DIRECTIONAL VARIOGRAMS FOR MULTIVARIATE EXTREMES

Manuel Hentschel, Frank Röttger, Johan Segers, Sebastian Engelke

LIDAM Discussion Paper ISBA
2026 / 27

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

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

Directional variograms for multivariate extremes

Manuel Hentschel[1], Frank Röttger[2], Johan Segers[3], and Sebastian Engelke[1]

[1]University of Geneva
[2]University of Twente
[3]KU Leuven and UCLouvain

03.07.2026

Abstract

Multivariate generalized Pareto distributions arise as limits of threshold exceedances and form a central model class for multivariate extremes. Existing inference methods based on the extremal variogram condition on the value of a single component, which can be statistically suboptimal. We generalize this approach by conditioning the multivariate generalized Pareto random vector Y to lie on arbitrary half-spaces. Specifically, for a direction vector v, we introduce the random vector $Y^v = (Y \mid v^\top Y > 0)$ and define the associated v-variogram $\Gamma^v_{ij} = \mathrm{Var}(Y^v_i - Y^v_j)$. We establish the decomposition $Y^v \overset{d}{=} W^v + E\mathbf{1}$ into the so-called v-extremal function W^v and an independent exponential random variable E, and derive several results relating these random variables to each other. For logistic, Dirichlet, and Hüsler–Reiss multivariate generalized Pareto models, we derive closed-form expressions for Γ^v. In the Hüsler–Reiss case, we further derive new density representations and identify a distinguished resistance-curvature vector v_0 that uniquely centers the Gaussian law of W^{v_0} while characterizing the least-mass half-space. On the statistical side, we introduce empirical v-variograms and show in a simulation study that the choice of v induces a pronounced bias-variance trade-off that is strongly related to the mass of the conditioning half-space. Moreover, combining information across multiple directions v can substantially reduce estimation variance relative to methods based on a single vector.

1 Introduction

Multivariate extreme value statistics investigates rare events in large systems with the aim of describing, estimating, and predicting their dependence structure and frequency of occurrence. Such methods are essential in applications where simultaneous or cascading extremes play a critical role, including environmental risk assessment, financial stability analysis, and engineering safety. Within this framework, the approach of threshold exceedances considers a vector as extreme whenever at least one of its components exceeds a high threshold.

Multivariate generalized Pareto distributions appear as the only possible limit of threshold exceedances and are therefore natural models for extreme events (Rootzén and Tajvidi, 2006; Kiriliouk et al., 2018). The left-hand side of Figure 1 shows a sample from a multivariate Pareto distribution $Y = (Y_1, \dots, Y_d)$, where every point exceeds the threshold in at least one component. The non-standard support of the random vector Y is often problematic, and several papers consider auxiliary vectors $Y^{(m)} = (Y \mid Y_m > 0)$ on a product-form set for some $m = 1, \dots, d$ to simplify estimation (Engelke et al., 2015) or enable the definition of extremal conditional independence (Segers, 2020; Engelke and Hitz, 2020). Based on these auxiliary vectors, Engelke and Volgushev (2022) introduce a dependence measure called the extremal variogram $\Gamma^{(m)}_{ij} = \mathrm{Var}(Y^{(m)}_i - Y^{(m)}_j)$, $i, j = 1, \dots, d$. The empirical variogram $\hat{\Gamma}^{(m)}$ is the sample version of $\Gamma^{(m)}$, which however only considers the subset of observations $y \in \mathbb{R}^d$ of Y that satisfy $y_m > 0$. In order to use all data, the most common estimator therefore averages over all components m to yield $\hat{\Gamma} = \frac{1}{m}\sum_{m=1}^{d} \hat{\Gamma}^{(m)}$. The conditioning along the axis is an arbitrary choice that might be suboptimal for estimation in terms of bias and/or variance.

In this work, for any vector v in the d-variate probability simplex, we introduce the random vector $Y^v = (Y \mid v^\top Y > 0)$ and the corresponding v-variogram matrix Γ^v with entries

$$\Gamma^v_{ij} = \mathrm{Var}(Y^v_i - Y^v_j), \quad i, j = 1, \dots, d,$$

where the original extremal variograms appear as special cases using the canonical unit vectors for v. For several of the most commonly used parametric models for multivariate Pareto distributions, namely the logistic, Dirichlet (Coles and Tawn, 1991), and Hüsler–Reiss (Hüsler and Reiss, 1989) models, we obtain closed-form expressions for Γ^v. We further derive the stochastic representation

$$Y^v \overset{d}{=} W^v + E\mathbf{1},$$

where $W^v \overset{d}{=} P_v Y^v$ is obtained by applying the oblique projection $P_v = \mathrm{I} - \mathbf{1}v^\top$ along $\mathbf{1}$ onto $v^\perp$ to the random vector Y^v. The vector W^v is called the v-extremal function and is a generalization of the classical extremal function obtained when v equals a unit vector (Dombry et al., 2013). The distributions and densities (if they exist) of the random vectors Y^v and W^v for different vectors v can be related through exponential tilting.

We study the important case of the Hüsler–Reiss distribution in more detail. This model is parameterized by a variogram matrix Γ and the v-extremal function follows a $(d-1)$-dimensional Gaussian distribution on the hyperplane $v^\top$. The v-variogram is in this case independent of the choice of v and satisfies $\Gamma^v = \Gamma$. As a main theoretical contribution, we derive new density representations for Y and Y^v in this model class and provide closed forms for the normalizing constants. We further identify a special vector v_0, called the resistance curvature vector, which corresponds to the half-space $\{y \mid v_0^\top y > 0\}$ that contains the lowest expected number of samples from Y. At the same time, it is the only vector such that the Gaussian distribution of W^{v_0} is centered, and leads to

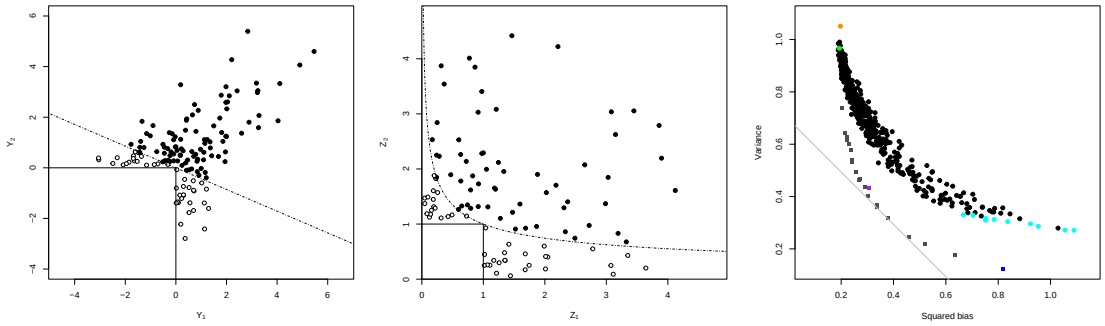


Figure 1: Left: sample from a Hüsler–Reiss multivariate Pareto distribution in $d = 2$ dimensions on exponential margins (all points), where filled points fall in the set $\{y \mid v^\top y > 0\}$ for $v = (0.3, 0.7)^\top$. Center: the same sample on Pareto margins, obtained by the transformation $Z = \exp(Y)$. Right: performance of different v-variogram estimators for data in the domain of attraction of a Hüsler–Reiss distribution. Ensemble estimators are shown as squares, and the MSE level curve in gray.

a particularly elegant density expression. The vector v_0 has also a geometric interpretation when considering Γ as a Euclidean distance matrix (Devriendt, 2022).

Estimation of the empirical version $\hat{\Gamma}^v$ of the v-variogram matrix is based on exceedances in the set $\{y \mid v^\top y > 0\}$, as shown by the filled points in the left panel of Figure 1 for the example of $v = (0.3, 0.7)^\top$ in $d = 2$ dimensions. The points in the right panel of Figure 1 represent squared biases and variances of estimates of the v-variogram matrices, averaged over all entries, for different vectors v. The cyan points correspond to the original extremal variograms estimates $\hat{\Gamma}^{(m)}$, and can be seen to have the largest biases but the smallest variances; the latter can be explained by the fact that the corresponding half-spaces have the largest expected sample size. On the other extreme, the estimator $\hat{\Gamma}^{v_0}$ of the resistance curvature vector, shown in orange, exhibits the smallest bias but the largest variance. The squares show ensemble estimators based on different sets of vectors v. These combined variograms can be seen to have the lowest mean squared error, indicated by the gray line.

We provide an extensive simulation study for inference based on our new v-variogram, with focus on models in the domain of attraction of a Hüsler–Reiss distribution. We observe that depending on the choice of the threshold, different v-variograms are optimal in terms of mean squared error. Furthermore, we find that ensembles of v-variograms permit a significant improvement of performance, and that the choice of the optimal ensemble depends on the threshold.

2 Preliminaries

2.1 Projections and half-spaces

For $v \in \mathbb{R}^d$, $v \neq \mathbf{0}$ we define the hyperplane $v^\perp = \{x \in \mathbb{R}^d \mid v^\top x = 0\}$ and the half-space $\mathcal{H}^v = \{x \in \mathbb{R}^d \mid v^\top x > 0\}$. For vectors satisfying $\mathbf{1}^\top v \neq 0$, we denote the oblique projection along the $\mathbf{1}$-direction onto $v^\perp$ as $P_v = \mathrm{I} - \mathbf{1}v^\top/(\mathbf{1}^\top v)$. Note that composed oblique projections with the same kernel simplify to the last one, i.e., $P_v P_u = P_v$ for any $\mathbf{1}^\top v, \mathbf{1}^\top u > 0$. Of particular interest will be vectors from the probability simplex $\Delta_{d-1} = \{v \in \mathbb{R}^d \mid \mathbf{1}^\top v = 1, v \geq \mathbf{0}\}$ or more generally

the affine hyperplane $\{v \mid \mathbf{1}^\top v = 1\}$. Note that for these vectors, the expression for the oblique projection matrix simplifies to $P_v = \mathrm{I} - \mathbf{1}v^\top$.

