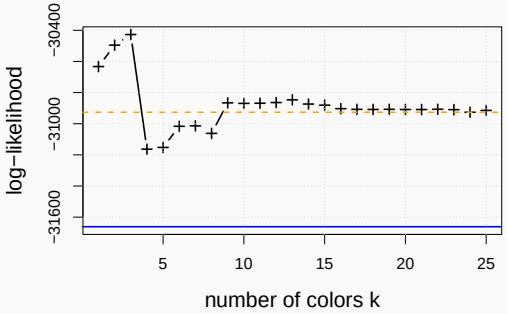


Figure 27: Eastern cluster likelihoods. The dashed line refers to the EMTP_2 estimator and the solid line to $\overline{\Gamma}$.

Best RVAR graph estimate

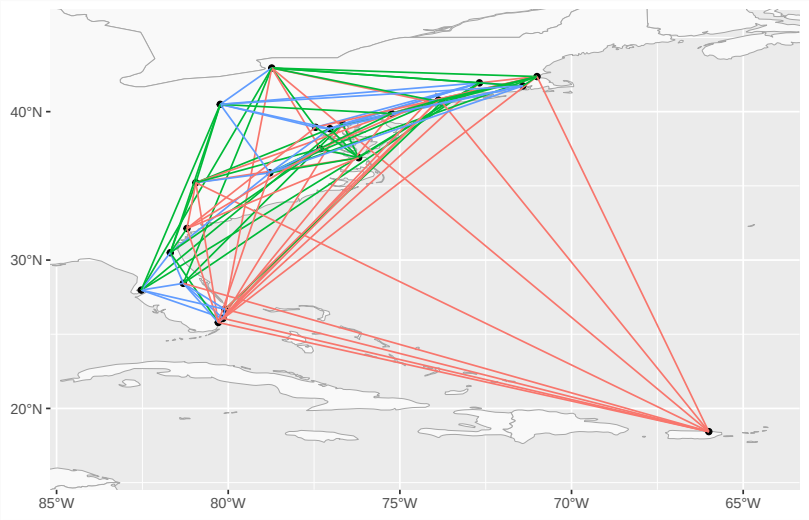


Figure 28: Eastern cluster graph estimate with 3 colors

Latent Hüsler–Reiss graphical models

Latent Hüsler–Reiss graphical models

- Engelke and Taeb (2024) introduced a latent Hüsler–Reiss graphical model.

Latent Hüsler–Reiss graphical models

- Engelke and Taeb (2024) introduced a latent Hüsler–Reiss graphical model.
- Let some random vector $X = (X_O, X_H)$ be in the domain of attraction of a Hüsler–Reiss vector Y with parameters

$$\Gamma = \begin{pmatrix} \Gamma_{OO} & \Gamma_{OH} \\ \Gamma_{HO} & \Gamma_{HH} \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_{OO} & \Theta_{OH} \\ \Theta_{HO} & \Theta_{HH} \end{pmatrix}.$$

Latent Hüsler–Reiss graphical models

- Engelke and Taeb (2024) introduced a latent Hüsler–Reiss graphical model.
- Let some random vector $X = (X_O, X_H)$ be in the domain of attraction of a Hüsler–Reiss vector Y with parameters

$$\Gamma = \begin{pmatrix} \Gamma_{OO} & \Gamma_{OH} \\ \Gamma_{HO} & \Gamma_{HH} \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_{OO} & \Theta_{OH} \\ \Theta_{HO} & \Theta_{HH} \end{pmatrix}.$$

- Then, the random vector X_O is in the domain of attraction of a Hüsler–Reiss distribution with variogram Γ_{OO}.

Latent Hüsler–Reiss graphical models

- Engelke and Taeb (2024) introduced a latent Hüsler–Reiss graphical model.
- Let some random vector $X = (X_O, X_H)$ be in the domain of attraction of a Hüsler–Reiss vector Y with parameters

$$\Gamma = \begin{pmatrix} \Gamma_{OO} & \Gamma_{OH} \\ \Gamma_{HO} & \Gamma_{HH} \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_{OO} & \Theta_{OH} \\ \Theta_{HO} & \Theta_{HH} \end{pmatrix}.$$

- Then, the random vector X_O is in the domain of attraction of a Hüsler–Reiss distribution with variogram Γ_{OO}.
- The precision matrix corresponding to Γ_{OO} can be obtained for example via the Fiedler–Bapat identity and calculates as the Schur complement

$$\widetilde{\Theta} = \Theta_{OO} - \Theta_{OH}(\Theta_{HH})^{-1}\Theta_{HO}.$$

Latent Hüsler–Reiss graphical models

- Engelke and Taeb (2024) introduced a latent Hüsler–Reiss graphical model.
- Let some random vector $X = (X_O, X_H)$ be in the domain of attraction of a Hüsler–Reiss vector Y with parameters

$$\Gamma = \begin{pmatrix} \Gamma_{OO} & \Gamma_{OH} \\ \Gamma_{HO} & \Gamma_{HH} \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_{OO} & \Theta_{OH} \\ \Theta_{HO} & \Theta_{HH} \end{pmatrix}.$$

- Then, the random vector X_O is in the domain of attraction of a Hüsler–Reiss distribution with variogram Γ_{OO}.
- The precision matrix corresponding to Γ_{OO} can be obtained for example via the Fiedler–Bapat identity and calculates as the Schur complement

$$\widetilde{\Theta} = \Theta_{OO} - \Theta_{OH}(\Theta_{HH})^{-1}\Theta_{HO}.$$

- Now, if X is in the domain of attraction of a sparse Hüsler–Reiss graphical model, the underlying parameter matrix Θ is sparse.

Latent Hüsler–Reiss graphical models

- Engelke and Taeb (2024) introduced a latent Hüsler–Reiss graphical model.
- Let some random vector $X = (X_O, X_H)$ be in the domain of attraction of a Hüsler–Reiss vector Y with parameters

$$\Gamma = \begin{pmatrix} \Gamma_{OO} & \Gamma_{OH} \\ \Gamma_{HO} & \Gamma_{HH} \end{pmatrix}, \qquad \Theta = \begin{pmatrix} \Theta_{OO} & \Theta_{OH} \\ \Theta_{HO} & \Theta_{HH} \end{pmatrix}.$$

- Then, the random vector X_O is in the domain of attraction of a Hüsler–Reiss distribution with variogram Γ_{OO}.
- The precision matrix corresponding to Γ_{OO} can be obtained for example via the Fiedler–Bapat identity and calculates as the Schur complement

$$\widetilde{\Theta} = \Theta_{OO} - \Theta_{OH}(\Theta_{HH})^{-1}\Theta_{HO}.$$

- Now, if X is in the domain of attraction of a sparse Hüsler–Reiss graphical model, the underlying parameter matrix Θ is sparse.

