

STATISTICAL INFERENCE FOR HÜSLER-REISS GRAPHICAL MODELS THROUGH MATRIX COMPLETIONS

Manuel Hentschel, Sebastian Engelke,
Johan Segers

REPRINT | 2024 / 19

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

Statistical Inference for Hüsler–Reiss Graphical Models Through Matrix Completions

Manuel Hentschel^{a,#}, Sebastian Engelke^a, Johan Segers^{b,*}

^aResearch Center for Statistics, University of Geneva

^bISBA/LIDAM, UCLouvain

*Funding from the F.R.S-FNRS in the form of a PDR grant (T020321F) is gratefully acknowledged.

#manuel.hentschel@unige.ch

Abstract

The severity of multivariate extreme events is driven by the dependence between the largest marginal observations. The Hüsler–Reiss distribution is a versatile model for this extremal dependence, and it is usually parameterized by a variogram matrix. In order to represent conditional independence relations and obtain sparse parameterizations, we introduce the novel Hüsler–Reiss precision matrix. Similarly to the Gaussian case, this matrix appears naturally in density representations of the Hüsler–Reiss Pareto distribution and encodes the extremal graphical structure through its zero pattern. For a given, arbitrary graph we prove the existence and uniqueness of the completion of a partially specified Hüsler–Reiss variogram matrix so that its precision matrix has zeros on non-edges in the graph. Using suitable estimators for the parameters on the edges, our theory provides the first consistent estimator of graph structured Hüsler–Reiss distributions. If the graph is unknown, our method can be combined with recent structure learning algorithms to jointly infer the graph and the corresponding parameter matrix. Based on our methodology, we propose new tools for statistical inference of sparse Hüsler–Reiss models and illustrate them on large flight delay data in the U.S., as well as Danube river flow data.

Keywords: extreme value analysis; multivariate generalized Pareto distribution; sparsity; variogram

1 Introduction

In statistical modelling, conditional independence and graphical models are well-established concepts for analyzing structural relationships in data (Lauritzen, 1996; Wainwright and Jordan, 2008). Particularly important are Gaussian graphical models, also known as Gaussian Markov random fields (Rue and Held, 2005). The graphical structure of a multivariate normal distribution with positive definite covariance matrix Σ can be read off from the zeros of its precision matrix Σ^{-1} .

For risk assessment in fields such as climate science, hydrology, or finance, the primary interest is in extreme observations, with attention to both the marginal tails and the dependence between multiple risk factors. In view of the growing complexity and dimensionality of modern data sets, sparsity and graphical models are becoming crucial notions for the analysis of extremes (e.g., Engelke and Ivanovs, 2021).

There are two different ways of defining graphical models for extreme value distributions. The first is based on max-linear models (Gissibl and Klüppelberg, 2018) and the second one studies multivariate Pareto distributions (Engelke and Hitz, 2020). We follow the second approach since their new notions of conditional independence and extremal graphical models link naturally to the well-known Hammersley–Clifford theorem for density factorizations. Moreover, in the case of tree graphs, Segers (2020) shows that the extremes of regularly varying Markov trees converge to these extremal tree models. Lee and Joe (2018) propose parsimonious models for extreme value copulas; the link with extremal graphical models is made in Asenova et al. (2021).

The class of extremal graphical models with Hüsler–Reiss Pareto distributions is of particular interest. In d dimensions, the parameter of this family is a variogram matrix $\Gamma \in \mathbb{R}^{d \times d}$. Because of their flexibility and stochastic properties, Hüsler–Reiss distributions can be seen as the counterpart of the Gaussian family for multivariate

extremes. In combination with extremal graphical models, the Hüsler–Reiss family constitutes a powerful tool for sparse extreme value modelling, with many open questions still to explore.

For a connected, undirected graph $G = (V, E)$ with nodes $V = \{1, \dots, d\}$ and edges E , Engelke and Hitz (2020) show that such a distribution’s graphical structure can be read off from a set of $(d-1) \times (d-1)$ precision matrices $\Theta^{(k)}$, for $k \in V$. While zeros in $\Theta^{(k)}$ correspond to extremal conditional independence of nodes $i, j \neq k$, the information on edges involving the k th node is encoded only indirectly through the row sums of this matrix. A natural question, appearing also in the discussion of Engelke and Hitz (2020), is if there exists a symmetric approach involving a single $d \times d$ precision matrix.

Statistical inference for Hüsler–Reiss graphical models is limited so far to the simple structures of trees and block graphs (Engelke and Volgushev, 2020; Asenova and Segers, 2023). The parameter matrix Γ is then additive on the graph and the maximum likelihood estimator is an explicit combination of the bivariate estimators on the edges. Since block graphs lack flexibility for general applications, several discussion contributions of Engelke and Hitz (2020) have emphasized the need for estimators suitable for more general graphs.

In this paper, we obtain new theoretical results on Hüsler–Reiss distributions that answer the two open questions above and enable statistical inference on extremal graphical models on decomposable and non-decomposable graphs. Our main contributions are threefold. Firstly, in Section 3, we introduce the Hüsler–Reiss precision matrix $\Theta \in \mathbb{R}^{d \times d}$ as $\Theta_{ij} = \Theta_{ij}^{(k)}$ for some $k \neq i, j$, a definition which—surprisingly—is independent of the particular choice of $k \in V$. This positive semi-definite matrix indeed reflects the sparsity of the extremal graph by zero off-diagonal entries. We give several characterizations of this matrix, one of them as the Moore–Penrose inverse of a projection of the parameter matrix Γ .

Secondly, we study how a Hüsler–Reiss distribution on a given, general graph $G = (V, E)$ and fixed marginal distributions on the edges of the graph can be constructed. Thanks to the new Hüsler–Reiss precision matrix, this task can be framed as a matrix completion problem, which aims to find a conditionally negative definite variogram matrix Γ that has specified values Γ_{ij} in the entries corresponding to the edges $(i, j) \in E$ of graph G and whose precision matrix Θ has zeros in the remaining entries, i.e., $\Theta_{ij} = 0$ for $(i, j) \notin E$. A concrete example in dimension $d = 4$ for the graph in Figure 1(b) is

$$\Gamma = \begin{pmatrix} 0 & 3 & ? & 1 \\ 3 & 0 & 10 & 2 \\ ? & 10 & 0 & ? \\ 1 & 2 & ? & 0 \end{pmatrix}, \quad \Theta = \begin{pmatrix} ? & ? & 0 & ? \\ ? & ? & ? & ? \\ 0 & ? & ? & 0 \\ ? & ? & 0 & ? \end{pmatrix}, \quad (1.1)$$

where the completion problem corresponds to finding the entries with "?". In Section 4, we show that such a completion exists and is unique. Our results can be seen as a semi-definite extension of matrix completion problems for the covariance matrix of Gaussian distributions studied in Speed and Kiiveri (1986) and Bakonyi and Woerdeman (2011).

Thirdly, we leverage our theoretical results to provide effective statistical tools for extremal graphical models. The Hüsler–Reiss precision matrix allows us to represent the maximum (surrogate-)likelihood estimate of Γ on the graph G as the maximizer of the constrained optimization problem

$$\log |\Theta|_+ + \frac{1}{2} \text{tr}(\hat{\Gamma} \Theta), \quad \text{s.t. } \Theta_{ij} = 0 \text{ if } (i, j) \notin E,$$

where $\hat{\Gamma}$ is the empirical variogram (Engelke and Volgushev, 2020) and $|\cdot|_+$ denotes the pseudo-determinant. In Section 5, we prove that the solution to this optimization problem is given by the solution of the above matrix completion problem. We further combine recent structure learning methods (Engelke et al., 2022c) with our completion to yield the first estimator that is jointly sparsistent for the extremal graph and consistent

for the model parameters. Our methodology enables new model assessment plots that allow for model interpretation and comparison of models with different degrees of sparsity. This is illustrated in a case study of large delays in the domestic U.S. air travel network in Section 6. The supplementary material contains another case study, together with mathematical details and proofs.

The Hüsler–Reiss precision matrix and the properties we derive in this paper have already been used for multiple purposes, including the parameterization of sparse statistical models (Röttger and Schmitz, 2023; Röttger et al., 2023a), structure estimation (Engelke et al., 2022c; Wan and Zhou, 2023), efficient statistical inference (Röttger et al., 2023b; Lederer and Oesting, 2023), and a characterization in terms of pairwise interaction models (Lalancette, 2023).

2 Extremal graphical models

2.1 Multivariate generalized Pareto distributions

Multivariate extreme value theory studies the tail behavior of a random vector $X = (X_1, \dots, X_d)$. A first summary of the extremal dependence structure of the bivariate margins for $i, j \in V := \{1, \dots, d\}$ is the extremal correlation

$$\chi_{ij} := \lim_{p \rightarrow 0} \chi_{ij}(p) := \lim_{p \rightarrow 0} \mathbb{P}(F_i(X_i) > 1 - p \mid F_j(X_j) > 1 - p) \in [0, 1], \quad (2.1)$$

defined whenever the limit exists and where F_i is the distribution function of X_i . When $\chi_{ij} > 0$ we say that X_i and X_j are asymptotically dependent, and when $\chi_{ij} = 0$ we speak of asymptotic independence. In the former case, there are two different, but closely related approaches for modelling extremal dependence: through component-wise maxima of independent copies of X leading to max-stable distributions (de Haan and Resnick, 1977); and through threshold exceedances of X resulting in multivariate generalized Pareto distributions (Rootzén and Tajvidi, 2006). Here, we concentrate on the threshold exceedance approach since it is well-suited for graphical modelling (Engelke and Hitz, 2020; Segers, 2020). For statistical models for asymptotic

independence, we refer to Heffernan and Tawn (2004), for instance, and to Papastathopoulos et al. (2017) in the context of extremes of Markov chains as well as Casey and Papastathopoulos (2023) for extremes in decomposable graphical models.