2.2 Multivariate generalized Pareto distribution

Let $X = (X_1, \ldots, X_d)^\top$ be a random vector with standard exponential margins, that is, $\mathbb{P}(X_i \leq x_i) = 1 - \exp(-x_i)$ for $i = 1, \ldots, d$ and $x_i \geq 0$. We assume standard exponential margins in order to focus on the extremal dependence structure. In practice we may need to employ empirical marginal transformations, leading to a rank-based procedure, which is done in extreme value theory at least since Drees and Huang (1998). The choice of standard exponential margins over the more common standard Pareto margins simplifies the notation for the objectives of this paper. For any $z \in \mathcal{L} = \{x \in \mathbb{R}^d \mid x \not\leq \mathbf{0}\}$, the limit of threshold exceedances

$$\mathbb{P}(Y \leq z) := \lim_{u \to \infty} \mathbb{P}(X - u\mathbf{1} \leq z \mid X \not\leq u\mathbf{1}) = \frac{\Lambda^c(z \wedge \mathbf{0}) - \Lambda^c(z)}{\Lambda^c(\mathbf{0})}, \tag{2.1}$$

if it exists, follows a multivariate generalized Pareto distribution (Rootzén and Tajvidi, 2006). We then say that the random vector X is in the domain of attraction of the multivariate generalized Pareto random vector $Y = (Y_1, \ldots, Y_d)^\top$. The distribution is characterized by an exponent measure Λ, and we write $\Lambda^c(z) := \Lambda([-\infty, \infty)^d \setminus [-\boldsymbol{\infty}, z])$. Because we started with normalized margins of X, the measure Λ is also normalized in the sense that $\Lambda(y_i > 0) = 1$ for all $i = 1, \ldots, d$. Moreover, it is homogeneous in the sense that $\Lambda(t\mathbf{1} + B) = \exp(-t)\Lambda(B)$. If the exponent measure allows a density with respect to the Lebesgue measure, we call this density $\lambda(y)$, and this density then satisfies $\lambda(y + t\mathbf{1}) = \exp(-t)\lambda(y)$ for all $y \in \mathbb{R}^d$ and all $t \in \mathbb{R}$.

A different but equivalent perspective on multivariate extremes is through point processes. We say that a random vector U on $\mathbb{R}^d$ with $\mathbb{E}(e^{U_i}) = 1$ for $i = 1, \ldots, d$, is a generator of Y, if the corresponding exponent measure Λ is the intensity of the Poisson point process (e.g., Resnick, 2008, Proposition 5.8)

$$\Pi = \sum_{j \in \mathbb{N}} \delta_{\mathbf{1}\xi_j + U_{(j)}},$$

where $(\xi_j)_{j \in \mathbb{N}}$ are the points of a Poisson point process on $\mathbb{R}$ with intensity $e^{-x}dx$ for $x \in \mathbb{R}$, and $U_{(j)}$ are independent copies of U. Moreover, we can represent the exponent measure as

$$\Lambda(A) = \int_{\mathbb{R}} e^{-x} \mathbb{P}(U + \mathbf{1}x \in A) dx, \tag{2.2}$$

for Borel sets $A \subseteq \mathbb{R}^d$. The random vector U can also serve as a U-generator of the multivariate generalized Pareto distribution in the sense of Rootzén et al. (2018, Proposition 9). Note that different generators U can result in the same exponent measure Λ. For more details see Appendix A and Corradini and Strokorb (2024, Section 2.1).

Several popular parametric models for multivariate generalized Pareto distributions exist.

Example 2.1. The extremal logistic model with parameter $\theta \in (0, 1)$ has generator $U = (U_1, \ldots, U_d)$ where U_i are independent with Gumbel distribution with scale θ and location $-\log \Gamma(1 - \theta)$ (e.g., Dombry et al., 2016).

Example 2.2. The Dirichlet model (Coles and Tawn, 1991) with parameters $\alpha_1, \ldots, \alpha_d > 0$ has generator $U = (\log V_1, \ldots, \log V_d)$ where V_i are independent Gamma$(\alpha_i, 1/\alpha_i)$ random variables (e.g., Kiriliouk et al., 2018; Corradini and Strokorb, 2024).

Example 2.3. The Hüsler–Reiss generalized Pareto distribution (Hüsler and Reiss, 1989) is parameterized by a conditionally negative definite variogram matrix Γ in the set $\mathcal{D}$ of all symmetric matrices Γ with zero diagonal and positive non-diagonal entries that satisfy $x^\top \Gamma x < 0$ for all $x \in \mathbf{1}^\perp$. Let $\Sigma = P_{\mathbf{1}}(-\frac{1}{2}\Gamma)P_{\mathbf{1}}$, $\Theta = \Sigma^+$ be its pseudoinverse, and $r_\Theta = -\frac{1}{2d}\Theta\Gamma\mathbf{1}$. The Hüsler–Reiss distribution has exponent measure density (Hentschel et al., 2025)

$$\lambda(y;\Gamma) \propto \exp(-\tfrac{1}{2}y^\top \Theta y + r_\Theta^\top y - d^{-1}\mathbf{1}^\top y), \qquad y \in \mathbb{R}^d. \tag{2.3}$$

For unit vectors e_m, $m = 1, \dots d$, similar expressions exist using $\Sigma^{e_m} := P_{e_m}(-\frac{1}{2}\Gamma)P_{e_m}^\top$ (Engelke et al., 2015). Corresponding to these density expressions, a Hüsler–Reiss random vector Y satisfies the stochastic representations

$$Y \mid \{Y_m > 0\} \overset{d}{=} W^{(m)} + \mathbf{1}E, \qquad Y \mid \{\mathbf{1}^\top Y > 0\} \overset{d}{=} W + \mathbf{1}E, \tag{2.4}$$

where $W^{(m)}$ and W are degenerate Gaussian vectors with covariance matrices Σ^{e_m} and Σ, respectively, and $E \sim \mathrm{Exp}(1)$ is an independent exponential random variable.

2.3 Variograms and Euclidean distance matrices

Let Z be a centered random vector taking values in $\mathbb{R}^d$ with finite covariance matrix $\Omega = \mathbb{E}(ZZ^\top)$. Its variogram matrix Γ is a symmetric $d \times d$-matrix given by the expected squared differences of the entries of Z, that is

$$\Gamma_{ij} = \mathbb{E}(Z_i - Z_j)^2 = \Omega_{ii} + \Omega_{jj} - 2\Omega_{ij}. \tag{2.5}$$

For any positive definite or positive semidefinite Ω satisfying $\ker(\Omega) \cap \{\mathbf{1}\}^\perp = \{\mathbf{0}\}$, the variogram matrix Γ has full rank and is conditionally negative definite. It is well known that the set of conditionally negative definite variogram matrices $\mathcal{D}$ is equivalent to the set of Euclidean distance matrices (Gower, 1982). This means that if Γ is conditionally negative definite there exists a set of d points in $\mathbb{R}^{d-1}$, which we denote as the row vectors x_i of a matrix $X \in \mathbb{R}^{d\times(d-1)}$, such that for all $1 \le i, j \le d$ we have

$$\Gamma_{ij} = \|x_i - x_j\|^2.$$

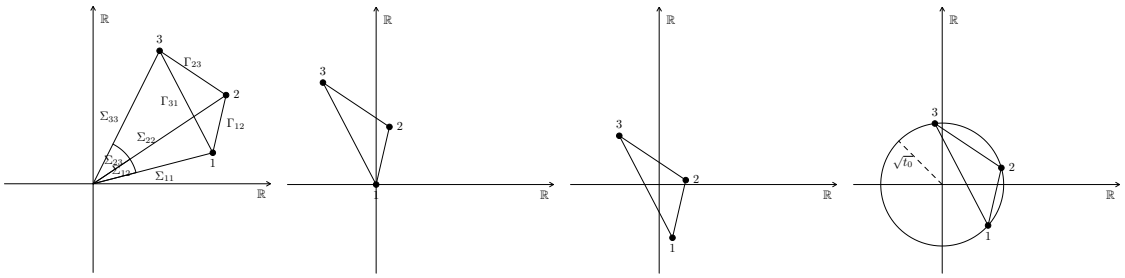
The covariance matrices Σ and Σ^{e_m}, as defined in Example 2.3 above, are equal to the inner products $XX^\top$ of these points for certain choices of X. As discussed for example in Gower (1982), the (squared) distances Γ_{ij} are invariant under translation of the points. Let $\mathbf{1}^\top v = 1$ and consider $X^v = P_v X$, which shifts the row vectors of X by the same vector $v^\top X$, resulting in a point cloud such that the linear combination of points given by v is the origin, i.e., $v^\top X = \mathbf{0}^\top \in \mathbb{R}^{d-1}$. The corresponding inner product matrix is

$$\Sigma^v = X^v (X^v)^\top.$$

In the context of a Hüsler–Reiss distribution with variogram matrix Γ, a canonical unit vector $v = \mathbf{e}_m$ yields the covariance matrix Σ^{e_m} of $W^{(m)}$ in (2.4). Similarly, choosing $v = d^{-1}\mathbf{1}$, we get the covariance matrix Σ of W in (2.4). Another interesting choice for v is

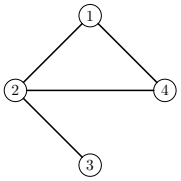
$$v_0 := 2t_0\Gamma^{-1}\mathbf{1}, \qquad \text{where} \qquad t_0 := \frac{1}{2}(\mathbf{1}^\top\Gamma^{-1}\mathbf{1})^{-1}. \tag{2.6}$$

In the literature, t_0 and v_0 are considered as resistance radius and resistance curvature of the Euclidean distance matrix Γ, as v_0 is a notion of discrete curvature, see Devriendt (2022). Clearly,



(a) $XX^\top$ (b) $v = \mathbf{e}_1$ (c) $v = d^{-1}\mathbf{1}$ (d) $v = v_0$

Figure 2: Initial matrix Σ and the effect of shifting the row vectors of X, where the points correspond to row vectors of X^v.



$$\Gamma = \begin{pmatrix} 0 & 9 & 20 & 1 \\ 9 & 0 & 11 & 6 \\ 20 & 11 & 0 & 17 \\ 1 & 6 & 17 & 0 \end{pmatrix} \qquad \Theta = \begin{pmatrix} 1.2 & 0.2 & 0 & -1.4 \\ 0.2 & 0.29 & -0.09 & -0.4 \\ 0 & -0.09 & 0.09 & 0 \\ -1.4 & -0.4 & 0 & 1.8 \end{pmatrix}$$

(a) Graph (b) Variogram (c) Precision matrix

Figure 3: Variogram and precision matrices for a Hüsler–Reiss graphical model with respect to the graph in (a).

the entries of v_0 sum up to one, because $v_0^\top \mathbf{1} = 2t_0 \mathbf{1}^\top \Gamma^{-1} \mathbf{1} = 1$. The term resistance radius comes from the fact that this value is the squared radius of the circumhypersphere that passes through all points of the simplex that spans the Euclidean distance matrix Γ.