$\Rightarrow$ We can learn the sparsity pattern using separate penalization of the sparse and low-rank components.

Directed graphical models in extremes

Directed graph terminology

- Let $G = (V, E)$ be a graph with vertex set V and edge set $E \subset V \times V$.

Directed graph terminology

- Let $G = (V, E)$ be a graph with vertex set V and edge set $E \subset V \times V$.
- A directed acyclic graph (DAG) is a directed graph without directed cycles.

Directed graph terminology

- Let $G = (V, E)$ be a graph with vertex set V and edge set $E \subset V \times V$.
- A directed acyclic graph (DAG) is a directed graph without directed cycles.

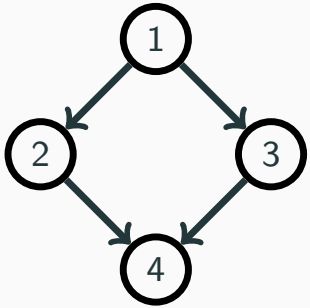


Figure 29: Diamond DAG

Directed graph terminology

- Let $G = (V, E)$ be a graph with vertex set V and edge set $E \subset V \times V$.
- A directed acyclic graph (DAG) is a directed graph without directed cycles.
- Let $\mathrm{pa}(i)$ be the set of parents of some $i \in V$.

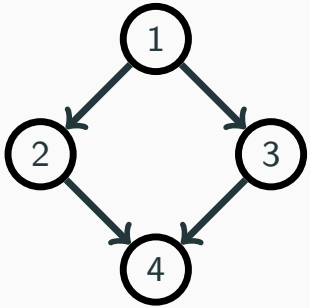


Figure 29: Diamond DAG

Directed graph terminology

- Let $G = (V, E)$ be a graph with vertex set V and edge set $E \subset V \times V$.
- A directed acyclic graph (DAG) is a directed graph without directed cycles.
- Let $\mathrm{pa}(i)$ be the set of parents of some $i \in V$.
- In this talk we only consider rooted DAGs!

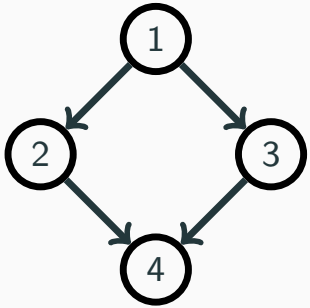


Figure 29: Diamond DAG

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal global Markov property with respect to a DAG $G = (V, E)$ if

$$A \perp\!\!\!\perp_{\mathcal{G}} B | C \Rightarrow Y_A \perp_e Y_B | Y_C,$$

for all disjoint index sets $A, B, C \subset V$.

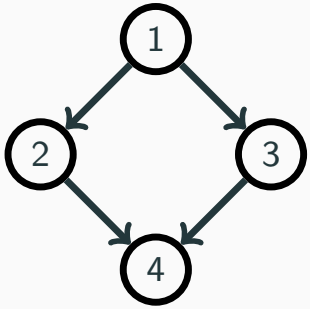


Figure 30: Diamond DAG

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal global Markov property with respect to a DAG $G = (V, E)$ if

$$A \perp\!\!\!\perp_{\mathcal{G}} B | C \Rightarrow Y_A \perp_e Y_B | Y_C,$$

for all disjoint index sets $A, B, C \subset V$. Then we call Y a directed extremal graphical model with respect to G.

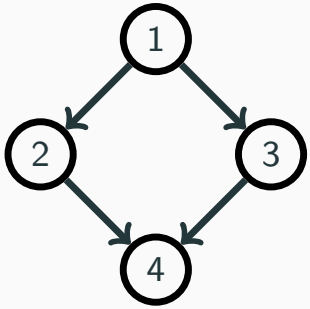


Figure 30: Diamond DAG

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal global Markov property with respect to a DAG $G = (V, E)$ if

$$A \perp\!\!\!\perp_{\mathcal{G}} B | C \Rightarrow Y_A \perp_e Y_B | Y_C,$$

for all disjoint index sets $A, B, C \subset V$. Then we call Y a directed extremal graphical model with respect to G.

Example:

- The set $\{2, 3\}$ blocks both paths from 1 to 4. Hence,

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\} \Rightarrow Y_1 \perp_e Y_4 | \{Y_2, Y_3\}.$$

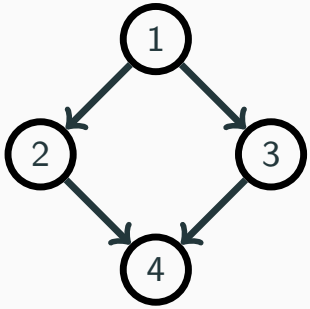


Figure 30: Diamond DAG

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal global Markov property with respect to a DAG $G = (V, E)$ if

$$A \perp\!\!\!\perp_{\mathcal{G}} B | C \Rightarrow Y_A \perp_e Y_B | Y_C,$$

for all disjoint index sets $A, B, C \subset V$. Then we call Y a directed extremal graphical model with respect to G.

Example:

- The set $\{2, 3\}$ blocks both paths from 1 to 4. Hence,

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\} \Rightarrow Y_1 \perp_e Y_4 | \{Y_2, Y_3\}.$$

- The singleton 1 blocks all paths between 2 and 3 because of the collider $2 \to 4 \leftarrow 3$. Hence,

$$2 \perp\!\!\!\perp_G 3 | 1 \Rightarrow Y_2 \perp_e Y_3 | Y_1.$$

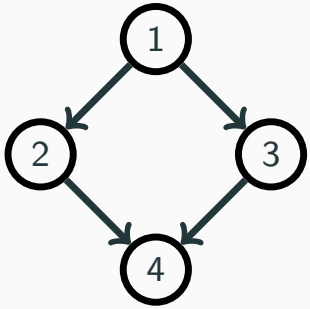


Figure 30: Diamond DAG

Conditional exponent measure density

Let Y be multivariate Pareto with exponent measure density λ.

Conditional exponent measure density

Let Y be multivariate Pareto with exponent measure density λ.

Definition

Let

$$\lambda(y_A \mid y_C) := \frac{\lambda_{A\cup C}(y_{A\cup C})}{\lambda_C(y_C)}$$

be the conditional exponent measure density of Y_A^k given $Y_C^k = y_C$.

Conditional exponent measure density

Let Y be multivariate Pareto with exponent measure density λ.

Definition

Let

$$\lambda(y_A \mid y_C) := \frac{\lambda_{A\cup C}(y_{A\cup C})}{\lambda_C(y_C)}$$

be the conditional exponent measure density of Y_A^k given $Y_C^k = y_C$.