To make abstraction of the univariate marginal distributions and concentrate on the extremal dependence, it is usually assumed that all variables X_i follow a given continuous distribution. Throughout, we use standard exponential margins, that is, $\mathbb{P}(X_i \leq x) = 1 - \exp(-x)$ for $x \geq 0$ and $i \in V$. Let $\mathbf{0}, \mathbf{1}$, and ∞ denote vectors of adequate size with all elements equal to 0, 1, and ∞ respectively. A random vector $Y = (Y_1, \dots, Y_d)$ is said to follow a multivariate generalized Pareto distribution (Rootzén and Tajvidi, 2006) if for any $z \in \mathcal{L} = \{x \in \mathbb{R}^d : x \not\leq \mathbf{0}\}$, we have

$$\mathbb{P}(Y \leq z) := \lim_{u \rightarrow \infty} \mathbb{P}(X - u\mathbf{1} \leq z \mid X \not\leq u\mathbf{1}) = \frac{\Lambda^c(z \wedge \mathbf{0}) - \Lambda^c(z)}{\Lambda^c(\mathbf{0})}, \quad (2.2)$$

for some random vector X , which is then said to be in the domain of attraction of Y ; the simple normalization by subtracting $u\mathbf{1}$ comes from the assumption of exponential margins of X . Multivariate generalized Pareto distributions are threshold-stable in the sense that for a vector $a \geq \mathbf{0}$, the conditional random vector $Y - a$ given $Y \not\leq a$ is again multivariate generalized Pareto. This also implies useful stochastic representations; see Rootzén et al. (2018) for details. The so-called exponent measure Λ is a measure on $[-\infty, \infty)^d \setminus \{-\infty\}$ that is finite on sets bounded away from $-\infty$, and we write $\Lambda^c(z) := \Lambda([-\infty, \infty)^d \setminus [-\infty, z])$. Multivariate generalized Pareto distributions are the only ones that can arise as limits of threshold exceedances as in (2.2).

We assume Λ to be absolutely continuous with respect to the d -dimensional Lebesgue measure and let λ denote its Radon–Nikodym derivative. The set of valid exponent measure densities λ is characterized by the following two properties:

$$\begin{aligned} \lambda(y + t\mathbf{1}) &= \exp(-t)\lambda(y), & \forall t \in \mathbb{R}^d, y \in \mathbb{R}^d, \\ \int_{y_i > 0} \lambda(y) dy &= 1, & \forall i \in V. \end{aligned}$$

Since the distribution of Y is proportional to the restriction of Λ to $\mathcal{L}$, its density f then also exists and is proportional to the exponent measure density λ as $f(y) = \lambda(y) / \Lambda^c(\mathbf{0})$ for all $y \in \mathcal{L}$, since $\Lambda^c(\mathbf{0}) = \Lambda(\mathcal{L})$.

When considering marginal distributions of multivariate generalized Pareto distributions, it is often more useful to work with the limit distributions that arise in (2.2) by replacing X with the sub-vector X_I , also in the conditioning event, for some non-empty $I \subset \{1, \dots, d\}$. The resulting “generalized Pareto marginal distribution”, here denoted $Y_{(I)}$, is supported on $\mathcal{L}_I = \{x \in \mathbb{R}^{|I|} : x \not\leq \mathbf{0}\}$, and its corresponding exponent measure and density are the actual marginals Λ_I and λ_I . Note that $Y_{(I)}$ is equal in distribution to Y_I conditioned on the event $Y_I \not\leq \mathbf{0}$. See Rootzén and Tajvidi (2006, Section 4), or Rootzén et al. (2018, Section 6), for a detailed discussion of these marginals. Due to the choice of standard exponential univariate margins for X , the univariate “generalized Pareto marginals” $Y_{(i)} = (Y_i | Y_i > 0)$ are standard exponential as well. For a discussion how to transform exponent measure densities between different choices of marginals, see Remark S.5.12.

2.2 Hüsler–Reiss distributions

A popular class of multivariate generalized Pareto distributions is the class of Hüsler–Reiss distributions, which are parametrized by symmetric conditionally negative definite matrices, here denoted by Γ (Hüsler and Reiss, 1989; Kabluchko et al., 2009). The set of these so-called variogram matrices is defined as

$$\mathcal{D}_d = \left\{ \Gamma \in \mathbb{R}^{d \times d} : \Gamma = \Gamma^\top, \text{diag}(\Gamma) = \mathbf{0}, v^\top \Gamma v < 0 \ \forall \mathbf{0} \neq v \perp \mathbf{1} \right\}. \quad (2.3)$$

Variogram matrices are closely related to covariance matrices and can be constructed from a centered random vector W with covariance matrix Σ as

$$\Gamma_{ij} = \mathbb{E}(W_i - W_j)^2 = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}, \quad \forall i, j \in V := \{1, \dots, d\}. \quad (2.4)$$

Notably, this so-called covariance transform (Farris et al., 1970) is not injective. One collection of covariance matrices corresponding to a given Γ , used for example in Engelke and Hitz (2020), can be obtained for $k \in V$ as

$$\tilde{\Sigma}^{(k)} = \tilde{\varphi}_k(\Gamma) := \frac{1}{2}(\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij})_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d}. \quad (2.5)$$

The k th row and column of these matrices are identically zero. Let $\Sigma^{(k)} \in \mathbb{R}^{(d-1) \times (d-1)}$ denote the matrices with these entries removed, conveniently indexed by $V \setminus \{k\}$, and $\varphi_k(\cdot)$ the corresponding mapping.

Definition 2.1. Let $\Gamma \in \mathcal{D}_d$ be a variogram matrix, and for any $k \in V$, let $\Sigma^{(k)} = \varphi_k(\Gamma)$, and $\mu^{(k)} = (-\frac{1}{2}\Gamma_{ik})_{i \neq k} \in \mathbb{R}^{d-1}$. A multivariate generalized Pareto random vector $Y = (Y_1, \dots, Y_d)$ is Hüsler–Reiss Pareto distributed with parameter matrix Γ if its exponent measure density λ satisfies

$$\lambda(y; \Gamma) = \frac{\exp(-y_k)}{\sqrt{(2\pi)^{d-1} |\Sigma^{(k)}|}} \exp\left(-\frac{1}{2} \left(\|y_{V \setminus \{k\}} - \mathbf{1}y_k - \mu^{(k)}\|_{\Theta^{(k)}}^2 \right)\right), \quad y \in \mathbb{R}^d,$$

with $\Theta^{(k)} = (\Sigma^{(k)})^{-1}$ and using the notation $\|v\|_M^2 = v^\top M v$. Notably, this expression does not depend on the choice of k (e.g., Engelke et al., 2015, Theorem 2), which is why it is omitted from the left-hand side.

The marginals $Y_{(I)}$ (in the generalized Pareto sense of Section 2.1) of a Hüsler–Reiss Pareto distributed vector with variogram matrix Γ are also Hüsler–Reiss Pareto distributed with variogram matrix $\Gamma_{I \times I}$ (Engelke and Hitz, 2020, Example 7).

The popularity of Hüsler–Reiss distributions has theoretical and practical origins. From known properties about this model class, it becomes obvious that the Hüsler–Reiss family takes the same role among multivariate generalized Pareto distributions as the Gaussian distribution in the non-extreme setting. The results of our paper will further shed light on this analogy. Moreover, Lalancette (2023) recently showed that Hüsler–

Reiss distributions constitute the only pairwise interaction models among multivariate generalized Pareto distributions. From a practical point of view, Hüsler–Reiss distributions have become a standard model in multivariate extreme value modeling because of their flexibility and tractability. Importantly, they are the finite dimensional distributions of the widely used spatial Brown–Resnick (Kabluchko et al., 2009) processes. We refer to Davison et al. (2012) and Engelke and Ivanovs (2021) for overviews of applications of these models.

2.3 Extremal conditional independence

In classical statistics, graphical models are defined through conditional independence relations between components of a random vector (e.g., Lauritzen, 1996). Since the support $\mathcal{L}$ of a multivariate generalized Pareto distribution is not a product space, this definition is impractical in this case. Instead, Engelke and Hitz (2020) introduce a novel notion of extremal conditional independence in terms of the vectors $Y^{(k)} = (Y_1^{(k)}, \dots, Y_d^{(k)})$ for $k \in V := \{1, \dots, d\}$, defined as the multivariate generalized Pareto vector Y conditioned on the event $\{Y_k > 0\}$, and supported on the product spaces $\mathcal{L}^{(k)} = \{x \in \mathcal{L} : x_k > 0\} = \{x \in \mathbb{R}^d : x_k > 0\}$. The extremal version of conditional independence between sub-vectors Y_A and Y_B given Y_C , for non-empty, disjoint sets $A \cup B \cup C = V$, is then defined as

$$\forall k \in V: Y_A^{(k)} \perp\!\!\!\perp Y_B^{(k)} | Y_C^{(k)}, \quad (2.6)$$

and it is denoted by $Y_A \perp_e Y_B | Y_C$.

We consider the index set V as a set of nodes and let $E \subseteq \mathcal{E}(V)$ be the set of undirected edges of a connected graph $G = (V, E)$, where $\mathcal{E}(V)$ denotes the set of all possible edges. Figure 1 shows four different graph structures in increasing generality; see Section S.1.3 for definitions of the related terms. We say that Y follows an extremal graphical model on G if for $i, j \in V$ with $i \neq j$ we have

$$(i, j) \notin E \Rightarrow Y_i \perp_e Y_j | Y_{V \setminus \{i, j\}}.$$

Engelke and Hitz (2020) show that the existence of a positive and continuous exponent measure density λ of Y implies connectedness of the corresponding graph G and the equivalence of (2.6) to the factorization

$$\lambda(y) \lambda_C(y_C) = \lambda_{A \cup C}(y_{A \cup C}) \lambda_{B \cup C}(y_{B \cup C}), \quad y \in \mathcal{L}.$$

If Y is an extremal graphical model on a decomposable graph G , this result, applied recursively, yields a Hammersley–Clifford theorem that states that the exponent measure density of Y factorizes on the graph into lower-dimensional marginal densities; for an extension to unconnected graphs see Engelke et al. (2022b).

The extremal conditional independence $Y_A \perp_e Y_B | Y_C$ has an intuitive interpretation in terms of classical conditional independence. Suppose that we observe an extreme event in variables corresponding to C , that is, $\max_{j \in C} Y_j > 0$, then, conditionally on Y_C , the sub-vectors Y_A and Y_B are independent in the usual sense. For an extremal graphical structure G , this translates into an extremal local prediction property of the form

$$(Y_i | Y_{\delta(i)} = y_{\delta(i)}) \stackrel{d}{=} (Y_i | Y_{\setminus \{i\}} = y_{\setminus \{i\}}), \quad y \in \mathbb{R}^d, \max_{j \in \delta(i)} y_j > 0,$$

where $\delta(i)$ denotes the neighborhood of vertex i in G (see Section S.1.3). Thus, knowing that an extreme event has occurred in the neighborhood of Y_i suffices to know the neighbors of Y_i for predicting its value.