In the context of extremal graphical models (Engelke and Hitz, 2020), it is often convenient to parameterize a Hüsler–Reiss random vector Y by the precision matrix $\Theta = (P_{\mathbf{1}}(-\frac{1}{2}\Gamma)P_{\mathbf{1}})^+$, introduced in Example 2.3, instead of the variogram matrix Γ. This matrix is a signed graph Laplacian matrix of the extremal conditional independence graph of Y (Hentschel et al., 2025), meaning that $\Theta_{ij} = 0$ if and only if $Y_i \perp_e Y_j \mid Y_{\{1,\dots,d\}\setminus\{i,j\}}$. In this context, the matrix Θ is typically referred to as Hüsler–Reiss precision matrix. We show an example for a Hüsler–Reiss graphical model in Figure 3.

3 Stochastic representations on non-negative half-spaces

Throughout this section, we consider vectors u, v from the probability simplex Δ_{d-1} as defined in Section 2.1. For such a vector v and a multivariate generalized Pareto random vector Y with standard exponential margins, we denote the restriction of Y to $\mathcal{H}^v = \{x \in \mathbb{R}^d \mid v^\top x > 0\}$ as $Y^v = Y \mid \{v^\top Y > 0\}$. In the special case $v = d^{-1}\mathbf{1}$, we write $Y^{\mathbf{1}}$ instead of $Y^{d^{-1}\mathbf{1}}$ as this is more concise and mathematically equivalent.

Remark 3.1. The requirement that $v \geq \mathbf{0}$ ensures that the half-space $\mathcal{H}^v$ is contained in $\mathcal{L} = \{x \in \mathbb{R}^d \mid x \not\leq 0\}$, the support of Y. Many of the results presented below can be extended to more general

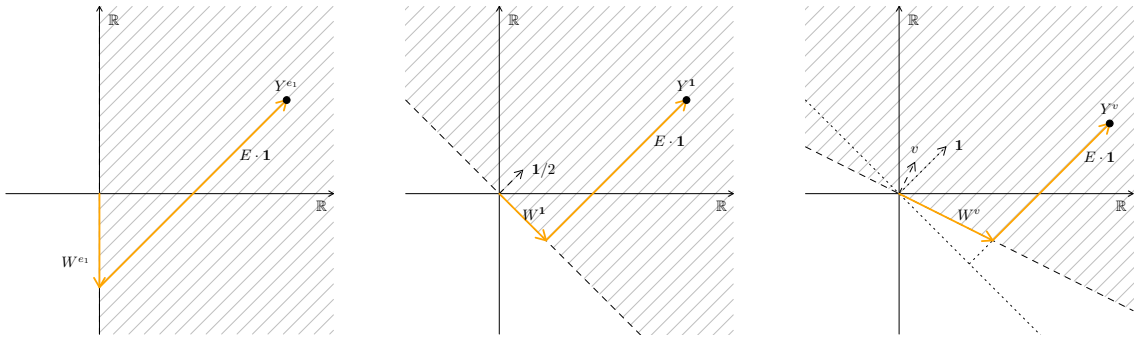


Figure 4: Geometric illustration of the stochastic representation in Proposition 3.2 for $v = e_1$ (left), $v = \mathbf{1}/2$ (center), and a general v (right), including the extremal function W^v and the corresponding multivariate generalized Pareto distributions Y^v.

v, satisfying only $\mathbf{1}^\top v = 1$, but we focus on the non-negative case to preserve a clear link between Y^v and the original multivariate generalized Pareto distribution Y.

3.1 Stochastic representation

In this section we derive stochastic representations of the random vectors Y^v in terms of a so-called v-extremal function W^v. We study the relation between versions of these random vectors for different $v \in \Delta_{d-1}$.

Proposition 3.2. *(See Proof.) Let Y be a multivariate generalized Pareto random vector, $v \in \Delta_{d-1}$, and $Y^v = Y \mid \{v^\top Y > 0\}$. Then we have the stochastic representation*

$$Y^v \overset{d}{=} W^v + E\mathbf{1}, \qquad \text{with}$$
$$W^v \overset{d}{=} P_v Y^v,$$

where E is a standard exponential random variable independent of W^v. We call W^v the v-extremal function of Y and have $\mathbb{P}(W^v \in v^\perp) = 1$.

Again, we use the more concise notation $W^{\mathbf{1}} = W^{d^{-1}\mathbf{1}}$. This decomposition is illustrated in Figure 4. The following results allow us to construct the v-extremal functions and relate them to each other. Recall the notion of a generator U of a multivariate generalized Pareto distribution as defined in Section 2.2.

Proposition 3.3. *(See Proof.) Let U be any generator of Y and let W^v be the v-extremal function defined in Proposition 3.2. We have the stochastic representation*

$$W^v \overset{d}{=} P_v U^v,$$

where U^v is the exponentially tilted random vector defined in distribution by

$$\mathbb{P}(U^v \in A) = \frac{1}{\mathbb{E}(\exp(v^\top U))} \mathbb{E}(\exp(v^\top U) \mathbb{1}\{U \in A\}),$$

for Borel sets $A \subseteq \mathbb{R}^d$. If U has density f_U, then U^v has density $f_{U^v}(y) = \frac{1}{\mathbb{E}(\exp(v^\top U))} \exp(v^\top y) f_U(y)$ for $y \in \mathbb{R}^d$.

Corollary 3.4. *(See Proof.) For any $v \in \Delta_{d-1}$ and v-extremal function W^v, the shifted random vector $W^v + \mathbf{1}c$ with $c = \log \mathbb{E}(\exp(v^\top U))$ is a generator of Y.*

Since the distribution of Y^v is defined by the restriction of the exponent measure Λ to the half-space $\mathcal{H}^v$, its density must be proportional to the exponent measure density λ, if it exists. In the proof of Proposition 3.3, the proportionality constant is computed to be $\Lambda(\mathcal{H}^v) = \mathbb{E}(\exp(v^\top U))$. Note that this normalization constant is independent of the chosen generator U of a given multivariate generalized Pareto distribution Y. We have the following representation of the density of Y^v.

Corollary 3.5. *(See Proof.) If Y has an absolutely continuous exponent measure Λ with density λ, then for $v \in \Delta_{d-1}$ the density of Y^v is*

$$f_{Y^v}(y) = \frac{1}{\mathbb{E}(\exp(v^\top U))} \mathbb{1}\{v^\top y > 0\} \lambda(y).$$

For $u \in \Delta_{d-1}$ and for y satisfying both $v^\top y > 0$ and $u^\top y > 0$, we have

$$f_{Y^v}(y) = f_{Y^u}(y) \frac{\mathbb{E}(\exp(u^\top U))}{\mathbb{E}(\exp(v^\top U))}.$$

Proposition 3.3 allows us to compute v-extremal functions if a generator U of Y is known. Furthermore, we can use the additivity of exponential tilting to obtain the following relationship between different extremal functions.

Lemma 3.6. *(See Proof.) Consider two vectors $u, v \in \Delta_{d-1}$ and the corresponding extremal functions W^u and W^v. Then, for any Borel set $A^v \subset v^\perp$, and with $A^u = P_u A^v \subset u^\perp$, we have*

$$\mathbb{P}(W^v \in A^v) = \frac{1}{\mathbb{E}(\exp(v^\top W^u))} \mathbb{E}(\exp(v^\top W^u) \mathbb{1}\{W^u \in A^u\}).$$

If W^u has density f_{W^u} on $u^\perp$, then W^v has the following density on $v^\perp$:

$$f_{W^v}(y) = \frac{\|u\|}{\|v\|} \frac{1}{\mathbb{E}(\exp(v^\top W^u))} \exp(-u^\top y) f_{W^u}(P_u y).$$

Remark 3.7. Throughout the paper, we consider densities of random vectors supported on hyperplanes $v^\perp$ with respect to the $(d-1)$-dimensional Hausdorff measure on these hyperplanes, that is, the pushforward measure obtained by the $(d-1)$-dimensional Lebesgue measure on $\mathbb{R}^{d-1}$ under any orthonormal transformation from $\mathbb{R}^{d-1}$ to $v^\perp$. See Appendix B for more details.

Since Y^v is the sum of independent random vectors W^v and $E\mathbf{1}$, which are supported on complementary subspaces, its density can be computed as the product of the densities of W^v and $E\mathbf{1}$, correcting for the Jacobian of the transformation.

Proposition 3.8. *(See Proof.) If W^v has density f_{W^v} on $v^\perp$, then the density of Y^v is given by*

$$f_{Y^v}(y) = f_{W^v}(P_v y) \|v\| \exp(-v^\top y), \quad y \in \mathcal{H}^v.$$

3.2 The v-variogram

The variogram is an important summary statistic of a random vector. We introduce the v-variogram that measures dependence in the vector Y^v.