Thus, extremal conditional independence $Y_A \perp_e Y_B \mid Y_C$ is equivalent to

$$\lambda(y_{A\cup B} \mid y_C) = \lambda(y_A \mid y_C)\lambda(y_B \mid y_C)$$

for all $y \in \mathcal{L} = \bigcup_{k\in V} \mathcal{L}^k$.

Markov factorization property

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal Markov factorization property with respect to a DAG $G = (V, E)$ if the exponent measure density λ satisfies

$$\lambda(\mathrm{y}) = \prod_{v \in V} \lambda(y_v \mid \mathrm{y}_{\mathrm{pa}(v)}), \quad \forall \mathrm{y} \in \mathcal{L}.$$

Markov factorization property

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal Markov factorization property with respect to a DAG $G = (V, E)$ if the exponent measure density λ satisfies

$$\lambda(\mathrm{y}) = \prod_{v \in V} \lambda(y_v \mid \mathrm{y}_{\mathrm{pa}(v)}), \quad \forall \mathrm{y} \in \mathcal{L}.$$

Then we call Y a directed extremal graphical model with respect to G.

Markov factorization property

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal Markov factorization property with respect to a DAG $G = (V, E)$ if the exponent measure density λ satisfies

$$\lambda(\mathrm{y}) = \prod_{v \in V} \lambda(y_v \mid \mathrm{y}_{\mathrm{pa}(v)}), \quad \forall \mathrm{y} \in \mathcal{L}.$$

Then we call Y a directed extremal graphical model with respect to G.

For the diamond DAG, this implies

$$\lambda(\mathrm{y}) = \lambda_1(y_1)\lambda(y_2|y_1)\lambda(y_3|y_1)\lambda(y_4|y_2, y_3)$$

for all $\mathrm{y} \in \mathcal{L}$.

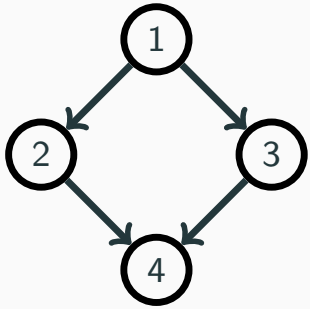


Figure 31: Diamond DAG

Definition

A multivariate Pareto distribution $\mathbb{P}_Y$ satisfies the extremal Markov factorization property with respect to a DAG $G = (V, E)$ if the exponent measure density λ satisfies

$$\lambda(\mathrm{y}) = \prod_{v \in V} \lambda(y_v \mid \mathrm{y}_{\mathrm{pa}(v)}), \quad \forall \mathrm{y} \in \mathcal{L}.$$

Then we call Y a directed extremal graphical model with respect to G.

For the diamond DAG, this implies

$$\lambda(\mathrm{y}) = \lambda_1(y_1)\lambda(y_2|y_1)\lambda(y_3|y_1)\lambda(y_4|y_2, y_3)$$

for all $\mathrm{y} \in \mathcal{L}$.

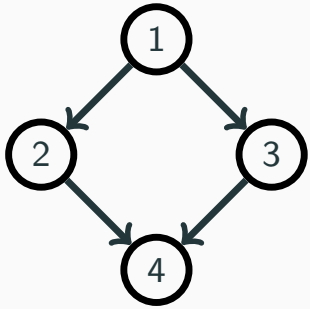


Figure 31: Diamond DAG

Corollary

The extremal global Markov and factorization properties are equivalent.

Conditional exponent measure density for Hüsler–Reiss

Hüsler–Reiss conditional exponent measure density

Let Y be Hüsler–Reiss distributed with variogram matrix Γ.

Conditional exponent measure density for Hüsler–Reiss

Hüsler–Reiss conditional exponent measure density

Let Y be Hüsler–Reiss distributed with variogram matrix Γ. For any disjoint $A, C \subset V$ it is

$$\lambda(y_A \mid y_C) = \frac{1}{\sqrt{\det(2\pi\Sigma_*)}} \exp\left(-\frac{1}{2}(y_A - \mu^*)^\top \Sigma_*^{-1}(y_A - \mu^*)\right),$$

Conditional exponent measure density for Hüsler–Reiss

Hüsler–Reiss conditional exponent measure density

Let Y be Hüsler–Reiss distributed with variogram matrix Γ. For any disjoint $A, C \subset V$ it is

$$\lambda(y_A \mid y_C) = \frac{1}{\sqrt{\det(2\pi\Sigma_*)}} \exp\left(-\frac{1}{2}(y_A - \mu^*)^\top \Sigma_*^{-1}(y_A - \mu^*)\right),$$

where

$$\Sigma_* = -\frac{1}{2}\Gamma_{A,A} - \begin{pmatrix} -\frac{1}{2}\Gamma_{A,C} & \mathbf{1} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}\Gamma_{C,C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2}\Gamma_{C,A} \\ \mathbf{1}^\top \end{pmatrix}$$

Conditional exponent measure density for Hüsler–Reiss

Hüsler–Reiss conditional exponent measure density

Let Y be Hüsler–Reiss distributed with variogram matrix Γ. For any disjoint $A, C \subset V$ it is

$$\lambda(y_A \mid y_C) = \frac{1}{\sqrt{\det(2\pi\Sigma_*)}} \exp\left(-\frac{1}{2}(y_A - \mu^*)^\top \Sigma_*^{-1}(y_A - \mu^*)\right),$$

where

$$\Sigma_* = -\frac{1}{2}\Gamma_{A,A} - \begin{pmatrix} -\frac{1}{2}\Gamma_{A,C} & \mathbf{1} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}\Gamma_{C,C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2}\Gamma_{C,A} \\ \mathbf{1}^\top \end{pmatrix}$$

and

$$\mu_* = \begin{pmatrix} -\frac{1}{2}\Gamma_{A,C}, \mathbf{1} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}\Gamma_{C,C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} y_C \\ 1 \end{pmatrix}.$$

The conditional exponent measure density is Gaussian! Furthermore, when $A \cup C = V$, it holds that $\Sigma_*^{-1} = \Theta_{A,A}$.

Structural equation models from threshold exceedances

Definition

Let G be a DAG rooted in k.