3 Hüsler–Reiss precision matrix

An important property of Gaussian graphical models with covariance matrix Σ is that the conditional independence structure can be read off from the zeros in the (Gaussian) precision matrix Σ^{-1} . Hüsler–Reiss graphical models satisfy a similar property. Indeed, let Y be a Hüsler–Reiss Pareto vector with parameter matrix $\Gamma \in \mathcal{D}_d$ that is an extremal graphical model on the undirected graph $G = (V, E)$. Engelke and Hitz (2020) show that the graph G is necessarily connected and that the graphical structure can be read off

from the precision matrices $\Theta^{(k)} = (\Sigma^{(k)})^{-1}$ for any $k \in V$; in the following we index these $(d-1) \times (d-1)$ matrices by $V \setminus \{k\}$ for the sake of simpler notation. Two distinct nodes $i, j \neq k$ are extremal conditionally independent in the sense that $Y_i \perp_e Y_j \mid Y_{V \setminus \{i,j\}}$ if and only if the corresponding entry $\Theta_{ij}^{(k)}$ is zero. If one of the nodes, say j , is equal to k , extremal conditional independence is equivalent to the row sum $\sum_{l \neq k} \Theta_{il}^{(k)}$ being zero.

Each $\Theta^{(k)}$ thus contains all information on conditional independence of Y . A natural question, already raised in the discussion of Engelke and Hitz (2020), is whether there is a single $d \times d$ precision matrix that contains this information but does not require a choice of k . In fact, such a matrix can be defined as follows.

Definition 3.1. Let $\Gamma \in \mathcal{D}_d$ be a variogram matrix. For $d \geq 3$ define $\Theta \in \mathbb{R}^{d \times d}$ by

$$\Theta_{ij} = \Theta_{ij}^{(k)}, \quad \text{for some } k \in V \setminus \{i, j\},$$

with $\Theta^{(k)}$ as above. In the case $d = 2$, set $\Theta_{11} = \Theta_{22} = -\Theta_{12} = -\Theta_{21} = 1/\Gamma_{12}$.

Lemma 1 in Engelke and Hitz (2020) implies that $\Theta_{ij}^{(k)} = \Theta^{(k)}[k']_{ij}$ for distinct $k, k' \in V \setminus \{i, j\}$. Hence, the matrix Θ is well-defined and represents extremal conditional independence of distinct nodes $i, j \in V$ in the corresponding Hüsler–Reiss model by

$$\Theta_{ij} = 0 \iff Y_i \perp_e Y_j \mid Y_{V \setminus \{i,j\}}. \quad (3.1)$$

While Definition 3.1 is a natural way to jointly represent the information contained in the $\Theta^{(k)}$ matrices, it remains to be shown that the matrix Θ allows for useful mathematical representations. In order to give a first such representation, let $(\cdot)^+$ denote the Moore–Penrose inverse (see Section S.1.1 for details), let I_d be the $d \times d$ identity matrix, write $\mathbf{e}_d = d^{-1}\mathbf{1}$, and let $\Pi = I_d - \mathbf{1}\mathbf{e}_d^\top$ be the $d \times d$ centering matrix, i.e., the projection matrix onto the orthogonal complement of $\mathbf{1}$. Recall (2.4), which can be written as a mapping for arbitrary matrices $S \in \mathbb{R}^{d \times d}$ as

$$\gamma(): S \mapsto \Gamma = \mathbf{1}\text{diag}(S)^\top + \text{diag}(S)\mathbf{1}^\top - 2S, \quad (3.2)$$

see Definition S.5.9 and Lemma S.5.10 for details.

Proposition 3.2. *Let $\Gamma \in \mathcal{D}_d$ and $S \in \mathbb{R}^{d \times d}$ satisfy $\gamma(S) = \Gamma$. Then*

$$\Pi S \Pi = \Pi \left(-\frac{1}{2} \Gamma \right) \Pi. \quad (3.3)$$

Furthermore, the matrix Θ from Definition 3.1 satisfies

$$\Theta = (\Pi S \Pi)^+ \quad (3.4)$$

$$= \lim_{t \rightarrow \infty} (t \mathbf{1} \mathbf{1}^\top + S)^{-1}. \quad (3.5)$$

Since the mapping γ preserves (a)symmetry and Γ is symmetric, the condition $\gamma(S) = \Gamma$ requires S to be a symmetric matrix as well. It turns out that, restricted to symmetric matrices, (3.3) is in fact equivalent to $\gamma(S) = \Gamma$; see Lemma S.5.1.

In practice, the matrix S is usually a (definite or semi-definite) covariance matrix, so it is natural to only consider symmetric matrices here. For the sake of completeness we show in the proof of Proposition 3.2 that (3.3) is a sufficient and necessary condition for (3.4), also allowing asymmetric matrices S . Lemma S.5.8 shows that (3.3) is further equivalent to S being a generalized inverse of Θ in the sense $\Theta S \Theta = \Theta$.

In light of (3.3), an obvious choice of S for a given Γ is setting $S = -\frac{1}{2} \Gamma$, while other interesting choices for S are the matrices $\tilde{\Sigma}^{(k)} = \tilde{\varphi}_k(\Gamma)$, defined in (2.5). Furthermore,

these matrices satisfy $\Pi \left(-\frac{1}{2} \Gamma \right) \Pi = d^{-1} \sum_{k=1}^d \tilde{\Sigma}^{(k)} - t(\Gamma) \mathbf{1} \mathbf{1}^\top$ where $t(\Gamma) = \frac{1}{2} d^{-2} \mathbf{1}^\top \Gamma \mathbf{1}$ is the

largest scalar t such that $d^{-1} \sum_{k=1}^d \tilde{\Sigma}^{(k)} - t \mathbf{1} \mathbf{1}^\top$ is positive semi-definite. Further connections between Γ , $\tilde{\Sigma}^{(k)}$ and Θ are investigated in Wan and Zhou (2023, Propositions 3.1 and 3.2).

To better understand representation (3.4), let $\mathcal{P}_d^1 \subset \mathbb{R}^{d \times d}$ denote the set of symmetric positive semi-definite matrices with kernel equal to $\text{span}(\{\mathbf{1}\})$. Let σ and θ be the mappings from a variogram matrix Γ to the corresponding matrices Σ and Θ , respectively:

$$\begin{aligned} \sigma: \quad \Gamma &\mapsto \Sigma := \Pi(-\frac{1}{2}\Gamma)\Pi, \\ \theta: \quad \Gamma &\mapsto \Theta := \Sigma^+ = \sigma(\Gamma)^+ = (\Pi(-\frac{1}{2}\Gamma)\Pi)^+. \end{aligned} \tag{3.6}$$

Proposition 3.3. *The mappings σ and θ are homeomorphisms between $\mathcal{D}_d$ and $\mathcal{P}_d^1$. Using $\gamma(\cdot)$ defined in (3.2), their inverses are*

$$\sigma^{-1}(\Sigma) = \gamma(\Sigma), \quad \theta^{-1}(\Theta) = \gamma(\Theta^+).$$

From this proposition it follows that the matrix Θ from Definition 3.1 is always positive semi-definite with kernel equal to $\text{span}(\{\mathbf{1}\})$. Furthermore, the class of Hüsler–Reiss distributions, parametrized by the set $\mathcal{D}_d$ of conditionally negative definite matrices, can just as well be parametrized by $\mathcal{P}_d^1$, interpreted either in the role of Θ or Σ . As discussed after Proposition 3.4, the matrix Σ is the degenerate covariance matrix of a particular transformation of the Hüsler–Reiss Pareto vector Y . Similarly to the Gaussian case, the precision matrix Θ can be obtained from this covariance matrix as $\Theta = \Sigma^+$, using the Moore–Penrose inverse due to its singularity.

For a given Γ , the characterizations in Proposition 3.2 give a straightforward way to

compute Θ in Definition 3.1 as $\Theta = \theta(\Gamma) = (\Pi(-\frac{1}{2}\Gamma)\Pi)^+$ and thus retrieve the conditional independence structure of the corresponding extremal graphical model. In a Gaussian model, another property of the precision matrix is that it can be used to express its probability density function in a concise way. The following result shows that a similar expression is also possible for Hüsler–Reiss distributions.

Proposition 3.4. Let $\Gamma \in \mathcal{D}_d$ and $\Theta = \theta(\Gamma)$. Then the exponent measure density $\lambda(\cdot; \Gamma)$ from Definition 2.1 can be expressed as

$$\lambda(y; \Gamma) = c_1 \cdot c_2 \cdot \exp\left(-\frac{1}{2} y^\top \Theta y + y^\top r_\Theta - y^\top \mathbf{e}_d\right), \quad y \in \mathbb{R}^d, \quad (3.7)$$

$$\text{with } r_\Theta = -\frac{1}{2} \Theta \Gamma \mathbf{e}_d, \quad c_1 = (2\pi)^{-(d-1)/2} (d^{-1} |\Theta|_+)^{1/2}, \quad \text{and } c_2 = \exp\left(-\frac{1}{8} \mathbf{e}_d^\top (\Gamma \Theta \Gamma + 2\Gamma) \mathbf{e}_d\right)$$

This expression is remarkably similar to a degenerate Gaussian density on the hyperplane $\{\mathbf{1}\}^\perp$, differing only by the value of the constant and the term $\exp(-y^\top \mathbf{e}_d)$. In fact, when restricting the support to the half-space $\{x \in \mathbb{R}^d : \mathbf{1}^\top x \geq 0\}$, this density is proportional to the probability density function of a random variable $Y_1 = W_\Sigma + R\mathbf{1}$, with $R \sim \text{Exp}(1)$ and W_Σ a degenerate normal vector with covariance matrix $\Sigma = \Pi(-\frac{1}{2}\Gamma)\Pi$, independent of R .

A careful look at other characterizations of λ such as in Wadsworth and Tawn (2014), or the slightly more general definitions of the Hüsler–Reiss exponent measure density in Ho and Dombry (2019) and Kiriliouk et al. (2018), shows that the precision matrix Θ appears naturally in these parameterizations as well.