Definition 3.9. *For a multivariate generalized Pareto random vector Y and a vector $v \in \Delta_{d-1}$, writing $Y^v = Y \mid \{v^\top Y > 0\}$, we define the v-variogram Γ^v as the $d \times d$-matrix with entries*

$$\Gamma^v_{ij} = \mathrm{Var}(Y^v_i - Y^v_j),$$

provided the variances exist and are finite.

Since we have the relations $P_v Y^v \overset{d}{=} W^v \overset{d}{=} P_v U^v$, the random vectors Y^v, W^v, and U^v differ in distribution only by an additive (random) value along the $\mathbf{1}$-direction, which cancels out in the computation of the v-variogram Γ^v, yielding the following result.

Corollary 3.10. *(See Proof.) For Y^v, W^v and U^v as above, we have*

$$\begin{aligned}\Gamma^v_{ij} &= \mathrm{Var}(Y^v_i - Y^v_j)\\ &= \mathrm{Var}(W^v_i - W^v_j)\\ &= \mathrm{Var}(U^v_i - U^v_j).\end{aligned}$$

Since many models are given through simple forms of their generators U, Corollary 3.10 together with Proposition 3.3 allow us to compute Γ^v for the three examples introduced in Section 2.2. Let $\psi^{(1)}$ denote the trigamma function.

Example 3.11. (See Proof.) For the logistic model (Example 2.1) with parameter $\theta \in (0, 1)$, we have for $i \neq j$

$$\Gamma^v_{ij} = \theta^2\psi^{(1)}(1 - v_i\theta) + \theta^2\psi^{(1)}(1 - v_j\theta).$$

Example 3.12. (See Proof.) For the Dirichlet model (Example 2.2) with parameters $\alpha_1, \ldots, \alpha_d > 0$, we have for $i \neq j$

$$\Gamma^v_{ij} = \psi^{(1)}(\alpha_i + v_i) + \psi^{(1)}(\alpha_j + v_j).$$

Note that for the m-th canonical unit vector $v = e_m$, Examples 3.11 and 3.12 recover Examples 2 and 3 from Engelke and Volgushev (2022).

Example 3.13. (See Proof.) For a covariance matrix Σ, let $U \sim \mathcal{N}(\mu, \Sigma)$, with $\mu = -\frac{1}{2}d_\Sigma$ and d_Σ denoting the vector of diagonal entries of Σ. This generator yields the Hüsler–Reiss generalized Pareto distribution in Example 2.3 parametrized by variogram $\Gamma = d_\Sigma \mathbf{1}^\top + \mathbf{1}d_\Sigma^\top - 2\Sigma$. Since both exponential tilting and projection preserve Gaussianity, the v-extremal functions can be computed to be

$$W^v \overset{d}{=} P_v U^v \sim \mathcal{N}(\mu_v, \Sigma^v), \tag{3.1}$$

with mean and covariance uniquely determined by Γ as

$$\mu_v = P_v\mu + P_v\Sigma v = -\tfrac{1}{2}P_v\Gamma v, \qquad \Sigma^v = P_v\Sigma P_v^\top = P_v(-\tfrac{1}{2}\Gamma)P_v^\top. \tag{3.2}$$

Since the variogram only depends on the covariance matrix and is invariant to shifts along the $\mathbf{1}$-direction, we find that the v-variogram Γ^v is independent of v and coincides with the parameter matrix Γ. By Corollary 3.10 we have

$$\Gamma^v_{ij} = \mathrm{Var}(U^v_i - U^v_j) = \Sigma^v_{ii} + \Sigma^v_{jj} - 2\Sigma^v_{ij} = \Gamma_{ij}.$$

4 Hüsler–Reiss density via noncanonical half-spaces

Among multivariate generalized Pareto distributions, the parametric Hüsler–Reiss family introduced in Example 2.3 has many special properties that are interesting for statistical modeling and inference. For example, Hüsler–Reiss models are parameterized by a variogram matrix Γ, which coincides with their v-variogram matrices for any $v \in \Lambda_d$. Furthermore, the Hüsler–Reiss precision matrix Θ permits parsimonious graphical modeling in extremes, which has inspired a recent series of research papers; see Engelke et al. (2024a) for an overview.

Throughout this section, let Y be a Hüsler–Reiss random vector with variogram matrix Γ. As in Example 2.3, we write $\Sigma = P_{\mathbf{1}}(-\frac{1}{2}\Gamma)P_{\mathbf{1}}$ and $\Theta = \Sigma^+$. Again consider vectors $u, v \in \Delta_{d-1}$ and let related objects such as Y^u and W^v be defined as in Section 3.

4.1 Exponent measure density expression

Using Proposition 3.8 and the density of the degenerate multivariate normal distribution in (3.1), we can directly express the density of Y^v as

$$f_{Y^v}(y) = \sqrt{(2\pi)^{-(d-1)}|\Sigma^v|_+^{-1}}\exp(-\tfrac{1}{2}\|P_v y - \mu_v\|^2_{(\Sigma^v)^+})\|v\|\exp(-v^\top y), \tag{4.1}$$

with Σ^v and μ_v as in Example 3.13, $\|y\|_A^2 := y^\top A y$, and $|\cdot|_+$ denoting the pseudo-determinant, that is, the product of the non-zero eigenvalues. See Lemma B.2 for a derivation of the density of the degenerate normal distribution on the hyperplane $v^\perp$. However, in order to express the exponent measure density λ similarly, it remains to compute the normalization constant $\Lambda(\mathcal{H}^v)$. Furthermore, it turns out that the projection $P_v y$ and subsequent multiplication with $(\Sigma^v)^+$ in the quadratic form naturally simplifies to multiplication with Θ, and the pseudo-determinant $|\Sigma^v|_+$ can be expressed in terms of $|\Theta|_+$. In the following Lemma, we provide a closed formula for $\Lambda(\mathcal{H}^v)$ for general $v \in \mathbb{R}^d$ satisfying $\mathbf{1}^\top v = 1$. For $v \in \Delta_{d-1}$, this gives the normalizing constant of the probability density in Corollary 3.5 for the Hüsler–Reiss distribution.

Lemma 4.1. *(See Proof.) Let Y follow a Hüsler–Reiss distribution with variogram matrix Γ and let $v \in \mathbb{R}^d$ with $\mathbf{1}^\top v = 1$. Then*

$$\Lambda(\mathcal{H}^v) = \mathbb{E}(\exp(v^\top U)) = \exp(-\tfrac{1}{4}v^\top \Gamma v).$$

For canonical unit vectors $v = e_m$, we recover the identity $\Lambda(\mathcal{H}^{e_m}) = 1$. For noncanonical $v \in \Delta_{d-1}$, the value of $\Lambda(\mathcal{H}^v)$ is strictly smaller than one, since Γ has strictly positive entries in off-diagonal entries. For $v = d^{-1}\mathbf{1}$, the exponent is proportional to the Kirchhoff index $\frac{1}{2}\mathbf{1}^\top\Gamma\mathbf{1} = d\,\mathrm{tr}(\Sigma)$, a scalar graph invariant, which is relevant for example in chemical graph theory (Devriendt, 2022).

As shown in Example 3.13, the extremal functions W^v have equivalent covariance matrices in the sense that they are all projections $\Sigma^v = P_v^\top(-\frac{1}{2}\Gamma)P_v$ of the same matrix. In order to relate their pseudo-determinants to each other and simplify the corresponding density expressions, we give the following result.

Lemma 4.2. *(See Proof.) For vectors u, v satisfying $\mathbf{1}^\top u = \mathbf{1}^\top v = 1$, the pseudo-determinants of Σ^u, Σ^v from Example 3.13 satisfy*

$$\frac{|\Sigma^u|_+}{|\Sigma^v|_+} = \frac{\|u\|^2}{\|v\|^2}.$$

Plugging in $u = d^{-1}\mathbf{1}$, this implies $|\Sigma^v|_+ = d\|v\|^2|\Sigma|_+$, and using $u = e_m$ it recovers $|\Sigma^{(m)}| = d|\Sigma|_+$ for the $(d-1)$-dimensional covariance matrices $\Sigma^{(m)} = \Sigma^{e_m}_{\setminus m,\setminus m}$ studied for example in Engelke and Hitz (2020). Combining these results with those from Section 3, we derive the following closed form expressions for the densities λ and f_{Y^v}.

Proposition 4.3. *(See Proof.) Let Y follow a Hüsler–Reiss distribution with variogram matrix Γ and $v \in \Delta_{d-1}$. The density of Y^v for $y \in \mathcal{H}^v$ can be expressed as*

$$f_{Y^v}(y) = \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+}\exp(-\tfrac{1}{2}\|y - \tilde{\mu}_v\|^2_\Theta)\exp(-v^\top y), \tag{4.2}$$

with $\tilde{\mu}_v := P_{\mathbf{1}}\mu_v = P_{\mathbf{1}}(-\frac{1}{2}\Gamma)v$. For the exponent measure density λ, we have

$$\lambda(y) = \exp(-\tfrac{1}{4}v^\top\Gamma v)f_{Y^v}(y) \tag{4.3}$$

for any $y \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$ with $\mathbf{1}^\top v = 1$, using the algebraic expression from (4.2) for values of y outside the support of f_{Y^v} or $v \notin \Delta_{d-1}$.

Note that the value of the right-hand side in (4.3) does not depend on $v \in \mathbb{R}^d$, as long as $\mathbf{1}^\top v = 1$, which is why v is omitted on the left-hand side. Using the general expression of f_{Y^v} in (4.1) and Proposition 4.3, we can recover some well-known density representations arising from particular choices of v.