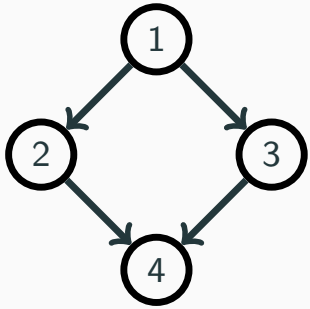


Figure 32: Example for a rooted DAG

Definition

Let G be a DAG rooted in k. If $Y^k (= Y | Y_k > 0)$ satisfies

$$\begin{cases} Y_k^k & \sim \text{Exp}(1) \\ Y_v^k & = \Psi_v(Y_{\text{pa}(v)}^k) + \varepsilon_v, \quad v \neq k, \end{cases}$$

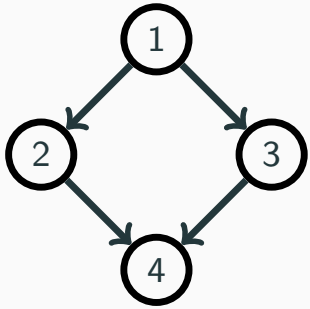


Figure 32: Example for a rooted DAG

Definition

Let G be a DAG rooted in k. If $Y^k (= Y | Y_k > 0)$ satisfies

$$\begin{cases} Y_k^k & \sim \text{Exp}(1) \\ Y_v^k & = \Psi_v(Y_{\text{pa}(v)}^k) + \varepsilon_v, \quad v \neq k, \end{cases}$$

for independent noise variables $(\varepsilon_v)_{v \neq k}$ and structural assignments Ψ_v with

$$\Psi_v(y_{\text{pa}(v)} + t1) = \Psi_v(y_{\text{pa}(v)}) + t,$$

for any $t \in \mathbb{R}$, we call Y an extremal SEM.

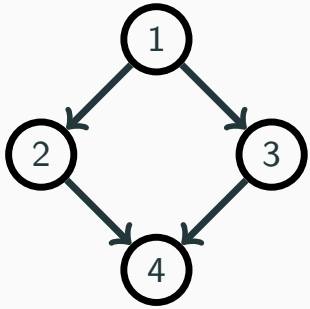


Figure 32: Example for a rooted DAG

Definition

Let G be a DAG rooted in k. If $Y^k (= Y | Y_k > 0)$ satisfies

$$\begin{cases} Y_k^k & \sim \text{Exp}(1) \\ Y_v^k & = \Psi_v(Y_{\text{pa}(v)}^k) + \varepsilon_v, \quad v \neq k, \end{cases}$$

for independent noise variables $(\varepsilon_v)_{v \neq k}$ and structural assignments Ψ_v with

$$\Psi_v(y_{\text{pa}(v)} + t1) = \Psi_v(y_{\text{pa}(v)}) + t,$$

for any $t \in \mathbb{R}$, we call Y an extremal SEM.

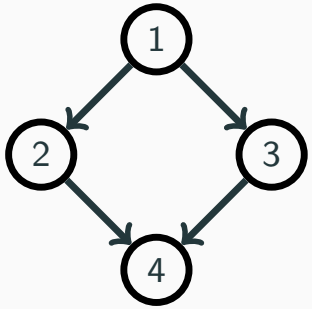


Figure 32: Example for a rooted DAG

Proposition

An extremal SEM is extremal Markov to its underlying rooted DAG.

Hüsler–Reiss structural causal models

- Let G be a rooted DAG with root node 1.

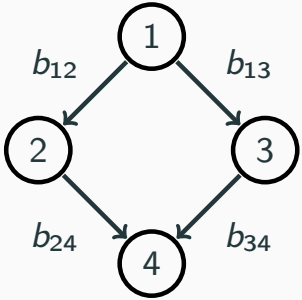


Figure 33: Diamond DAG with edge weights

- Let G be a rooted DAG with root node 1.
- Let $B = (b_{ij})_{1 \leq i,j \leq d}$ be the adjacency matrix of G.

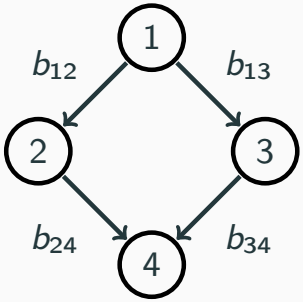


Figure 33: Diamond DAG with edge weights

$$B = \begin{pmatrix} 0 & b_{12} & b_{13} & 0 \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Let G be a rooted DAG with root node 1.
- Let $B = (b_{ij})_{1 \le i,j \le d}$ be the adjacency matrix of G.
- Assume $B_{\setminus 1}^T \mathbf{1} = \mathbf{1}$

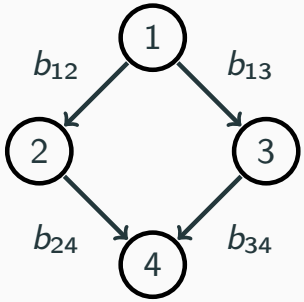


Figure 33: Diamond DAG with edge weights

$$B = \begin{pmatrix} 0 & b_{12} & b_{13} & 0 \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Let G be a rooted DAG with root node 1.
- Let $B = (b_{ij})_{1 \le i,j \le d}$ be the adjacency matrix of G.
- Assume $B_{\setminus 1}^T \mathbf{1} = \mathbf{1}$ and let

$$Y_1^1 := R,$$
$$Y_v^1 := \sum_{i \in \mathrm{pa}(v)} b_{iv} Y_i^1 + \varepsilon_v, \qquad \forall v \neq 1,$$

where $R \sim \mathrm{Exp}(1)$ and ε_v are ind. Gaussians.

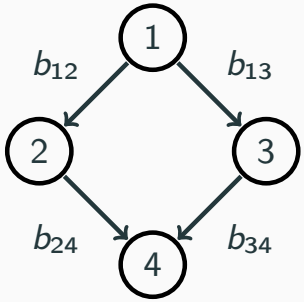


Figure 33: Diamond DAG with edge weights

$$B = \begin{pmatrix} 0 & b_{12} & b_{13} & 0 \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Let G be a rooted DAG with root node 1.
- Let $B = (b_{ij})_{1 \le i,j \le d}$ be the adjacency matrix of G.
- Assume $B_{\setminus 1}^T \mathbf{1} = \mathbf{1}$ and let

$$Y_1^1 := R,$$
$$Y_v^1 := \sum_{i \in \mathrm{pa}(v)} b_{iv} Y_i^1 + \varepsilon_v, \qquad \forall v \neq 1,$$

where $R \sim \mathrm{Exp}(1)$ and ε_v are ind. Gaussians.

Hüsler–Reiss structural causal model

We call the resulting multivariate Pareto vector Y a Hüsler–Reiss structural causal model (SCM) with respect to G.