Remark 3.1. Besides the extremal conditional independence structure and exponent measure density, the precision matrix Θ is also useful to describe other stochastic properties of Hüsler–Reiss distributions. Based on the results of the present paper, Röttger et al. (2023b) and Röttger and Schmitz (2023) show that multivariate total positivity of order two, a notion of positive dependence, can be encoded as $\Theta_{ij} \leq 0$ for all $i \neq j$. Engelke et al. (2022c) suggest that the precision matrix can be used to estimate extremal graphical structures by penalizing its L^1 -norm $\|\Theta\|_1$, and a similar approach is taken in Wan and Zhou (2023) and Lederer and Oesting (2023). Compared to the Gaussian case, a difficulty of dealing with Θ in mathematical derivations and statistical implementations is that it is not invertible.

4 Matrix completion problems

A well-studied problem related to Gaussian graphical models is the task of constructing a model with a given graphical structure and specified marginal distributions on the fully connected subsets of vertices, or equivalently, to complete a partially defined covariance matrix such that its precision matrix has zeros where there are no edges in the specified graph (Speed and Kiiveri, 1986; Bakonyi and Woerdeman, 2011). For extremal Hüsler–Reiss models, the same problem can be posed, expressed as a matrix completion problem on the variogram and precision matrix.

To formalize the notion of a partially specified matrix, we define the set $\mathring{\mathbb{R}}^d := \mathbb{R}^d \cup \{\text{"?"}\}$, consisting of the real numbers and the placeholder "?" for unspecified values (see e.g., Bakonyi and Woerdeman, 2011, for this use of "?"). For an undirected graph $G = (V, E)$, a matrix is said to be “specified on G ” if it is specified on the diagonal and the entries corresponding to the edges of G . A matrix is said to be “partially conditionally negative definite” if it is symmetric, its diagonal is fully specified, and all fully specified principal submatrices are conditionally negative definite (a principal submatrix is any submatrix obtained by removing the same index set from both the columns and rows of the matrix). In computations involving partially specified matrices, an entry in the result is “?” as soon as any of the entries used to compute it is itself “?”. For definitions of graph theoretical concepts used in this section, we refer to Section S.1.3.

Definition 4.1. Let $G = (V, E)$ be an undirected graph and let $\overset{\circ}{\Gamma} \in \overset{\circ}{\mathbb{R}}^{d \times d}$ be a partially conditionally negative definite matrix, specified on G . The corresponding matrix completion problem is to find a conditionally negative definite matrix $\Gamma \in \mathcal{D}_d$ and $\Theta = \theta(\Gamma)$ such that

$$\begin{aligned} \Gamma_{ij} &= \overset{\circ}{\Gamma}_{ij} & \forall (i, j) \in \bar{E}, \\ \Theta_{ij} &= 0 & \forall (i, j) \notin \bar{E}, \end{aligned} \quad (4.1)$$

where $\bar{E}$ denotes the edge set E augmented by the diagonal entries $\{(i, i) : i \in V\}$.

An example for such a matrix completion problem on the graph in Figure 1(b) is given in (1.1). In the Gaussian case, a similar problem can be posed with the covariance matrix Σ instead of Γ and with $\Theta = \Sigma^{-1}$. To this positive definite completion problem, an explicit solution for decomposable graphs and a convergent algorithm for general graphs is given in Speed and Kiiveri (1986). In this section we discuss semi-definite matrix completion problems for Hüsler–Reiss models as in Definition 4.1, starting from simple graph structures such as trees and finishing with general graphs. We assume throughout that the graph G is connected since only those can be associated to non-degenerate Hüsler–Reiss models; see Section 2.3.

Example 4.2. Block graphs are simple graph structures where the separator sets contain only single nodes. Trees are a special case of this class where all cliques consist of exactly two nodes. The particular structure of block graphs makes them appealing for the construction of parametric models and for statistical inference (Engelke and Hitz, 2020; Asenova et al., 2021). In fact, this structure also yields a simple explicit solution to the matrix completion problem in Definition 4.1. If $\overset{\circ}{\Gamma}$ is a partially conditionally negative definite matrix on the connected block graph $G = (V, E)$, then a unique completion exists (Engelke and Hitz, 2020, Proposition 4) and can be expressed as

$$\Gamma_{ij} = \sum_{(s,t) \in \text{path}(i,j)} \overset{\circ}{\Gamma}_{st}, \quad (4.2)$$

where $\text{path}(i, j)$ is the unique shortest path between i and j in G ; see also Engelke and Volgushev (2020) and Asenova and Segers (2023). Using this result, the missing entries in (1.1) can be computed to be $\Gamma_{13} = \Gamma_{31} = 13$ and $\Gamma_{34} = \Gamma_{43} = 12$.

4.1 Decomposable graphs

In this section, a solution to the matrix completion problem will be given for connected, decomposable graphs. First, consider the simplest (non-trivial) example from this class of graphs, a graph consisting of exactly two cliques. Recall $\Sigma^{(k)}, \tilde{\Sigma}^{(k)}$ and φ_k from (2.5), γ from (3.2), and observe that $\varphi_k^{-1}(\Sigma^{(k)}) = \gamma(\tilde{\Sigma}^{(k)})$.

Lemma 4.3. *Let $G = (V, E)$ be a connected decomposable graph consisting of two cliques C_1 and C_2 , separated by $D_2 = C_1 \cap C_2 \neq \emptyset$. Let $\overset{\circ}{\Gamma}$ be a partially conditionally negative definite matrix, specified on G . For some $k \in D_2$, let $\overset{\circ}{\Sigma}^{(k)} = \varphi_k(\overset{\circ}{\Gamma})$ and let $\Sigma^{(k)}$ be its unique positive definite completion with graphical structure $G|_{V \setminus \{k\}}$. Then*

$$\Gamma := \varphi_k^{-1}(\Sigma^{(k)}) = \gamma(\tilde{\Sigma}^{(k)}),$$

is the unique solution of the matrix completion problem 4.1 for $\overset{\circ}{\Gamma}$ and graph G .

The existence and uniqueness of such a positive definite completion for general graph structures is shown in Speed and Kiiveri (1986), and a short proof of the case with two cliques is given in Lemma S.5.13. Using this result, a completion for a general decomposable graph G can be constructed as follows. Let $\{C_1, \dots, C_N\}$ be the cliques of G , ordered according to the running intersection property (cf. Section S.1.3), and, without losing generality, assume that the vertices in V are ordered accordingly, in the

sense $i \leq j$ for all $i \in C_k$ and $j \in C_l$ with $k < l$. Let $\Gamma_1 = \overset{\circ}{\Gamma}|_{C_1}$, $V_n = C_1 \cup \dots \cup C_n$, and

$K_n = V_n \times V_n$. Further, for $n = 2, \dots, N$, iteratively define $\overset{\circ}{\Gamma}_n \in \mathbb{R}^{|V_n| \times |V_n|}$ with entries

$$(\overset{\circ}{\Gamma}_n)_{ij} = \begin{cases} (\Gamma_{n-1})_{ij} & (i, j) \in K_{n-1}, \\ \overset{\circ}{\Gamma}_{ij} & (i, j) \in \bar{E}|_{V_n \setminus K_{n-1}}, \\ \text{"?"} & \text{otherwise,} \end{cases}$$

and $\Gamma_n = \mathfrak{C}(\Gamma_n^\circ) \in \mathbb{R}^{|V_n| \times |V_n|}$, where $\mathfrak{C}(\cdot)$ denotes the completion from Lemma 4.3.

Proposition 4.4. *Let $G = (V, E)$ be a connected, decomposable graph and Γ° a partially conditionally negative definite matrix, specified on G . Then there exists a unique solution to the matrix completion problem 4.1 with Γ° and G . This solution can be computed explicitly as Γ_N in the construction above.*

The class of decomposable graphs is a significant extension of the class of block graphs. For instance, a non-decomposable graph can be approximated in a non-trivial fashion by computing a so-called decomposable completion. In contrast, the only block graph completion of any biconnected graph (i.e., a graph that remains connected after removal of any one vertex) is already the complete graph, since the only biconnected block graph is the complete graph, and adding edges to a graph preserves connectivity. See for instance Lauritzen (1996, Section 5.3) for further details about decomposable graphical models.

Example 4.2. Figure 2 shows an example of the matrix completion algorithm from Proposition 4.4. Starting with the clique $\{1, 2, 3\}$, the missing values are computed clique by clique. Edges whose corresponding matrix entries were already computed, are considered as part of the graph in subsequent steps, such that each computation is a direct application of Lemma 4.3. Thanks to the running intersection property, the required conditional independence structure is preserved.

4.2 General graphs

The class of general connected graphs is much larger than the class of decomposable ones; see Figure 1(d) for an example. In applications, when the graph is estimated from data without restrictions, non-decomposable structures often arise. The following results provide a more general but slightly weaker solution to the matrix completion problem in Definition 4.1, where, as before, $\bar{E}$ is equal to $E \cup \{(i, i) : i \in V\}$.

Proposition 4.6. *Let $G = (V, E)$ be a connected graph and let $\overset{\circ}{\Gamma} \in \mathbb{R}^{\overset{\circ}{d} \times \overset{\circ}{d}}$ be specified on G , such that there exists a fully specified conditionally negative definite matrix that agrees with $\overset{\circ}{\Gamma}$ on the entries $(i, j) \in \bar{E}$. Then there exists a unique conditionally negative definite matrix Γ that solves the matrix completion problem (4.1) for $\overset{\circ}{\Gamma}$ and G .*

This result provides the same theoretical existence of a unique solution as Proposition 4.4 and allows the definition of the following mapping. We use the notation $\Gamma|_G \in \mathbb{R}^{\overset{\circ}{d} \times \overset{\circ}{d}}$ to denote the restriction of a fully specified matrix Γ to a graph $G = (V, E)$, in the sense

$$(\Gamma|_G)_{ij} = \begin{cases} \Gamma_{ij} & (i, j) \in \bar{E}, \\ \text{"?"} & (i, j) \notin \bar{E}. \end{cases}$$

Definition 4.7. *For a connected graph $G = (V, E)$ let $\overset{\circ}{\mathcal{D}}_G = \{(\Gamma'|_G) : \Gamma' \in \mathcal{D}_d\}$ be the restriction of conditionally negative definite matrices to G . The function*

$$\mathfrak{C}_G() : \overset{\circ}{\mathcal{D}}_G \rightarrow \mathcal{D}_d, \quad \overset{\circ}{\Gamma} \mapsto \Gamma,$$

maps a partial matrix $\overset{\circ}{\Gamma} \in \overset{\circ}{\mathcal{D}}_G$ to its unique completion Γ satisfying the matrix completion problem in Definition 4.1 with respect to G .