Example 4.4. Let $v = e_m$ be a canonical unit vector and let μ_{e_m} and Σ^{e_m} be as in (3.2). Simplifying (4.1) then yields

$$f_{Y^{e_m}}(y) = \sqrt{(2\pi)^{-(d-1)}|\Sigma^{e_m}|_+^{-1}}\exp(-\tfrac{1}{2}\|y - \mu_{e_m}\|^2_{\Theta_{e_m}} - y_m).$$

Observing that $e_m^\top\Gamma e_m = 0$, (4.3) furthermore implies that the same expression gives the exponent measure density $\lambda(y)$. After transforming to standard Pareto margins, this recovers the exponent measure density expression used for example in expression (9) of Engelke and Hitz (2020). Setting $v = d^{-1}\mathbf{1}$ recovers a similar expression to (2.3) and Proposition 3.4 in Hentschel et al. (2025).

Example 4.5. Let $v = d^{-1}\mathbf{1}$ and observe that $\mu_{\mathbf{1}}$ from (3.2) and $\tilde{\mu}_{\mathbf{1}}$ from Proposition 4.3 coincide for this choice of v. Then (4.2) yields

$$f_{Y^{\mathbf{1}}}(y) = \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+}\exp(-\tfrac{1}{2}\|y - \mu_{\mathbf{1}}\|^2_\Theta - d^{-1}\mathbf{1}^\top y).$$

4.2 The resistance curvature vector

In the exponent measure density expression in Proposition 4.3, the vector $v \in \mathbb{R}^d$ can be any vector that satisfies $\mathbf{1}^\top v = 1$. Standard choices include the canonical unit vectors and the vector $d^{-1}\mathbf{1}$. In this section we will show that the resistance curvature vector v_0 introduced in (2.6) leads to a particularly elegant exponent measure density expression. Our simulation study in Section 5 gives evidence for interesting statistical properties for this choice.

Let $\Gamma \in \mathcal{D}$ be a variogram and define the matrix $M_t = t\mathbf{1}\mathbf{1}^\top - \frac{1}{2}\Gamma$ for $t \in \mathbb{R}$. By Hentschel et al. (2025, Proposition 3.2), M_t is invertible for all $t \neq t_0$, where $t_0 = \frac{1}{2}(\mathbf{1}^\top\Gamma\mathbf{1})^{-1}$ is the resistance radius

of Γ. Furthermore, M_t is connected to the Hüsler–Reiss precision matrix Θ through the relationship $\Theta = \lim_{t\to\infty} M_t^{-1}$. We observe that the resistance curvature v_0 satisfies

$$M_{t_0} v_0 = (t_0 \mathbf{1}\mathbf{1}^\top - \tfrac{1}{2}\Gamma)v_0 = t_0\mathbf{1} - t_0\mathbf{1} = \mathbf{0},$$

implying that v_0 spans the one-dimensional kernel of M_{t_0}. In the following result we find that the vector v_0, if chosen as the defining vector in Proposition 4.3, leads to a particularly simple representation of the Hüsler–Reiss exponent measure density. It has already been observed to play a special role in optimal prediction in Hüsler–Reiss models where it naturally appears in the kriging formula (Bolin et al., 2025, Proposition 5.4). Furthermore, we obtain a particularly simple stochastic representation of the Hüsler–Reiss distribution restricted to the half-space spanned by v_0.

Corollary 4.6. *(See Proof.) The exponent measure density λ of a Hüsler–Reiss distribution with variogram Γ and precision matrix Θ can be expressed as*

$$\lambda(y) = \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+} \cdot \exp(-\tfrac{1}{2}t_0) \cdot \exp(-\tfrac{1}{2}y^\top \Theta y - v_0^\top y), \quad y \in \mathbb{R}^d.$$

If $v_0 \in \Delta_{d-1}$ it holds that $W^{v_0} \sim \mathcal{N}(\mathbf{0}, \Sigma^{v_0})$ with $\Sigma^{v_0} = t_0\mathbf{1}\mathbf{1}^\top - \frac{1}{2}\Gamma$.

In particular, this means that if $v_0 \in \Delta_{d-1}$, the v_0-extremal function W^{v_0} is a centered Gaussian vector. The representation of λ allows rewriting the exponent into a quadratic form involving t_0, v_0 and Θ (Engelke et al., 2025, Example 5). Since, by construction, Σ^{v_0} has a constant diagonal, it can be standardized to a correlation matrix. This gives rise to a standardized variogram and an analog of the (Gaussian) elliptope, that is, the bounded set of all positive semi-definite correlation matrices, see Devriendt et al. (2026). Furthermore, we find that the half-space spanned by v_0 has the smallest volume with respect to the exponent measure among all vectors v that sum up to one.

Corollary 4.7. *(See Proof.) For $v \in \Delta_{d-1}$ and $m = 1, \dots, d$ we have*

$$\exp(-\tfrac{1}{2}t_0) = \Lambda(\mathcal{H}^{v_0}) \le \Lambda(\mathcal{H}^{v}) \le \Lambda(\mathcal{H}^{e_m}) = 1.$$

Furthermore, the first inequality also holds for all $v \in \mathbb{R}^d$ with $\mathbf{1}^\top v = 1$ and is strict for $v \neq v_0$. The second inequality is strict for any $v \in \Delta_{d-1}$ that is not a canonical unit vector.

This result has important consequences for statistical inference, since it states that the half-space $\mathcal{H}^{v_0}$ contains in expectation the smallest number of observations of Y. For the empirical version $\hat{\Gamma}^v$ of Γ^v, the variance of $\hat{\Gamma}^{v_0}$ is thus expected to be the largest across all $\hat{\Gamma}^v$ for $v \in \Delta_{d-1}$. On the other hand, as we will see in the simulation study in Section 5, $\hat{\Gamma}^{v_0}$ often has a much smaller bias.

Remark 4.8. In the context of statistical inference, we only consider vectors $v \in \Delta_{d-1}$, since only these define half-spaces $\mathcal{H}^v$ contained in the support $\mathcal{L} = \{x \in \mathbb{R}^d \mid x \not\le \mathbf{0}\}$ of the multivariate generalized Pareto distribution Y. In the case that v_0 has negative entries, we therefore consider an approximation $v_0^* \in \Delta_{d-1}$ which is closest to v_0 in the sense that it defines a half-space $\mathcal{H}^{v_0^*} \subseteq \mathcal{L}$ with a minimal value of the Hüsler–Reiss exponent measure $\Lambda(\mathcal{H}^{v_0^*})$, that is,

$$v_0^* = \operatorname*{arg\,min}_{v\in\Delta_{d-1}} \Lambda(\mathcal{H}^v) = \operatorname*{arg\,min}_{v\in\Delta_{d-1}} -v^\top \Gamma v. \tag{4.4}$$

The second equality follows from Lemma 4.1. Since Γ is a conditionally negative definite matrix and v is required to satisfy $\mathbf{1}^\top v = 1$, this optimization problem is convex and thus efficiently solvable using standard solvers. By Corollary 4.7, v_0^* agrees with v_0 if the latter is non-negative.

5 Inference based on noncanonical half-spaces

In this section we introduce an estimator for v-variograms and provide an extensive simulation study that illustrates its application and performance.

5.1 The empirical v-variogram

In the following definition we generalize the empirical variogram of Engelke and Volgushev (2022) to an estimator of the v-variogram as introduced in Definition 3.9. We further introduce an ensemble estimator for collections of vectors in Δ_{d-1}.

Definition 5.1. *Let $Y^{(1)}, \dots, Y^{(n)}$ be i.i.d. copies of a multivariate generalized Pareto vector Y. We define the empirical v-variogram as $\hat{\Gamma}^v \in \mathbb{R}^{d\times d}$ with entries*

$$\begin{aligned}\hat{\Gamma}^v_{ij} &= \widehat{\mathrm{Var}}(Y_i - Y_j \mid v^\top Y > 0),\\ &= \frac{1}{|\mathcal{J}^v| - 1} \sum_{k\in\mathcal{J}^v} ((Y_i^{(k)} - Y_j^{(k)}) - (\overline{Y}_i - \overline{Y}_j))^2,\end{aligned}$$

where $\mathcal{J}^v = \{j = 1, \dots, n \mid v^\top Y^{(j)} > 0\}$, and $\overline{Y}_i = |\mathcal{J}^v|^{-1} \sum_{k\in\mathcal{J}} Y_i^{(k)}$. For a finite, non-empty set of vectors $V \subset \Delta_{d-1}$, define the ensemble estimator

$$\hat{\Gamma}^V = \frac{1}{|V|} \sum_{v\in V} \hat{\Gamma}^v. \tag{5.1}$$

Since $\hat{\Gamma}^v_{ij}$ is the sample version of the true variance $\mathrm{Var}(Y_i^v - Y_j^v)$, we immediately obtain consistency and unbiasedness of these estimators as long as the variance Γ^v_{ij} exists.

In practice, we cannot directly observe data from a multivariate generalized Pareto distribution. Instead, we apply the following steps in order to obtain an approximate sample for the multivariate generalized Pareto distribution, see e.g. Röttger et al. (2023, Section 7.1). Let $\tilde{X}$ be a data-generating process with continuous marginal cumulative distribution functions $F_1, \dots, F_d$. Then, the transformed random vector X with $X_i = -\log(1 - F_i(\tilde{X}_i))$ has standard exponential margins. We assume that X satisfies the limit (2.1) for some multivariate generalized Pareto vector Y, or, equivalently, that X lies in the domain of attraction of Y.