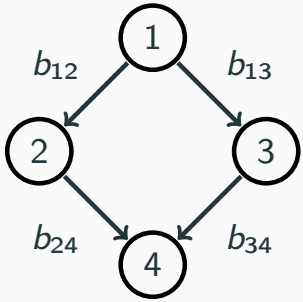


Figure 33: Diamond DAG with edge weights

$$B = \begin{pmatrix} 0 & b_{12} & b_{13} & 0 \\ 0 & 0 & 0 & b_{24} \\ 0 & 0 & 0 & b_{34} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Let Y be a Hüsler–Reiss SCM with respect to a rooted DAG G with root 1.

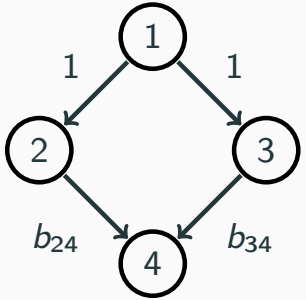


Figure 34: Diamond DAG, assume $b_{24} + b_{34} = 1$.

Hüsler–Reiss structural causal models

- Let Y be a Hüsler–Reiss SCM with respect to a rooted DAG G with root 1.
- Y is a directed Hüsler–Reiss graphical model with respect to G.

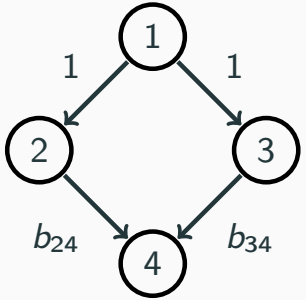


Figure 34: Diamond DAG, assume $b_{24} + b_{34} = 1$.

Hüsler–Reiss structural causal models

- Let Y be a Hüsler–Reiss SCM with respect to a rooted DAG G with root 1.
- Y is a directed Hüsler–Reiss graphical model with respect to G.
- Let B be the adjacency matrix and let $\nu_2^2, \dots \nu_d^2$ be the noise variances.

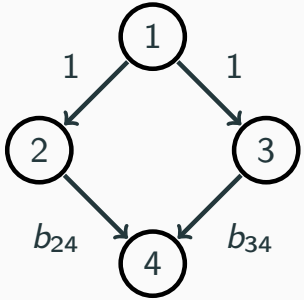


Figure 34: Diamond DAG, assume $b_{24} + b_{34} = 1$.

Hüsler–Reiss structural causal models

- Let Y be a Hüsler–Reiss SCM with respect to a rooted DAG G with root 1.
- Y is a directed Hüsler–Reiss graphical model with respect to G.
- Let B be the adjacency matrix and let $\nu_2^2, \dots \nu_d^2$ be the noise variances.

Hüsler–Reiss SCM precision matrix

The Hüsler–Reiss precision matrix Θ can be factorized as

$$\Theta = (I - B^\top)_{\setminus 1,V}^\top \begin{pmatrix} \frac{1}{\nu_2^2} & & \\ & \ddots & \\ & & \frac{1}{\nu_d^2} \end{pmatrix} (I - B^\top)_{\setminus 1,V}.$$

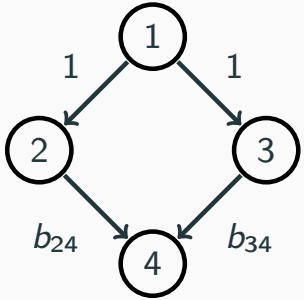


Figure 34: Diamond DAG, assume $b_{24} + b_{34} = 1$.

- Let Y be a Hüsler–Reiss SCM with respect to a rooted DAG G with root 1.
- Y is a directed Hüsler–Reiss graphical model with respect to G.
- Let B be the adjacency matrix and let $\nu_2^2, \dots \nu_d^2$ be the noise variances.

Hüsler–Reiss SCM precision matrix

The Hüsler–Reiss precision matrix Θ can be factorized as

$$\Theta = (I - B^\top)_{\setminus 1,V}^\top \begin{pmatrix} \frac{1}{\nu_2^2} & & \\ & \ddots & \\ & & \frac{1}{\nu_d^2} \end{pmatrix} (I - B^\top)_{\setminus 1,V}.$$

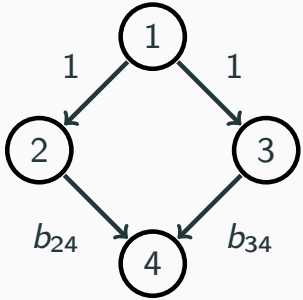


Figure 34: Diamond DAG, assume $b_{24} + b_{34} = 1$.

Example:

$$\Theta = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 0 & -b_{24} \\ 0 & 1 & -b_{34} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\nu_2^2} & & \\ & \frac{1}{\nu_3^2} & \\ & & \frac{1}{\nu_4^2} \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -b_{24} & -b_{34} & 1 \end{pmatrix}$$

Structure learning

Structure learning

- We can adapt the PC-algorithm to this setting!

Structure learning

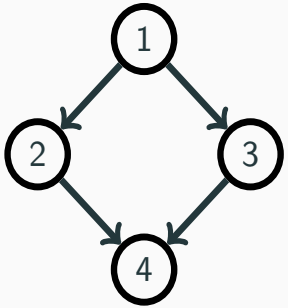
- We can adapt the PC-algorithm to this setting!
- Furthermore, if X is an SCM wrt to a DAG G, the limiting multivariate Pareto Y will be an extremal SCM wrt a subgraph of G.

Structure learning

- We can adapt the PC-algorithm to this setting!
- Furthermore, if X is an SCM wrt to a DAG G, the limiting multivariate Pareto Y will be an extremal SCM wrt a subgraph of G.
- This gives rise to extremal pruning.

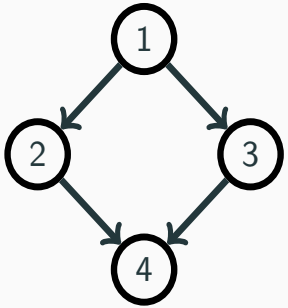
Extremal pruning

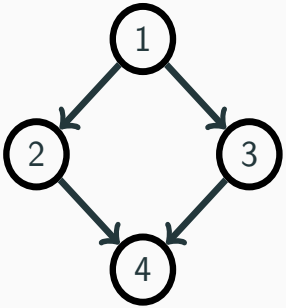
Data generating process X is SCM wrt G



Extremal pruning

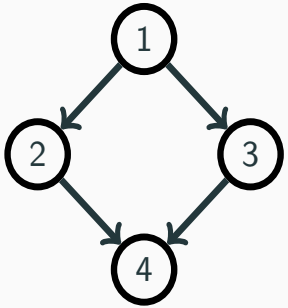
Data generating process X is SCM wrt G $\Longrightarrow$ Multivariate Pareto Y is SCM wrt pruning of G