The uniqueness of this completion can be understood intuitively by counting the number of parameters and constraints involved in (4.1). Notably, the existence of any

conditionally negative definite completion of $\overset{\circ}{\Gamma}$ is sufficient for the existence of a graphical completion. In the decomposable case, this can be guaranteed by verifying the definiteness of all fully specified principal submatrices, but in the general case this criterion does not work, as the following counter-example shows.

Example 4.8. For $d \geq 4$, let $\overset{\circ}{\Gamma} \in \mathbb{R}^{\overset{\circ}{d} \times \overset{\circ}{d}}$ be a partial matrix on the d -dimensional ring graph with entries $\overset{\circ}{\Gamma}_{ij} = 1$ if $|i - j| = 1$, $\overset{\circ}{\Gamma}_{1d} = \overset{\circ}{\Gamma}_{d1} = (2d)^2$, zeros on the diagonal, and "???" elsewhere. Then all fully specified principal submatrices of this $\overset{\circ}{\Gamma}$ are conditionally negative definite, but there exists no conditionally negative definite completion of the entire matrix (see Section S.5.5.3 for details).

In general, computing the completion is less straightforward than in the decomposable case, but the following algorithm follows from the proof of Proposition 4.6. This procedure is similar to the one described in Speed and Kiiveri (1986, Section 4), using the above results instead of positive definite matrix completions.

Corollary 4.9. Let G and $\overset{\circ}{\Gamma}$ be as in Proposition 4.2, and let $\Gamma_0 \in \mathcal{D}_d$ be such that $(\Gamma_0)|_G = \overset{\circ}{\Gamma}$. For some fixed $m \in \mathbb{N}$, let $G_i = (V, E_i)$, $i = 1, \dots, m$, be decomposable graphs such that $\bigcap_i E_i = E$ and define $(\Gamma_n)_{n \geq 1}$ recursively by

$$\Gamma_n = \mathfrak{C}_{G_n}((\Gamma_{n-1})|_{G_n}),$$

with $t_n \in \{1, \dots, m\}$, $t_n \equiv n \pmod{m}$, and $\mathfrak{C}_{G_n}()$ computed as in Proposition 4.4. Then Γ_n converges to the unique completion $\mathfrak{C}_G(\overset{\circ}{\Gamma})$ as $n \rightarrow \infty$.

Remark 4.10. An easy way to construct a suitable set of graphs $\{G_1, \dots, G_m\}$ is to put $\{e_1, \dots, e_m\} = \mathcal{E}(V) \setminus E$ and use $E_i = \mathcal{E}(V) \setminus \{e_i\}$, where $\mathcal{E}(V)$ denotes the set of all possible edges, see Section S.1.3. However, this choice leads to slow convergence

since each iteration only updates one entry in Γ while requiring inversion of a matrix of dimension $(d-3) \times (d-3)$. Better performance can be achieved by choosing the G_i to be decomposable completions of G , with separator sets being as small as possible; see Baz et al. (2022) for details and further optimizations.

In order to apply Corollary 4.9, an initial (non-graphical) completion of Γ° , to be used as Γ_0 , is required. The problem of finding such a matrix is also known as the Euclidean distance matrix completion problem, with solution algorithms for example in Bakonyi and Johnson (1995) and Fang and O’Leary (2012).

Example 4.11. To illustrate the algorithm from Corollary 4.9, consider the matrix Γ_0 below, which does not have any non-trivial graphical structure, and its “completion” Γ , whose conditional independence structure is described by the graph in Figure 3.

$$\Gamma_0 = \begin{pmatrix} 0 & 0.23 & \underline{0.08} & 0.09 & 0.21 \\ 0.23 & 0 & 0.14 & \underline{0.23} & \underline{0.19} \\ \underline{0.08} & 0.14 & 0 & 0.11 & \underline{0.20} \\ 0.09 & \underline{0.23} & 0.11 & 0 & 0.16 \\ 0.21 & \underline{0.19} & \underline{0.20} & 0.16 & 0 \end{pmatrix} \mapsto \Gamma = \begin{pmatrix} 0 & 0.23 & \underline{0.17} & 0.09 & 0.21 \\ 0.23 & 0 & 0.14 & \underline{0.20} & \underline{0.35} \\ \underline{0.17} & 0.14 & 0 & 0.11 & \underline{0.26} \\ 0.09 & \underline{0.20} & 0.11 & 0 & 0.16 \\ 0.21 & \underline{0.35} & \underline{0.26} & 0.16 & 0 \end{pmatrix}$$

These two matrices differ only in the highlighted entries, corresponding to non-edges in G . During the computation of Γ , the entries of Γ_0 that correspond to edges in G do not change, and the convergence of the remaining entries in Θ to zero is plotted in Figure 3. It can be seen that the entries corresponding to edges (2, 5) and (3, 5) stay at zero (up to numerical precision of magnitude 10^{-15}) after the first few iterations, whereas the maximum of entries (2, 4) and (1, 3) converges to zero at a slower rate. This observation is in line with the fact that vertices $\{1, 2, 3, 4\}$ induce the chordless cycle that makes the graph non-decomposable.

5 Statistical inference

Estimation of a Hüsler–Reiss parameter matrix Γ that is an extremal graphical model on a given graph $G = (V, E)$ is currently restricted to the simple structures of trees (Engelke and Volgushev, 2020; Hu et al., 2022), latent trees (Asenova et al., 2021) and block graphs (Engelke and Hitz, 2020; Asenova and Segers, 2023). In the discussion of Engelke and Hitz (2020), this has been pointed out as too restrictive in practice, since many data sets require more general graph structures.

In this section we solve this issue by applying our results on matrix completions for variograms. In particular, we show how any consistent estimator $\hat{\Gamma}$ can be transformed into a consistent estimator $\hat{\Gamma}^G$ with desired graph structure. When the graph is unknown, it can be recovered by existing structure learning methods $\hat{G}$. Our completion then produces the first estimator $\hat{\Gamma}^{\hat{G}}$ that jointly estimates the graph and the variogram matrix for general graphs G in a consistent way.

5.1 Matrix completion as likelihood optimization

For a more statistical perspective on the matrix completions in Section 4, we first characterize them as constrained maximum likelihood estimators; see Uhler (2017) for the Gaussian case. For a variogram matrix $\bar{\Gamma} \in \mathcal{D}_d$ and a connected graph $G = (V, E)$, we will show that the maximizer of the positive semi-definite Gaussian log-likelihood

$$\log|\Theta|_+ + \frac{1}{2} \text{tr}(\bar{\Gamma}\Theta), \quad (5.1)$$

under suitable graph constraints, is equal to our completion operator $\mathfrak{C}_G()$ from

Definition 4.7 applied to $\bar{\Gamma}|_G$. The connection to likelihood estimation for Hüsler–Reiss distributions arises by choosing for $\bar{\Gamma}$ the empirical variogram $\hat{\Gamma}$ (Engelke and Volgushev, 2020). In this case, (5.1) is the (surrogate) log-likelihood of the Hüsler–Reiss model parameterized in terms of the precision matrix Θ (Röttger et al., 2023b, Section

5.1); see Section S.2.1 for a simple derivation. Recall $\bar{E} = E \cup \{(i, i) : i \in V\}$ for $E \subseteq \mathcal{E}(V)$ as well as the map θ and its inverse in Proposition 3.3.

Proposition 5.1. *Let $\bar{\Gamma} \in \mathcal{D}_d$ be a variogram matrix and $G = (V, E)$ a connected graph. Then $\mathfrak{C}_G(\bar{\Gamma}|_G) = \theta^{-1}(\bar{\Theta}_G)$ where $\bar{\Theta}_G$ is the unique maximizer of (5.1) over all Hüsler–Reiss precision matrices $\Theta \in \mathcal{P}_d^1$ under the constraint that $\Theta_{ij} = 0$ for all $(i, j) \notin \bar{E}$.*

We note that the result holds for any variogram matrix $\bar{\Gamma}$, but only for the empirical variogram $\hat{\Gamma}$ there is an interpretation in terms of maximum (surrogate) likelihood estimation. In practice, solving this optimization provides an alternative to Corollary 4.1 to compute the graphical completion of a matrix. We leave this approach for future research.

5.2 Consistency

In order to show consistency results, a useful property of the completion $\mathfrak{C}_G(\cdot)$ is that it is a continuous mapping from the space $\mathring{\mathcal{D}}_G$ of partially specified variogram matrices to the space of variogram matrices $\mathcal{D}$.

Lemma 5.2. *The mapping $\mathfrak{C}_G(\cdot)$ from Definition 4.7 is continuous.*

By continuity of $\mathfrak{C}_G(\cdot)$, a consistent estimator on the edges of a Hüsler–Reiss graphical model can be extended to a consistent estimator of the whole parameter matrix.