Let $\tilde{x} \in \mathbb{R}^{n\times d}$ be a data matrix of n i.i.d. observations of $\tilde{X}$. In order to obtain a matrix $x \in \mathbb{R}^{n\times d}$ with approximate observations of X, it is common in extremal dependence modeling to employ empirical cumulative distribution functions, see e.g. Einmahl and Segers (2009). For each $i = 1, \dots, d$, let $\hat{F}_i$ be the empirical distribution function of $\tilde{x}_{i1}, \dots, \tilde{x}_{in}$, that is of the i-th column of $\tilde{x}$. We can now obtain an approximate data matrix $x \in \mathbb{R}^{n\times d}$ via transformations $x_{ji} = -\log(1 - \frac{n}{n+1}\hat{F}_i(\tilde{x}_{ji}))$ for all $j = 1, \dots, n$ and $i = 1, \dots, d$. Finally, we threshold and standardize the observations in x to construct an approximate sample for the limiting multivariate generalized Pareto vector Y. To this end, we select a high quantile of the standard exponential distribution q_p for $p \in (0, 1)$ close to one, and define

$$y_j = x_j - q_p \mathbf{1} \quad \forall j \in \mathcal{I} = \{\ell = 1, \dots, n \mid \max x_\ell > q_p\}.$$

The resulting data matrix $y \in \mathbb{R}^{m\times d}$, for some (random) $m = |\mathcal{I}| \le n$, is now considered as an approximate sample for the multivariate generalized Pareto vector Y. This permits the approximate calculation of the empirical variogram $\hat{\Gamma}^v$ as in Definition 5.1 for any $v \in \Delta_{d-1}$.

Model	Limiting Model				Domain of Attraction			
Vector(s)	e_1	v_0	$\{e_1, e_2\}$	$\{v_1, \dots, v_{11}\}$	e_1	v_0	$\{e_1, e_2\}$	$\{v_1, \dots, v_{11}\}$
Variance	0.047	0.057	0.025	0.032	0.033	0.082	0.027	0.049
Bias	0.004	-0.004	0.000	-0.003	-0.380	-0.140	-0.382	-0.241
RMSE	0.216	0.239	0.157	0.178	0.421	0.319	0.416	0.327

Table 1: Bias, variance, RMSE of different estimators. Sets of vectors denote ensemble estimators which average over the corresponding v-variograms.

5.2 Simulation setup

In the remainder of this section we present the results of a simulation study to explore how the empirical v-variogram from Definition 5.1 can be used for statistical inference and how the choice of v affects the resulting estimator $\hat{\Gamma}^v$. We focus on the Hüsler–Reiss distribution, for which all v-variograms Γ^v coincide with the parameter matrix Γ, independently of $v \in \Delta_{d-1}$.

For a given dimension d and variogram Γ, we consider two distributions: The limiting multivariate generalized Pareto distribution with exponential margins as defined in Example 2.3, and a distribution in the domain of attraction of this distribution in the sense of (2.1), from which we obtain an approximate sample as described in Section 5.1 above. There are many such distributions, and we use the max-stable Hüsler–Reiss distribution with the same variogram Γ since it is easy to simulate from. For the limiting distribution, we directly apply the definition of $\hat{\Gamma}^v$ from Definition 5.1, whereas for the distribution in the domain of attraction, we apply the definition of the empirical variogram to the approximate sample obtained after thresholding and transformation.

Throughout, we consider different choices of the dimension d, number of observations n, and threshold quantile p for the distribution in the domain of attraction. To ensure comparability of the results, the same randomly generated variogram Γ is considered in all setups with the same dimension d. In order to estimate the bias and variance of the estimators $\hat{\Gamma}^v$, we generate $M = 2000$ datasets for each setup.

The `R` package `graphicalExtremes` (Engelke et al., 2024b) is used to sample the variograms Γ, generate data from the considered distributions, and perform the thresholding and transformation procedure.

5.3 Two-dimensional case

First, we perform a simulation study for the two-dimensional case. This setup is particularly easy to visualize, since any two-dimensional variogram Γ is uniquely parametrized by its off-diagonal entry $\gamma := \Gamma_{12} = \Gamma_{21}$, and the allowed directions v can be parametrized by a single value $\eta \in [0, 1]$ as $v = (\eta, 1 - \eta) \in \Delta_2$. Due to symmetry, we have $v_0 = v_0^* = (0.5, 0.5)$ for any value of γ. We consider regularly spaced values of $\eta \in [0, 1]$ and use $\gamma = 1.5$ as the true value for the variogram parameter.

Figure 5 illustrates the performance of the corresponding estimators $\hat{\Gamma}^v$. Furthermore, we consider the ensemble estimator defined in (5.1), based on $V = \{e_1, e_2\}$, equivalent to the empirical variogram from Engelke and Volgushev (2022), and V consisting of the evenly spaced vectors v considered above. Table 1 shows the biases, variances, and RMSEs of the resulting estimators for e_1, v_0, and the two ensembles.

In the limiting model, we observe that all estimators $\hat{\Gamma}^v$ are unbiased and differ only in their variance. In fact, the different variances can be perfectly explained by the different half-space

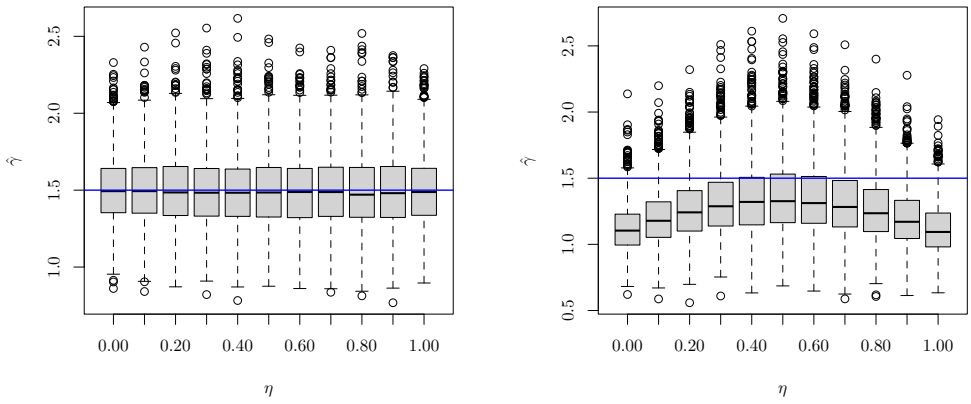


Figure 5: Performance of estimators $\hat{\Gamma}^v$ for different choices of v and $M = 2000$ dataset realizations, based on the limiting Hüsler–Reiss distribution (left) as well as thresholded data in the domain of attraction (right). The true value of $\gamma = \Gamma_{12}$ is indicated by the blue line.

measures $\Lambda(\mathcal{H}^v)$, which are proportional to the expected number of observations used to compute $\hat{\Gamma}^v$ for each v. For the ensemble estimators we observe a significant reduction in variance.

For the thresholded data in the domain of attraction, we observe a more complex picture. Vectors near the unit vectors e_m yield estimators with low variance but high bias, whereas vectors near v_0^* yield estimators with low bias and high variance. Considering ensembles of vectors reduces the variance as expected.

This illustrates a classical issue in multivariate extreme value theory. For data from a random vector X in the domain of attraction of a multivariate generalized Pareto distribution Y, components X_j that do not exceed the threshold might still be far away (in distribution) from their limit. One way to mitigate the resulting estimation bias is to perform censoring (e.g., Wadsworth and Tawn, 2014), which however becomes computationally very costly even in moderate dimensions. An intuitive explanation for the lower bias of the empirical variogram with vector v_0 is that it selects points in lower-density regions of the exponent measure, which are therefore more extreme.

5.4 Inference based on different half-spaces

The simulation setup in higher dimensions is more complex and harder to visualize than in the two-dimensional case. We consider dimension $d = 10$ and a fixed, randomly generated variogram $\Gamma \in \mathbb{R}^{d\times d}$ as true parameter. The resistance vector v_0 of this variogram Γ does not satisfy $v_0 \geq \mathbf{0}$, and therefore we use the adapted vector v_0^* in (4.4) instead.

In order to consider not just vectors from the interior of the simplex Δ_{d-1}, we enforce different sparsity levels for the vectors v, by randomly setting $0, 1, \ldots, d-2$ entries of v to zero. The remaining non-zero entries are then sampled uniformly from the corresponding face of the probability simplex. Furthermore, we consider the vectors v_0^*, $d^{-1}\mathbf{1}$, and the unit vectors e_m for $m = 1, \ldots, d$. In total, we consider the same 500 distinct vectors for each setup.

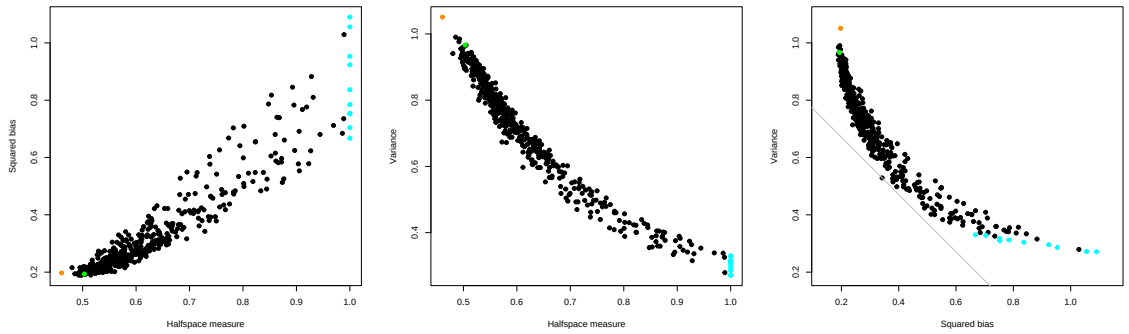


Figure 6: Estimator performance for data in the domain of attraction, using $n = 2000$ observations per dataset and a threshold of $p = 0.98$. Each point corresponds to a choice of $v \in \Delta_{d-1}$. Shown are the squared bias (left) and variance (center) against the half-space measure $\Lambda(\mathcal{H}^v)$, and variance against squared bias (right) for the different vectors v. The unit vectors e_m are shown in cyan, v_0^* in orange, and $d^{-1}\mathbf{1}$ in green.