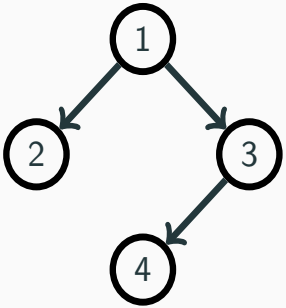




Extremal pruning

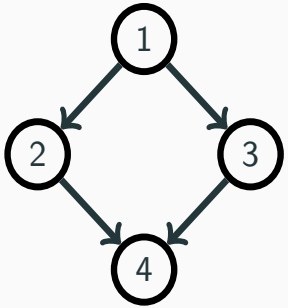
Data generating process X is SCM wrt G $\Longrightarrow$ Multivariate Pareto Y is SCM wrt pruning of G

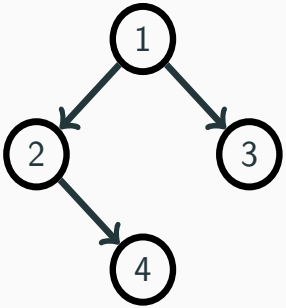




Extremal pruning

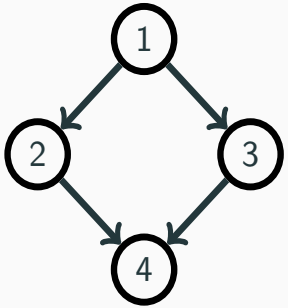
Data generating process X is SCM wrt G $\Longrightarrow$ Multivariate Pareto Y is SCM wrt pruning of G

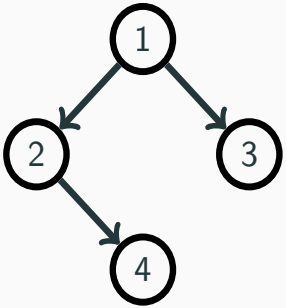




Extremal pruning

Data generating process X is SCM wrt G $\Longrightarrow$ Multivariate Pareto Y is SCM wrt pruning of G

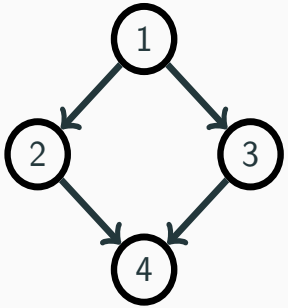


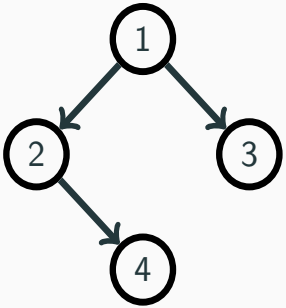


- A structure estimate for X gives a starting point for "extremal pruning", where we try to remove edges.

Extremal pruning

Data generating process X is SCM wrt G $\Longrightarrow$ Multivariate Pareto Y is SCM wrt pruning of G





- A structure estimate for X gives a starting point for "extremal pruning", where we try to remove edges.
- For each prunable edge, we employ a parametric partial correlation test (based on Hüsler–Reiss).

- $Y_i \perp_e Y_j \mid Y_S$ is equivalent to the vanishing of the conditional covariance

$$-\frac{1}{2}\Gamma_{i,j} - \begin{pmatrix} -\frac{1}{2}\Gamma_{i,S} & \mathbf{1} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}\Gamma_{S,S} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2}\Gamma_{S,j} \\ \mathbf{1}^\top \end{pmatrix}.$$

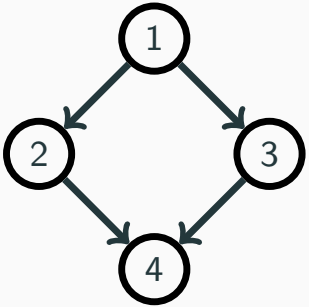


Figure 35: Diamond DAG

- $Y_i \perp_e Y_j \mid Y_S$ is equivalent to the vanishing of the conditional covariance

$$-\frac{1}{2}\Gamma_{i,j} - \begin{pmatrix} -\frac{1}{2}\Gamma_{i,S} & \mathbf{1} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}\Gamma_{S,S} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2}\Gamma_{S,j} \\ \mathbf{1}^\top \end{pmatrix}.$$

$\Rightarrow$ Closed form for a partial correlation coefficient $\rho_{ij|S}$ (only depending on Γ) from the conditional covariance matrix.

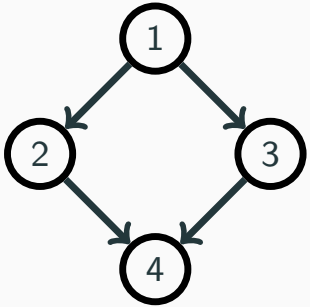


Figure 35: Diamond DAG

- $Y_i \perp_e Y_j \mid Y_S$ is equivalent to the vanishing of the conditional covariance

$$-\frac{1}{2}\Gamma_{i,j} - \begin{pmatrix} -\frac{1}{2}\Gamma_{i,S} & \mathbf{1} \end{pmatrix} \begin{pmatrix} -\frac{1}{2}\Gamma_{S,S} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2}\Gamma_{S,j} \\ \mathbf{1}^\top \end{pmatrix}.$$

$\Rightarrow$ Closed form for a partial correlation coefficient $\rho_{ij|S}$ (only depending on Γ) from the conditional covariance matrix.

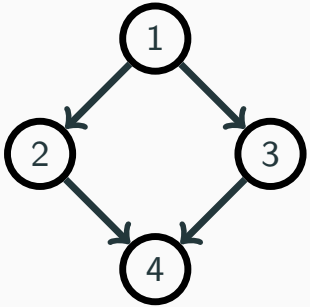


Figure 35: Diamond DAG

$$Y_1 \perp_e Y_4 | \{Y_2, Y_3\} \iff \rho_{14|23} = 0$$

$$Y_2 \perp_e Y_3 | \{Y_1\} \iff \rho_{23|1} = 0$$

Hüsler–Reiss partial correlation test

Hüsler–Reiss conditional independence test

For two indices $i, j \in V$ and a conditioning set $S \subseteq V \setminus \{i, j\}$, we would like to perform the hypothesis test

$$H_0 : Y_i \perp_e Y_j \mid Y_S \quad \text{versus} \quad H_1 : Y_i \not\perp_e Y_j \mid Y_S.$$

Hüsler–Reiss partial correlation test

Hüsler–Reiss conditional independence test

For two indices $i, j \in V$ and a conditioning set $S \subseteq V \setminus \{i, j\}$, we would like to perform the hypothesis test

$$H_0 : Y_i \perp_e Y_j \mid Y_S \quad \text{versus} \quad H_1 : Y_i \not\perp_e Y_j \mid Y_S.$$

$\Rightarrow$ We found a closed form for a partial correlation coefficient $\rho_{ij|S}$ based on Γ.