Theorem 5.3. *Consider a Hüsler–Reiss graphical model with graphical structure $G = (V, E)$ and variogram matrix Γ . Let $\mathring{\Gamma} \equiv \mathring{\Gamma}_n$ be an estimator sequence of $\Gamma|_G$ which satisfies $\text{diag}(\mathring{\Gamma}) \equiv \mathbf{0}$, is symmetric, and is consistent in the sense that for all $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\max_{(i,j) \in E} \left| \mathring{\Gamma}_{ij} - \Gamma_{ij} \right| < \varepsilon\right) \rightarrow 1, \quad n \rightarrow \infty. \quad (5.2)$$

Then with probability tending to one there exists a completion of $\overset{\circ}{\Gamma}$, that is,
 $\mathbb{P}(\overset{\circ}{\Gamma} \in \overset{\circ}{\mathcal{D}}_G) \rightarrow 1$ as $n \rightarrow \infty$. Let $\hat{\Gamma}^G = \mathfrak{C}_G(\overset{\circ}{\Gamma})$ denote this completion and $\hat{\Theta}^G$ the
corresponding precision matrix (and set both matrices to the zero matrix if the
completion does not yet exist). Then $\hat{\Gamma}^G$ is a consistent estimator for Γ with the correct
graph structure, that is, $\hat{\Theta}_{ij}^G = 0$ if $(i, j) \notin \bar{E}$ and for all $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{(i,j) \in V \times V} |\hat{\Gamma}_{ij}^G - \Gamma_{ij}| < \varepsilon\right) \rightarrow 1, \quad n \rightarrow \infty.$$

Theorem 5.3 provides the first consistent estimator of a Hüsler–Reiss variogram matrix that respects the structure of a general graph $G = (V, E)$. Indeed, the only ingredient that is needed is a consistent estimator of Γ on the edge set E . There are many different possibilities for such estimators in the literature. A natural and computationally efficient estimator is the empirical variogram (Engelke et al., 2015; Engelke and Volgushev, 2020), which for the Hüsler–Reiss distributions is the empirical version of the parameter matrix Γ . Other proposals include M-estimators (Einmahl et al., 2012, 2016; Lalancette et al., 2021), proper scoring rules (de Fondeville and Davison, 2018) and likelihood methods. The latter have the advantage that they can incorporate censoring of components of the data that are not extreme (Ledford and Tawn, 1996; Wadsworth and Tawn, 2014). Pairwise likelihood methods (Padoan et al., 2010) reduce the high computational cost of censoring.

In order to guarantee that even for a fixed sample size n there exists a completion, it can be advantageous to start with a consistent estimator $\hat{\Gamma} \in \mathcal{D}$ of the entire parameter matrix Γ . Since $\hat{\Gamma}|_{G \in \overset{\circ}{\mathcal{D}}_G}$ by definition, Proposition 4.6 ensures that

$$\hat{\Gamma}^G = \mathfrak{C}_G(\hat{\Gamma}|_G) \quad (5.3)$$

exists and is a consistent estimator with graph structure G . In larger dimensions $d = |V|$, estimating all entries of Γ by likelihood optimization or censoring is often infeasible, and

the only option is then the empirical variogram or estimators based on simple summary statistics such as extremal coefficients (Einmahl et al., 2018).

If the graph G is sparse, then there is an efficient alternative to estimating every entry of Γ , which requires only estimation on all cliques separately; see Section S.2.2.

Furthermore, in Section S.2.3, we perform a simulation study that illustrates the performance gains achieved by this method, while yielding a similar quality of estimation of Γ .

The graph G is in practice often unknown and has to be estimated from the data. Consistent structure estimation methods for extremal graphs exist for trees (Engelke and Volgushev, 2020; Hu et al., 2022), and for general graphs based on lasso-type L^1 penalization (Engelke et al., 2022c; Wan and Zhou, 2023). The latter paper proposes the `EGlearn` method and shows, under certain conditions, its sparsistency even in the high-dimensional case where the dimension may grow with the sample size. More precisely, if G is the graph implied by the zero entries in the true Θ , and $\hat{G} = (V, \hat{E})$ is the estimated graph from `EGlearn` then

$$\mathbb{P}(\hat{G} = G) \rightarrow 1, \quad n \rightarrow \infty. \quad (5.4)$$

While they consistently recover a general graph structure, they do not obtain an estimate of the Γ matrix on the estimated graph structure.

Our theory complements the structure estimation in Engelke et al. (2022c). In combination, we are now able to estimate jointly any graph structure and the corresponding Hüsler–Reiss parameter matrix consistently.

Corollary 5.4. *Consider a Hüsler–Reiss graphical model with graphical structure $G = (V, E)$ and variogram matrix Γ . Let $\hat{\Gamma}$ be a consistent estimator for Γ and $\hat{G} = (V, \hat{E})$ a sparsistent estimator of G as in (5.4). If $\hat{\Gamma}^{\hat{G}} = \mathfrak{C}_{\hat{G}}(\hat{\Gamma}|_{\hat{G}})$ is the completion on $\hat{G}$ and $\hat{\Theta}^{\hat{G}}$ its precision matrix, then $\hat{\Gamma}^{\hat{G}}$ is a consistent and sparsistent estimator for Γ , that is, for all $\varepsilon > 0$,*

$$\mathbb{P}(\max_{(i,j) \in V \times V} |\hat{\Gamma}_{ij}^{\hat{G}} - \Gamma_{ij}| < \varepsilon, \hat{\Theta}_{ij}^{\hat{G}} = 0 \forall (i,j) \notin \bar{E}) \rightarrow 1, \quad n \rightarrow \infty.$$

In Corollary 5.4, not only is $\hat{\Gamma}^{\hat{G}}$ a consistent estimator of Γ , it is also, with probability going to one, a graphical model with respect to the true graph G .

6 Application

Evaluation of the risk of climate extremes such as heatwaves (e.g., Reich et al., 2014) or floods (e.g., Asadi et al., 2015; Cooley and Thibaud, 2019) is the bread and butter of extreme value analysis, and our methods are perfectly suitable for such data. In order to illustrate our methods on a widely used data set and to underline their applicability to a broad range of domains, we study the river flow data of Asadi et al. (2015) in the Supplementary Material (Section S.4). In this section, we extend the range of possible applications and use our methodology to study large flight delays in the U.S. flight network. Excessive delays have a variety of negative effects, ranging from inconveniences for passengers and congestion of critical airport infrastructure to financial losses for involved parties.

6.1 Data and exploratory analysis

The United States Bureau of Transportation Statistics ¹ provides records of domestic flights in the U.S. that are operated by major carriers (at least 1% market share) at airports accounting for at least 1% of domestic enplanements. We use this data from 2005 to 2020, selecting only airports from the contiguous U.S. with a minimum of 1000 flights per year. For each airport, we compute the accumulated (positive) flight delays (in minutes) on a given day. As there are a few days for which no data are available, this results in a data set with $n = 5601$ observations $x_1, \dots, x_n \in \mathbb{R}^d$ of daily accumulated flight delays for the 16 years at $d = 169$ airports. This pre-processed data set is available in the R-package `graphicalExtremes` (Engelke et al., 2022a).

The models in this paper are suitable for variables where the largest observations are asymptotically dependent. For our data, such dependence between the largest daily

accumulated flight delays at different airports may result from several factors. For instance, large delays at a hub in the network may induce delays at other airports that have frequent connections to that hub. On the other hand, independently of flight connections, meteorological events such as severe storms or heavy snowfall can cause simultaneous delays at airports in the same geographical area.

In order to find groups of airports at risk of simultaneous large delays, we first run a k -medoids clustering (Kaufman and Rousseeuw, 2009), where similarities are defined in terms of the strength of extremal dependence. Clustering is frequently used in the extreme value literature to identify regions that are homogeneous in terms of dependence properties (e.g., Bernard et al., 2013; Saunders et al., 2020; Vignotto et al., 2021). We follow these approaches but use a different dissimilarity measure, namely the empirical version $\hat{\chi}_{ij}(p)$ of the extremal correlation (2.1) at probability level $p = 0.95$. The clusters and subsequent results are stable with respect to the exact probability level in a range of about $p \in [0.8, 0.95]$, yielding both a sufficiently high threshold and enough exceedances.

The resulting clusters are shown in Figure 4, for an ad-hoc choice of four clusters. Even though no information on the airport locations is used, the resulting groups exhibit strong geographical proximity, supporting our intuition on the importance of flight connectivity and meteorological factors. Figure S.3 shows that the extremal dependence of large delays is much stronger within clusters than between different clusters. In the following we focus on the southern cluster around Texas, consisting of 29 airports, which is particularly stable even when modifying the number k of clusters and threshold value p ; similar analyses can be conducted for the other clusters. The IATA codes of this cluster are given in Table S.4. The existence of a (regular) flight connection between two airports defines a natural graph, denoted by G_{flight} and shown in the left panel of Figure 5.

6.2 Graphical modeling

The flight graph G_{flight} is based on domain knowledge and is certainly a good first candidate for statistical modeling. It is however not necessarily the best graph in terms of conditional independence properties. We therefore also consider extremal graph structures estimated from data. To evaluate the fitted models out-of-sample, all rows with missing values were removed, and the data was split into a training set with 1764 observations (2005-01-01 to 2010-12-31) for estimation, and a validation set with 1839 observations (2011-01-01 to 2020-12-31) for selection of tuning parameters and model comparisons. As a base estimator, we use the empirical extremal variogram $\hat{\Gamma}$ computed at probability threshold $p = 0.95$ on the training data.

The first, sparse graph is obtained as the minimum spanning tree with weight matrix $\hat{\Gamma}$ as proposed in Engelke and Volgushev (2020); this tree is denoted by $\hat{T}$ and is shown in the center of Figure 5. Alternatively, the empirical extremal correlation $\hat{\chi}$ as weight matrix can also recover an underlying tree (Engelke and Volgushev, 2020; Hu et al., 2022).

As a second family of estimated extremal graph structures $\hat{G}_\rho$ we apply the `EGlearn` algorithm (Engelke et al., 2022c) to the flights data set with different regularization parameters $\rho \geq 0$. The latter governs the amount of sparsity in the estimated graph, where larger ρ values correspond to sparser graphs. `EGlearn` produces a whole sequence of estimated graphical models, but without estimates of the corresponding parameter matrices.

We apply our methodology to fit a Hüsler–Reiss multivariate generalized Pareto distribution with different extremal graphical structures. As before, the empirical variogram $\hat{\Gamma}$ acts as base estimator. Applying the completion operator $\mathfrak{C}_G()$ to the restriction of $\hat{\Gamma}$ to a graph G as in (5.3), we obtain graph structured estimators $\hat{\Gamma}^{G_{\text{flight}}}, \hat{\Gamma}^{\hat{T}}$ and $\hat{\Gamma}^{\hat{G}_\rho}$ for the flight graph, the extremal tree and the ρ -regularized general graph, respectively. We note that, previously, only parameter estimation on the tree was

possible through the tree metric property (4.2), but not on the more general graphs. Our method thus allows to estimate the Hüsler–Reiss parameters of an extremal graphical model based on a graph learned from the data, the only restriction being that the graph needs to be connected.