To evaluate the performance of different estimators $\hat{\Gamma}^v$, we compute their (average) squared bias and variance as

$$\text{Bias}^2(\hat{\Gamma}^v) = \frac{1}{d(d-1)} \sum_{i \neq j} (\mathbb{E}\hat{\Gamma}^v_{ij} - \Gamma_{ij})^2,$$
$$\text{Variance}(\hat{\Gamma}^v) = \frac{1}{d(d-1)} \sum_{i \neq j} \text{Var}(\hat{\Gamma}^v_{ij}),$$

replacing expectations and variances by their sample versions based on the M dataset realizations. It turns out that the half-space measure $\Lambda(\mathcal{H}^v)$ is strongly correlated with both the variance and bias of the corresponding estimator $\hat{\Gamma}^v$, as shown in the left and center panel of Figure 6. We also plot the estimators in terms of their bias and variance in the right panel of that figure. We observe a clear bias-variance tradeoff, with the unit vectors e_m yielding low variance but high bias, and the vector v_0^* yielding low bias but high variance. Other vectors fall between these two extremes, depending on their half-space measure $\Lambda(\mathcal{H}^v)$. This behavior can be explained by implicit censoring, similarly as in the bivariate case above.

The optimal choice of vector depends strongly on the setup. Figure 7 shows the bias-variance tradeoff for different threshold quantiles p. For each dataset, the sample size n is chosen such that a fixed number of $k = 100$ observations exceeds the threshold. This is not a practical scenario, as n is usually fixed in real applications, but it allows us to isolate the effect of "more extreme" data from the effect of having less data for higher thresholds. The level curve of the best MSE is indicated by the gray line, showing that the quality of the estimation improves as the threshold increases.

Figure 8 shows the more realistic scenario of having a fixed sample size $n = 500$ and having fewer threshold exceedances for higher thresholds p. Here, we observe the same general trend, with respect to the bias-variance tradeoff for different vectors. However, due to the decreasing number of threshold exceedances for higher thresholds, the overall performance first improves and then degrades as seen by the increasing level curve of the best MSE.

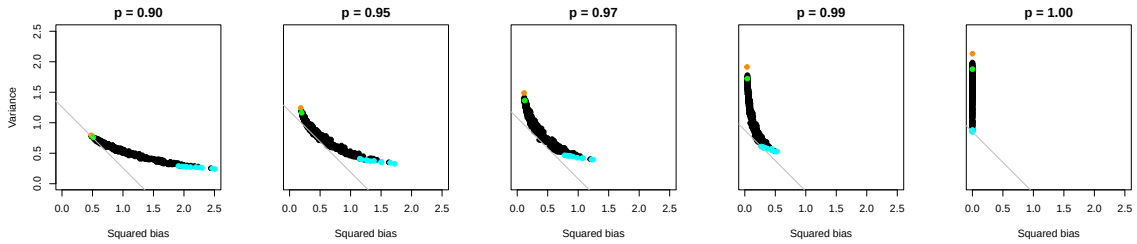


Figure 7: Bias-variance tradeoff for different threshold quantiles p, with a fixed number of $k = 100$ threshold exceedances for each dataset. The limiting multivariate generalized Pareto distribution is included as the $p = 1.0$ quantile. Highlighted vectors are colored as in Figure 6.

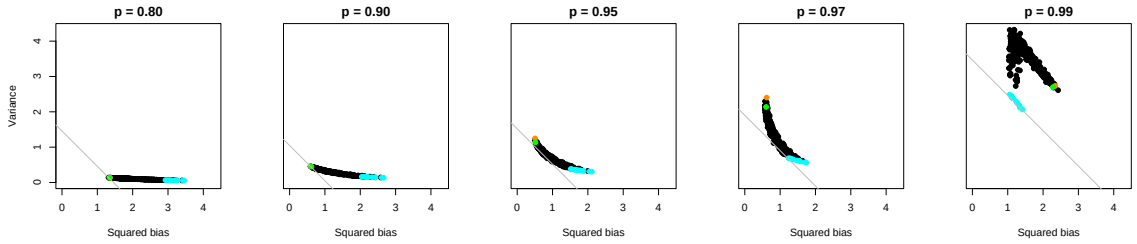


Figure 8: Bias-variance tradeoff for different threshold quantiles p, with a fixed sample size of $n = 500$ for each dataset. Highlighted vectors are colored as in Figure 6.

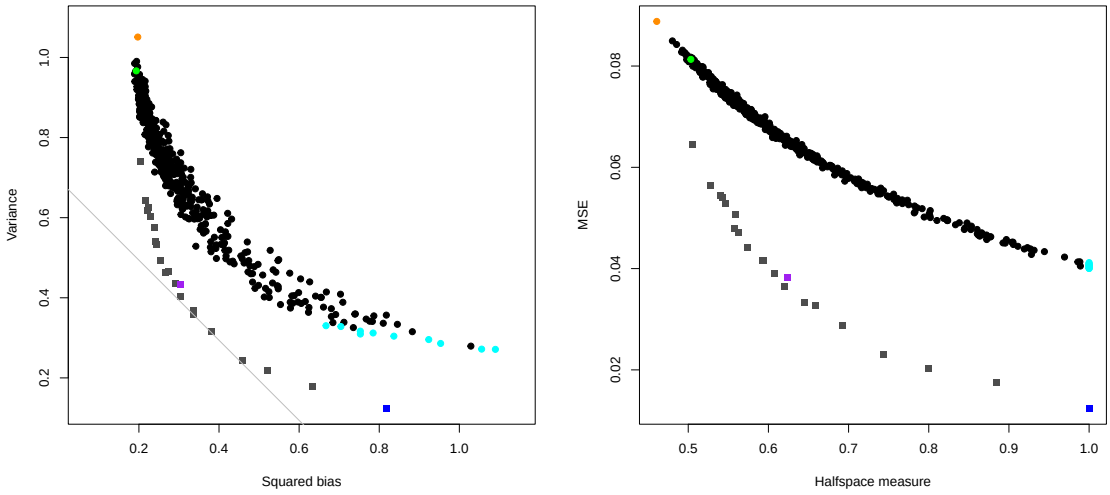


Figure 9: Performance of ensemble estimators, compared to the performance of single vector estimators. Individual vectors are highlighted as in Figure 6. Bags are shown as squares, the ensemble of all 500 vectors v is highlighted in purple, and the empirical variogram from Engelke and Volgushev (2022) in blue. Left: data in the domain of attraction with $n = 2000$ and $p = 0.98$. Right: data from the limiting multivariate generalized Pareto distribution with $n = 2000$. Note that we show MSE vs. half-space measure for the limiting distribution, as in this scenario all estimators are unbiased.

In both Figures 7 and 8, we observe that for low thresholds p, the bias dominates and vectors with smaller half-space measures perform better. As the threshold increases, the variance starts to dominate and for large thresholds, the unit vectors e_m perform best as they use the most data. In the limiting model, included as the $p = 1.0$ quantile in Figure 7, the bias is zero and the performance of the estimators only depends on the variance.

5.5 Ensemble estimators

Similar to the empirical variogram from Engelke and Volgushev (2022), we define ensemble estimators by averaging over multiple v-variograms $\hat{\Gamma}^v$. Based on the strong correlation between the half-space measure $\Lambda(\mathcal{H}^v)$ and the bias and variance of the corresponding estimator $\hat{\Gamma}^v$, we consider a grouping strategy based on the half-space measures $\Lambda(\mathcal{H}^v)$, dividing the vectors into ten groups of equal size based on their half-space measure. As this quantity is not known in practice, we also include a grouping strategy based on the sparsity of the vectors v, which happens to be a reasonable proxy for $\Lambda(\mathcal{H}^v)$. The empirical variogram from Engelke and Volgushev (2022) is included here as the ensemble of vectors with sparsity $d - 1$, each containing only a single non-zero entry. Furthermore, we consider an ensemble of all vectors from above. In total, we consider 21 different ensemble estimators: ten based on half-space measure, ten based on sparsity, and the ensemble of all vectors.

Figure 9 shows that the ensemble estimators perform significantly better than single vector

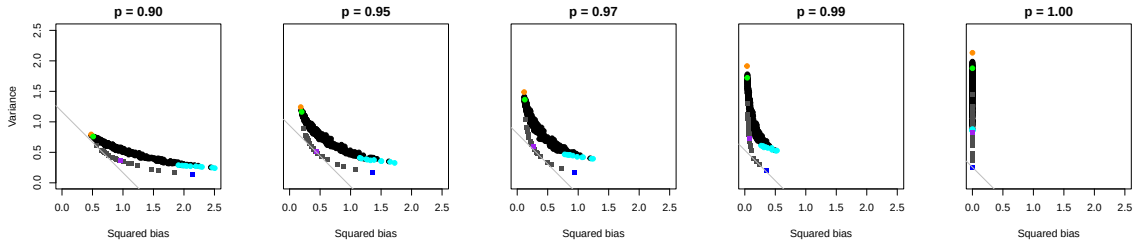


Figure 10: Bias-variance tradeoff for different threshold quantiles p and fixed number of $k = 100$ threshold exceedances for each dataset, including ensemble estimators. Color highlights and shapes are as in Figure 9.

estimators, reducing their variance and averaging the biases. This is in line with expectations, since different $\hat{\Gamma}^v$ are based on different but overlapping sets of observations, leading to positively correlated estimators. Since we grouped the vectors based on their half-space measures, the same bias-variance tradeoff from before is preserved for the ensemble estimators, with the optimal choice of vectors depending on the setup, see also Figure 10.

Acknowledgements

The authors thank Ignacio Echave-Sustaeta Rodríguez for helpful discussions and comments.