Hüsler–Reiss conditional independence test

For two indices $i, j \in V$ and a conditioning set $S \subseteq V \setminus \{i, j\}$, we would like to perform the hypothesis test

$$H_0 : Y_i \perp_e Y_j \mid Y_S \quad \text{versus} \quad H_1 : Y_i \not\perp_e Y_j \mid Y_S.$$

$\Rightarrow$ We found a closed form for a partial correlation coefficient $\rho_{ij|S}$ based on Γ.

- We can restate the test $H_0 : \rho_{ij|S} = 0$ versus $H_1 : \rho_{ij|S} \neq 0$.

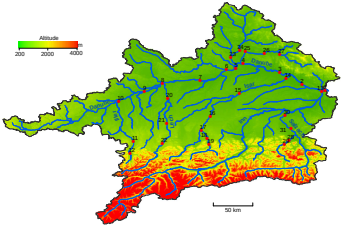


Figure 36: Topographic map of the upper Danube basin.

Hüsler–Reiss conditional independence test

For two indices $i, j \in V$ and a conditioning set $S \subseteq V \setminus \{i, j\}$, we would like to perform the hypothesis test

$$H_0 : Y_i \perp_e Y_j \mid Y_S \quad \text{versus} \quad H_1 : Y_i \not\perp_e Y_j \mid Y_S.$$

$\Rightarrow$ We found a closed form for a partial correlation coefficient $\rho_{ij|S}$ based on Γ.

- We can restate the test $H_0 : \rho_{ij|S} = 0$ versus $H_1 : \rho_{ij|S} \neq 0$.
- Plugging in an estimator $\widehat{\Gamma}$ then gives rise to a test statistic.

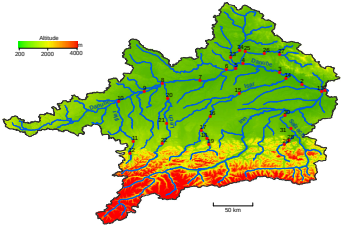


Figure 36: Topographic map of the upper Danube basin.

Hüsler–Reiss conditional independence test

For two indices $i, j \in V$ and a conditioning set $S \subseteq V \setminus \{i, j\}$, we would like to perform the hypothesis test

$$H_0 : Y_i \perp_e Y_j \mid Y_S \quad \text{versus} \quad H_1 : Y_i \not\perp_e Y_j \mid Y_S.$$

$\Rightarrow$ We found a closed form for a partial correlation coefficient $\rho_{ij|S}$ based on Γ.

- We can restate the test $H_0 : \rho_{ij|S} = 0$ versus $H_1 : \rho_{ij|S} \neq 0$.
- Plugging in an estimator $\widehat{\Gamma}$ then gives rise to a test statistic.

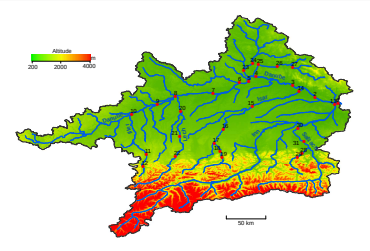
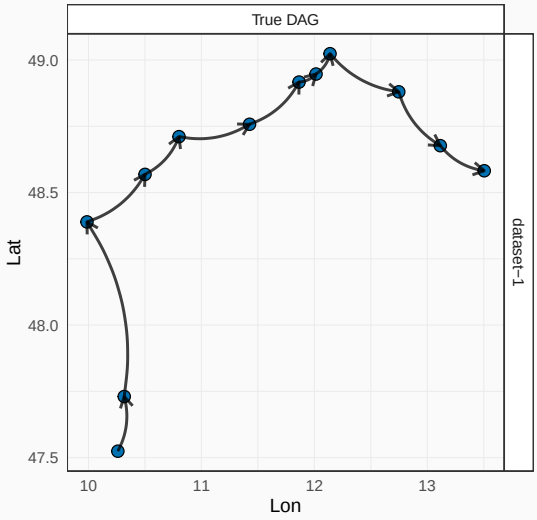


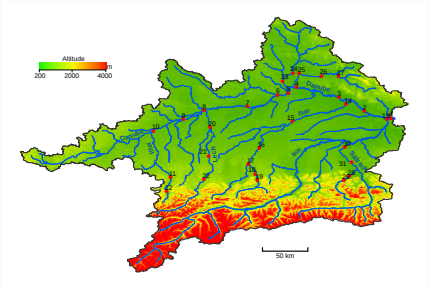
Figure 36: Topographic map of the upper Danube basin.