A first benefit of our approach is that it complements structure learning methods such as `EGlearn` by allowing for a data-driven sparsity tuning. One approach would be to compare the Hüsler–Reiss log-likelihood values of the different parameter matrices $\hat{\Gamma}^{\hat{G}_\rho}$, for instance by AIC or BIC applied to the training data. Instead, we directly compare the log-likelihood values on the validation set and plot them against the tuning parameter ρ in Figure 6. The likelihood is computed based on the Hüsler–Reiss Pareto density, and since we are mainly interested in the dependencies between different airports, the univariate marginals were normalized to the standard exponential scale using the empirical distribution functions; Figure S.4 shows the univariate shape parameters before normalization. The structural estimators in this section are all part of the family of Hüsler–Reiss distributions, allowing the comparison of the log-likelihood values, in the same way as normal distributions with different degrees of sparsity are compared in classical graphical modelling (e.g. Friedman et al., 2007). The best fitting model on the independent validation data set comes from `EGlearn` at $\rho^* = 0.15$, and the corresponding graph with 101 edges is shown on the right-hand side of Figure 5. The flight graph performs significantly worse, and the tree seems to be too sparse. The Hüsler–Reiss precision matrix defined in Section 3 can be used to track how sparsity is induced in the `EGlearn` algorithm. Figure 6 shows the entries of the precision matrices $\hat{\Theta}^{\hat{G}_\rho}$ as a function of the tuning parameter ρ . Similar to a usual lasso, we observe that they tend to zero in a possibly non-monotone way.

The completed variogram matrices also serve to assess the fitted model and evaluate its goodness of fit. Figure 7 compares the values of the empirical extremal variogram $\hat{\Gamma}$ to the variogram estimates implied by the fitted graphical model $\hat{\Gamma}^G$ for graph G being the flight graph, the extremal minimum spanning tree and the optimal `EGlearn` graph,

respectively. The extremal correlations are $\chi = 2 - 2\Phi(\sqrt{\Gamma}/2)$ for a Hüsler–Reiss distribution with variogram matrix Γ , where Φ is the standard normal distribution function and all functions are applied componentwise. The results confirm the likelihood considerations and, in particular, convincingly show that a tree model is not flexible enough to capture all extremal dependencies in this data set. It is worthwhile to note that this finding is in contrast to the results of the application to river flow data in Section S.4, where the performance of the tree model is very similar to the `EGlearn` method, outperforming both the complete graph and the physical flow graph of the river network. This stresses the usefulness and applicability of our methodology in various situations with different amounts of sparsity.

Acknowledgments

Financial support from the F.R.S-FNRS in the form of a PDR grant (T020321F, J. Segers) and the Swiss National Science Foundation (S. Engelke) is gratefully acknowledged. The authors report there are no competing interests to declare.

Notes

¹ <https://www.bts.dot.gov/browse-statistical-products-and-data/bts-publications/airline-service-quality-performance-234-time>

References

- Asadi, P., Davison, A. C., and Engelke, S. (2015). Extremes on river networks. *Ann. Appl. Statist.*, 9(4):2023–2050.
- Asenova, S., Mazo, G., and Segers, J. (2021). Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes*, 24:461–500.
- Asenova, S. and Segers, J. (2023). Extremes of Markov random fields on block graphs: max-stable limits and structured Hüsler–Reiss distributions. *Extremes*, 26:433–468.
- Bakonyi, M. and Johnson, C. R. (1995). The euclidian distance matrix completion problem. *SIAM J. Matrix Anal. Appl.*, 16(2):646–654.
- Bakonyi, M. and Woerdeman, H. J. (2011). *Matrix Completions, Moments, and Sums of Hermitian Squares*. Princeton University Press.
- Baz, J., Díaz, I., Montes, S., and Pérez-Fernández, R. (2022). Some results on the gaussian markov random field construction problem based on the use of invariant subgraphs. *TEST*.
- Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: spatial dependencies among heavy rainfall in france. *J Climate*, 26(20):7929–7937.
- Casey, A. and Papasthatopoulos, I. (2023). Decomposable tail graphical models. Available from <https://arxiv.org/abs/2302.05182>.
- Cooley, D. and Thibaud, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3):587–604.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.*, 27:161–186.

de Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105:575–592.

de Haan, L. and Resnick, S. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 40:317–337.

Einmahl, J. H. J., Kiriliouk, A., Krajina, A., and Segers, J. (2016). An M-estimator of spatial tail dependence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 78:275–298.

Einmahl, J. H. J., Kiriliouk, A., and Segers, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, 21(2):205–233.

Einmahl, J. H. J., Krajina, A., and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.*, 40:1764–1793.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 82(4):871–932.

Engelke, S., Hitz, A. S., Gnecco, N., and Hentschel, M. (2022a). *graphicalExtremes: Statistical Methodology for Graphical Extreme Value Models*. Available from <https://github.com/sebastian-engelke/graphicalExtremes>, R package version 0.1.0.9000.

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.*, 8:241–270.

Engelke, S., Ivanovs, J., and Storkorb, K. (2022b). Graphical models for infinite measures with applications to extremes and Lévy processes. Available from <https://arxiv.org/abs/2211.15769>.

Engelke, S., Lalancette, M., and Volgushev, S. (2022c). Learning extremal graphical structures in high dimensions. Available from <https://arxiv.org/abs/2111.00840>.

- Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 77(1):239–265.
- Engelke, S. and Volgushev, S. (2020). Structure learning for extremal tree models. Available from <https://arxiv.org/abs/2012.06179>.
- Fang, H.-r. and O’Leary, D. P. (2012). Euclidean distance matrix completion problems. *Optim. Methods Softw.*, 27(4-5):695–717.
- Farris, J. S., Kluge, A. G., and Eckardt, M. J. (1970). A numerical approach to phylogenetic systematics. *Syst Zool.*, 19(2):172–189.
- Friedman, J., Hastie, T., and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.
- Gissibl, N. and Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, 24:2693–2720.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):497–546.
- Ho, Z. W. O. and Dombry, C. (2019). Simple models for multivariate regular variation and the Hüsler–Reiss Pareto distribution. *J. Multivariate Anal.*, 173:525–550.
- Hu, S., Peng, Z., and Segers, J. (2022). Modelling multivariate extreme value distributions via markov trees. Available from <https://arxiv.org/abs/2208.02627>.
- Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Prob. Letters*, 7(4):283–286.
- Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.*, 37(5):2042 – 2065.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kiriliouk, A., Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, 61:123–135.

Lalancette, M. (2023). On pairwise interaction multivariate Pareto models. Available from <https://arxiv.org/abs/2306.12074>.

Lalancette, M., Engelke, S., and Volgushev, S. (2021). Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes. *Ann. Statist.*, 49(5):2552 – 2576.

Lauritzen, S. L. (1996). *Graphical models*, volume 17 of *Oxford statistical science series*. Clarendon Press, Oxford.

Lederer, J. and Oesting, M. (2023). Extremes in high dimensions: Methods and scalable algorithms. Available from <https://arxiv.org/abs/2303.04258>.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187.

Lee, D. and Joe, H. (2018). Multivariate extreme value copulas with factor and tree dependence structures. *Extremes*, 21:147–176.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.*, 105:263–277.

Papastathopoulos, I., Strokorb, K., Tawn, J. A., and Butler, A. (2017). Extreme events of Markov chains. *Adv. in Appl. Probab.*, 49(1):134–161.

Reich, B. J., Shaby, B. A., and Cooley, D. (2014). A hierarchical model for serially-dependent extremes: A study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.*, 19:119–135.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *J. Multivariate Anal.*, 165:117–131.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12:917–930.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*, volume 104 of *Monographs on statistics and applied probability*. Chapman & Hall/CRC, Boca Raton, Fla.

Röttger, F., Coons, J. I., and Grosdos, A. (2023a). Parametric and nonparametric symmetries in graphical models for extremes. Available from <https://arxiv.org/abs/2306.00703>.

Röttger, F., Engelke, S., and Zwiernik, P. (2023b). Total positivity in multivariate extremes. *Ann. Statist.*, 51(3):962 – 1004.

Röttger, F. and Schmitz, Q. (2023). On the local metric property in multivariate extremes. Available from <https://arxiv.org/abs/2212.10350>.

Saunders, K. R., Stephenson, A. G., and Karoly, D. J. (2020). A regionalisation approach for rainfall based on extremal dependence. *Extremes*.

Segers, J. (2020). One- versus multi-component regular variation and extremes of Markov trees. *Adv. in Appl. Probab.*, 52:855–878.

Speed, T. P. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.*, 14(1):138–150.

Uhler, C. (2017). Gaussian graphical models: An algebraic and geometric perspective. Available from <https://arxiv.org/abs/1707.04345>.

Vignotto, E., Engelke, S., and Zscheischler, J. (2021). Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland. *Weather Clim. Extremes*, 32:100318.

Wadsworth, J. and Tawn, J. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika*, 101(1):1–15.

Wainwright, M. and Jordan, M. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1(1–2):1–305.

Wan, P. and Zhou, C. (2023). Graphical lasso for extremes. Available from <https://arxiv.org/abs/2307.15054>.

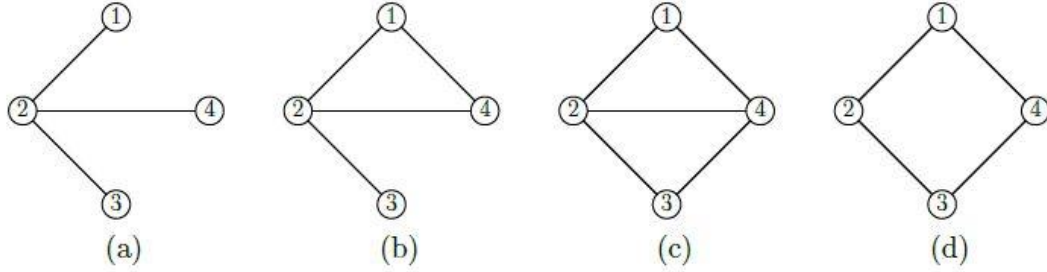


Fig. 1 Different undirected, connected graphs on four nodes: (a) is a tree, (b) is a block graph, (c) is decomposable, (d) is non-decomposable.

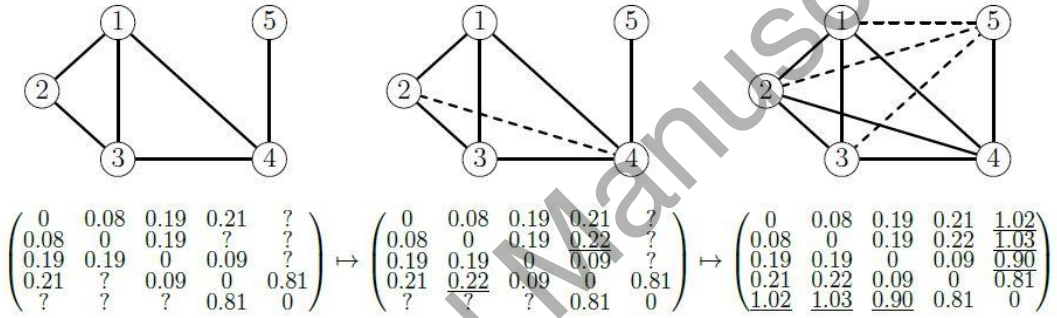


Fig. 2 Illustration of Example 4.2. On the left-hand side is the initial partial matrix Γ and the corresponding decomposable graph. In each of the following steps, dashed edges (top) correspond to newly computed matrix entries (underlined, bottom).

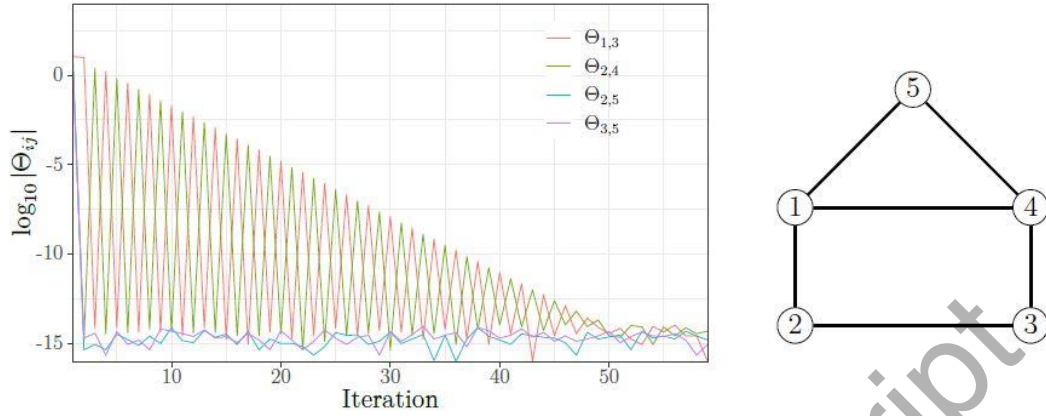


Fig. 3 Illustration of Proposition 4.2 in Example 4.11. A non-decomposable graph $G = (V, E)$ (right), and the convergence of Θ_{ij} to zero for $(i, j) \notin E$ as the graphical completion of Γ is computed (left). In each iteration, the graph defined by edge set $E_1 = E \cup \{(1, 3)\}$ or $E_2 = E \cup \{(2, 4)\}$ is completed using Proposition 4.4.

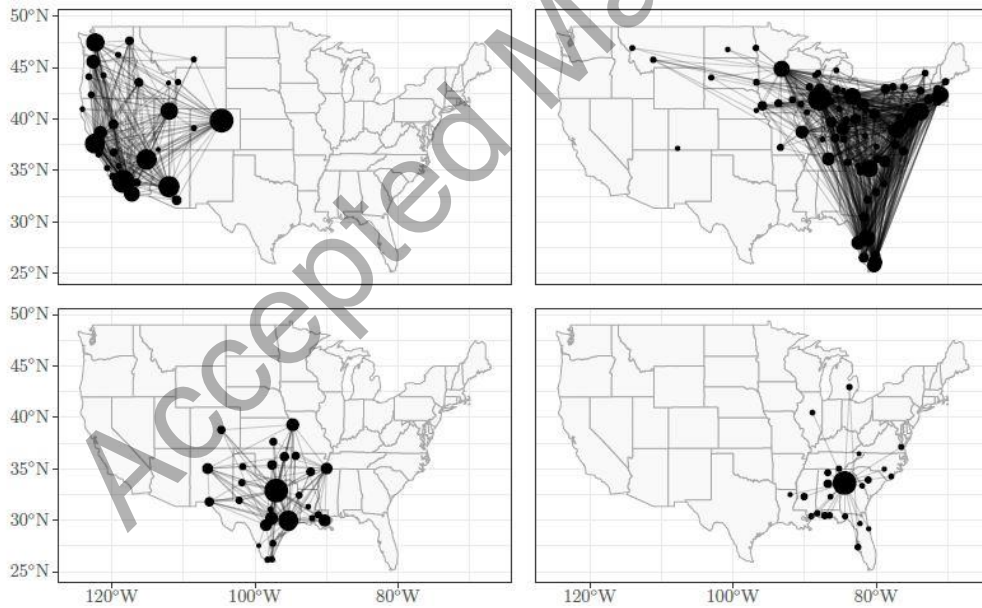


Fig. 4 Clusters of airports in the contiguous U.S., using k -medoids clustering with empirical extremal correlation as dissimilarity. Edges represent airports connected by (regular) flights. Airport sizes are proportional to the average number of daily flights.

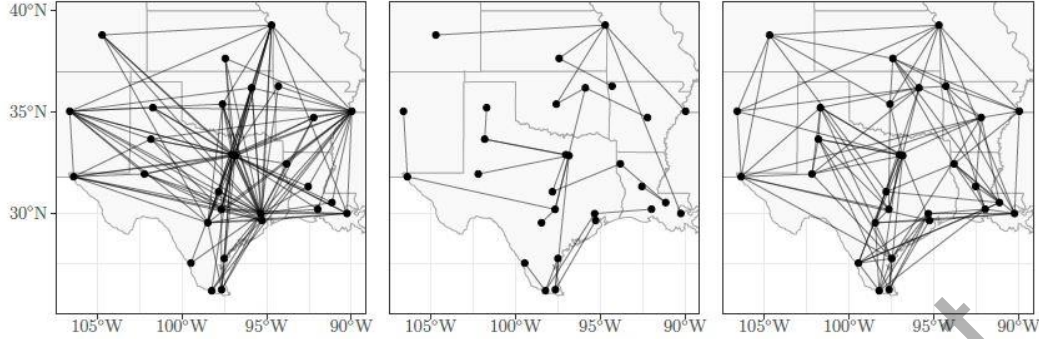


Fig. 5 The flight graph G_{flight} (left), the estimated tree graph $\hat{T}$ (center), and the graph estimated using EGlearn , $\hat{G}_{\rho^*}$, for $\rho^* = 0.15$ (right).

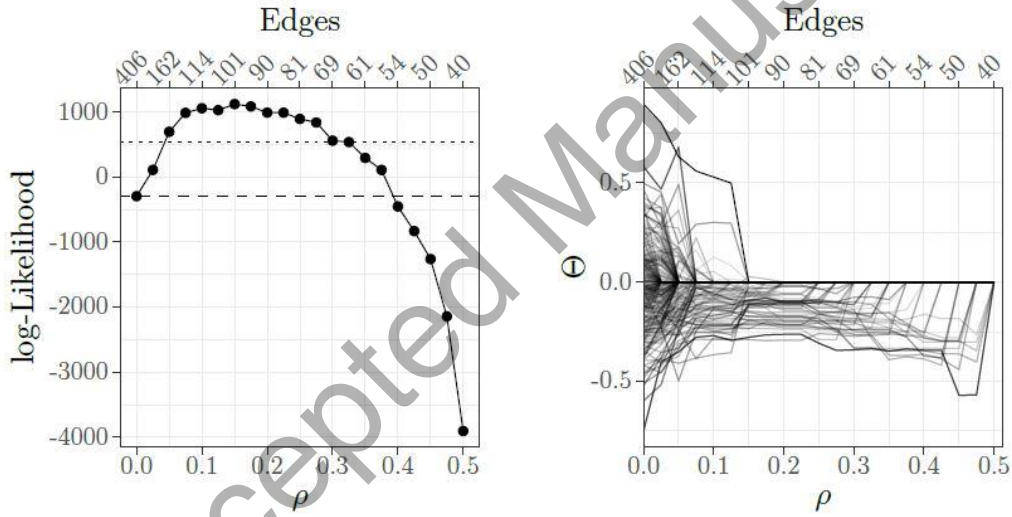


Fig. 6 Left: Log-likelihood based on the validation data set for different regularization parameters ρ . Horizontal lines indicate the log-likelihoods of the complete graph (406 edges, dashed line) and the flight graph (129 edges, dotted line). The log-likelihood of the tree graph (28 edges) is -6435 and not shown. Right: Evolution of $\hat{\theta}^{\hat{G}_{\rho}}$ entries. Only entries that vanish at $\rho = 0.5$ are shown. Both plots show the number of edges in the corresponding graph on top.

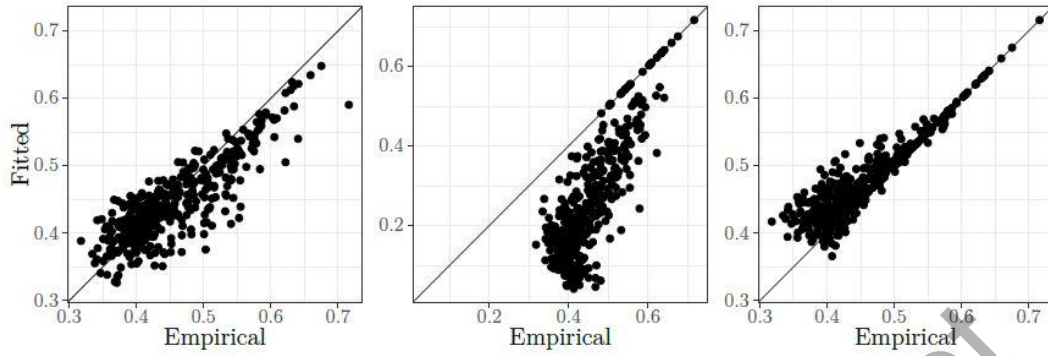


Fig. 7 Extremal correlations based on the empirical and fitted extremal variogram for the flight graph G_{flight} (left), the estimated tree graph $\hat{T}$ (center), and the graph estimated using EGlearn , $\hat{G}_{\rho^*}$, for $\rho^* = 0.15$ (right).