Stations	p-value
1-2	0.042
1-12	0.694
27-12	0.640

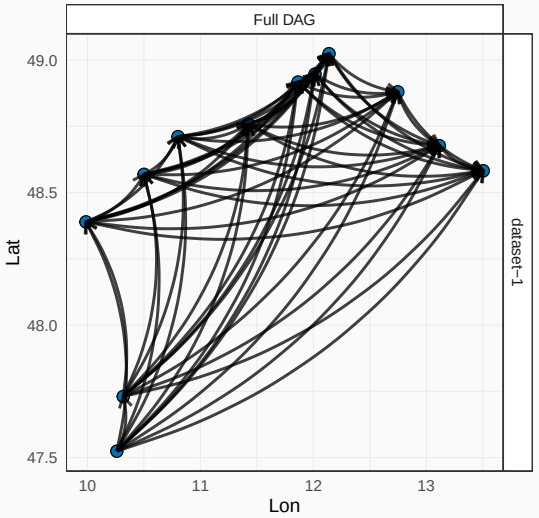
Application

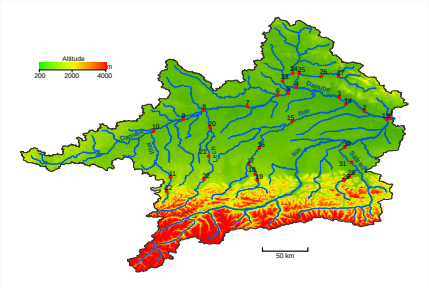
Extremal pruning for the path 1-12



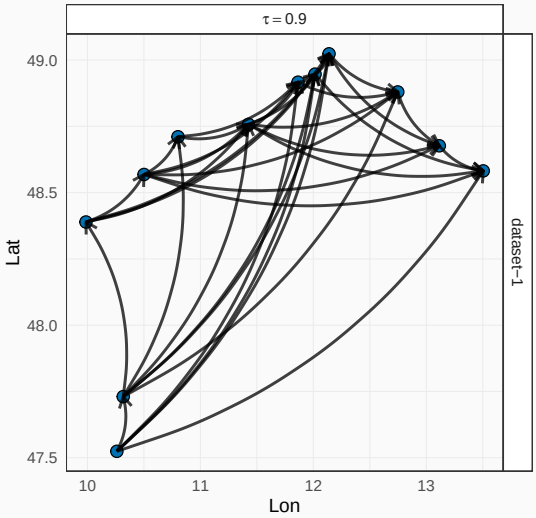


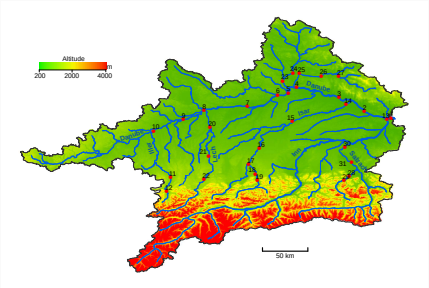
Extremal pruning for the path 1-12



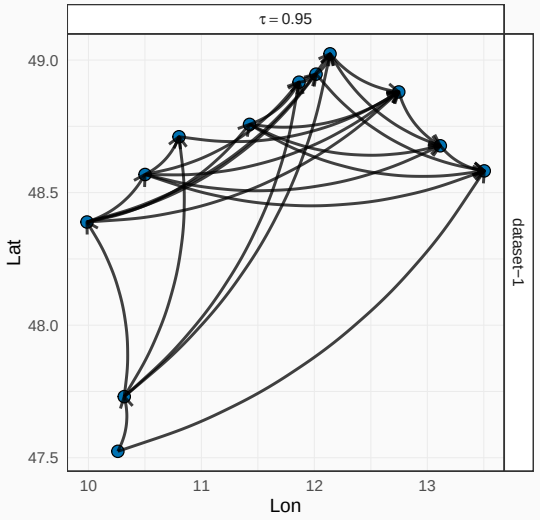


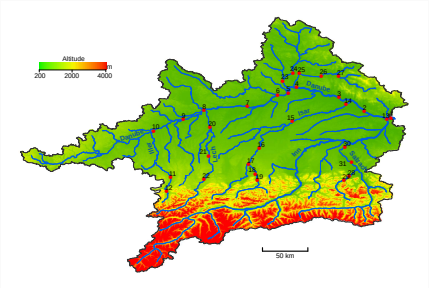
Extremal pruning for the path 1-12



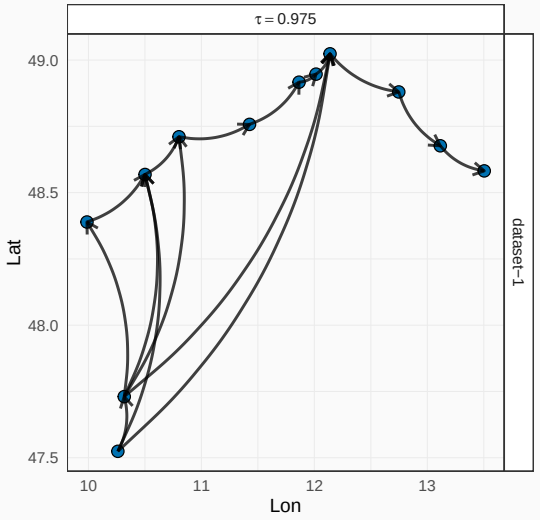


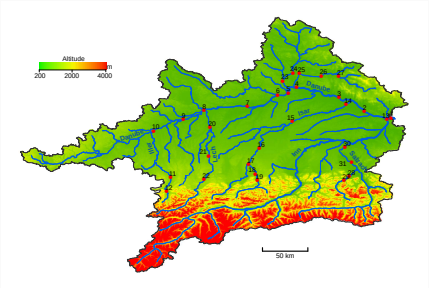
Extremal pruning for the path 1-12





Extremal pruning for the path 1-12





Outlook and discussion

Outlook and discussion

Future work:

- Extremal SCMs:
  - Generalization to arbitrary DAG structures
  - LiNGAM approaches
  - (Asymptotic) distribution of the test statistic?
  - Inclusion of latent variables
- Understand better the Laplacian parametrization
- Find further real-world applications

References

Améndola, C., J. I. Coons, A. Grosdos, and F. Röttger (2026).
Algebraic statistics of Hüsler-Reiss graphical models in multivariate extremes.

Asadi, P., A. C. Davison, and S. Engelke (2015).
Extremes on river networks.
The Annals of Applied Statistics 9(4), 2023 – 2050.

Devriendt, K. (2022).
Effective resistance is more than distance: Laplacians, simplices and the Schur complement.
Linear Algebra and its Applications 639, 24–49.

Devriendt, K., I. E.-S. Rodríguez, and F. Röttger (2026).
Extremal conditional independence for Hüsler-Reiss distributions via modular functions.

Drton, M. and M. H. Maathuis (2017).
Structure learning in graphical modeling.
Annual Review of Statistics and Its Application 4(1), 365–393.

References

Echave-Sustaeta Rodríguez, I. and F. Röttger (2025).
Latent Gaussian and Hüsler-Reiss graphical models with Golazo penalty.
Internat. J. Approx. Reason. 185, Paper No. 109468, 24.

Engelke, S., N. Gnecco, and F. Röttger (2025).
Extremes of structural causal models.
Preprint arXiv:2503.06536.

Engelke, S., M. Hentschel, M. Lalancette, and F. Röttger (2024).
Graphical models for multivariate extremes.

Engelke, S. and A. S. Hitz (2020).
Graphical models for extremes.
Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82(4), 871–932.

Hentschel, M., S. Engelke, and J. Segers (2025).
Statistical inference for Hüsler-Reiss graphical models through matrix completions.
J. Amer. Statist. Assoc. 120(550), 909–921.

References


Lauritzen, S. L. (1996).
***Graphical models*, Volume 17 of *Oxford Statistical Science Series*.**
The Clarendon Press, Oxford University Press, New York.
Oxford Science Publications.


Lederer, J. and M. Oesting (2023).
Extremes in high dimensions: Methods and scalable algorithms.


Peters, J., D. Janzing, and B. Schölkopf (2017).
***Elements of causal inference: foundations and learning algorithms*.**
The MIT Press.


Röttger, F., S. Engelke, and P. Zwiernik (2023).
Total positivity in multivariate extremes.
The Annals of Statistics 51(3), 962 – 1004.


Röttger, F., J. I. Coons, and A. Grosdos (2026).
Parametric and nonparametric symmetries in graphical models for extremes.
To appear in JRSSB.

References


Studený, M. (2005).
Probabilistic Conditional Independence Structures.
Springer.


Wan, P. and C. Zhou (2023).
Graphical lasso for extremes.

Thank You!