A Details on the Poisson point process construction

Consider the Poisson point processes Π^0 on $\mathbb{R} \times \mathbb{R}^d$ given by

$$\Pi^0 = \sum_{i \in \mathbb{N}} \delta_{(\xi_i, U_{(i)})}$$

with intensity $\exp(-x)dx\mathbb{P}_U$, where U is a (possibly degenerate) d-dimensional random vector U with $\mathbb{E}(e^{U_i}) = 1$, $i = 1, \dots, d$ and $U_{(i)}$ are independent copies of U. From this, we construct the point process Π on $\mathbb{R}^d$ as

$$\Pi = \sum_{i \in \mathbb{N}} \delta_{(\xi_i + U_{(i)})},$$

which has intensity $\Lambda(A) = \int_{\mathbb{R}} e^{-x} \mathbb{P}(U + \mathbf{1}x \in A)dx$. The corresponding max-stable process Z on exponential scale is given by

$$Z = \max_{i \in \mathbb{N}} \xi_i + U_{(i)}.$$

The corresponding multivariate Pareto distribution Y is defined by the restriction of Π to the set $\mathcal{L} = \{y \in \mathbb{R}^d \mid y \not\leq \mathbf{0}\}$, that is

$$\mathbb{P}(Y \in A) = \Lambda(A \cap \mathcal{L})/\Lambda(\mathcal{L}).$$

For details on this construction, see for example Resnick (2008).

B Details on densities on hyperplanes

When defining densities on hyperplanes, we consider them with respect to the following dominating measure.

Definition B.1. *Let λ_{d-1} be the $(d-1)$-dimensional Lebesgue measure on $\mathbb{R}^{d-1}$. For $v \in \mathbb{R}^d$ define μ_v as*

$$\mu_v(A) = \lambda_{d-1}(Q_v^\top A),$$

for measurable $A \subseteq v^\perp$, and $Q_v \in \mathbb{R}^{d \times (d-1)}$ with columns forming an orthonormal basis of $v^\perp$.

Note that the measure is independent of the choice of the orthonormal basis Q_v. This measure is the Hausdorff measure on the hyperplane $v^\perp$, using scaling such that it coincides with the $(d-1)$-dimensional Lebesgue measure on $\mathbb{R}^{d-1}$ when $v = \mathbf{e}_m$.

Lemma B.2. *Let $v \in \mathbb{R}^d$ and X be a multivariate normal random vector supported on $v^\perp$, with mean $\mu \perp v$ and covariance Σ, satisfying $\ker(\Sigma) = \operatorname{span}(v)$. Then its density with respect to μ_v is*

$$f(x) = (2\pi)^{-(d-1)/2} |\Sigma|_+^{-1/2} \exp(-\tfrac{1}{2}(x-\mu)^\top \Sigma^+ (x-\mu)),$$

for $x \in v^\perp$.

Proof of Lemma B.2. Let the columns of $Q_v \in \mathbb{R}^{d \times (d-1)}$ form an orthonormal basis of $v^\perp$ and write $C = Q_v^\top \Sigma Q_v \in \mathbb{R}^{(d-1)\times(d-1)}$. Let $Z \sim \mathcal{N}(\mathbf{0}, C)$ and observe that $X \overset{d}{=} \mu + Q_v Z$, since $Q_v Q_v^\top$ is the orthogonal projection matrix onto $v^\perp$. Since C is invertible, Z has density

$$f_Z(z) = (2\pi)^{-(d-1)/2} |C|^{-1/2} \exp(-\tfrac{1}{2} z^\top C^{-1} z),$$

with respect to λ_{d-1}. Hence, the density of X with respect to μ_v is given by

$$\begin{aligned} f(x) = f_Z(Q_v^\top(x-\mu)) &= (2\pi)^{-(d-1)/2}|C|^{-1/2}\exp(-\tfrac{1}{2}(x-\mu)^\top Q_v C^{-1} Q_v^\top(x-\mu)) \\ &= (2\pi)^{-(d-1)/2}|\Sigma|_+^{-1/2}\exp(-\tfrac{1}{2}(x-\mu)^\top \Sigma^+(x-\mu)). \end{aligned}$$

In the last equality we use that $|C| = |\Sigma|_+$ and $\Sigma^+ = Q_v C^{-1} Q_v^\top$, which follows from the fact that Q_v is orthonormal and basic properties of the pseudodeterminant and pseudoinverse. □

Lemma B.3. *Let $u, v \in \mathbb{R}^d$, satisfying $\mathbf{1}^\top u = \mathbf{1}^\top v = 1$, and $P_v = \mathrm{I} - \mathbf{1}v^\top$ be the oblique projection along $\mathbf{1}$ onto $v^\perp$. Let J_{uv} denote the Jacobian determinant of the mapping $P_v : u^\perp \to v^\perp$. Then*

$$J_{uv} = \frac{\|v\|}{\|u\|}.$$

Proof of Lemma B.3. We give a geometric proof, using the fact that the Jacobian determinant of a map is the factor by which it scales volumes. For a full-rank matrix $A \in \mathbb{R}^{d\times k}$, let $V(A)$ denote the k-dimensional volume (Hausdorff measure) of the parallelepiped spanned by its columns.

Next, consider $A_u \in \mathbb{R}^{d\times(d-1)}$ with columns forming a basis of $u^\perp$, and the parallelepiped spanned by the columns of A_u and $\mathbf{1}$. Its volume $V([A_u, \mathbf{1}])$ can be computed in two ways: by multiplying the volume of its base $V(A_u)$ with its height $\|\mathbf{1} - (\mathrm{I} - uu^\top/(uu^\top))\mathbf{1}\|$, and as the determinant of the matrix $[A_u, \mathbf{1}]$. Hence, we have

$$|[A_u, \mathbf{1}]| = V([A_u, \mathbf{1}]) = V(A_u)\cdot\|\mathbf{1} - (\mathrm{I} - uu^\top/(uu^\top))\mathbf{1}\| = V(A_u)\cdot\|uu^\top\mathbf{1}/(u^\top u)\| = V(A_u)/\|u\|,$$

implying $V(A_u) = \|u\|\cdot|[A_u, \mathbf{1}]|$. Similarly, for $A_v = P_v A_u$ we have $V(A_v) = \|v\|\cdot|[P_v A_u, \mathbf{1}]|$. Note that the projection $A_u \mapsto P_v A_u$ adds a multiple of $\mathbf{1}$ to each column of A_u, and hence does not change the determinant of the matrix $[A_u, \mathbf{1}]$, i.e., $|[P_v A_u, \mathbf{1}]| = |[A_u, \mathbf{1}]|$.

Putting everything together, we can compute the Jacobian determinant of the mapping $P_v : u^\perp \to v^\perp$ as

$$J_{uv} = \frac{V(A_v)}{V(A_u)} = \frac{\|v\|\cdot|[P_v A_u, \mathbf{1}]|}{\|u\|\cdot|[A_u, \mathbf{1}]|} = \frac{\|v\|}{\|u\|}.$$ □

C Proofs

C.1 Proofs from Section 3

Proof of Proposition 3.2. The proof follows along the lines of the proof of Proposition 3.3 in Wan (2026), generalizing the vector $\mathbf{1}$ to any $v \in \Delta_{d-1}$, and adjusting projections and half-spaces accordingly. Let $Q(x) = x - \mathbf{1}\max(x)$ denote the (non-linear) projection onto the set $\{x \in \mathbb{R}^d \mid \max(x) = 0\}$. Let $\tilde{W}^v \overset{d}{=} P_v Y$, supported on $v^\perp$, and $S \overset{d}{=} Q(Y) \overset{d}{=} Q(\tilde{W}^v)$. From Theorem 7 in Rootzén et al. (2018) we have the stochastic representation

$$\begin{aligned} Y &\overset{d}{=} S + \mathbf{1}E \\ &= \tilde{W}^v - \mathbf{1}\max(\tilde{W}^v) + \mathbf{1}E, \end{aligned}$$

where E follows a standard exponential distribution, and S and E are independent. Using $\mathbf{1}^\top v = 1$, we have

$$v^\top Y \geq 0 \iff E - \max(\tilde{W}^v) \geq 0$$

and

$$Y^v \stackrel{d}{=} \mathbf{1}(E - \max(\tilde{W}^v)) + \tilde{W}^v \mid \{E \geq \max(\tilde{W}^v)\}.$$

Next, consider sets of the shape $A = B + \mathbf{1} \cdot [s, \infty)$, for $B \subset v^\perp$ and $s \in \mathbb{R}$. We have

$$\begin{aligned}
\mathbb{P}(Y^v \in A) &= \mathbb{P}(Y \in A \mid v^\top Y \geq 0) \\
&= \mathbb{P}(\tilde{W}^v \in B, v^\top Y \geq s \mid v^\top Y \geq 0) \\
&= \mathbb{P}(\tilde{W}^v \in B, E - \max(\tilde{W}^v) \geq s \mid v^\top Y \geq 0) \\
&= \frac{\int_s^\infty e^{-t} \mathbb{P}(\tilde{W}^v \in B, \max(\tilde{W}^v) \leq t - s) dt}{\int_0^\infty e^{-t} \mathbb{P}(\max(\tilde{W}^v) \leq t) dt} \\
&= \frac{\int_0^\infty e^{-(u+s)} \mathbb{P}(\tilde{W}^v \in B, \max(\tilde{W}^v) \leq u) du}{\int_0^\infty e^{-t} \mathbb{P}(\max(\tilde{W}^v) \leq t) dt} \\
&= e^{-s} \cdot \frac{\int_0^\infty e^{-u} \mathbb{P}(\tilde{W}^v \in B, \max(\tilde{W}^v) \leq u) du}{\int_0^\infty e^{-t} \mathbb{P}(\max(\tilde{W}^v) \leq t) dt}.
\end{aligned}$$

For $B = v^\perp$ we get

$$\mathbb{P}(E - \max(\tilde{W}^v) \geq s \mid v^\top Y \geq 0) = e^{-s},$$

and for $s = 0$ we obtain

$$\mathbb{P}(\tilde{W}^v \in B \mid v^\top Y \geq 0) = \frac{\int_0^\infty e^{-u} \mathbb{P}(\tilde{W}^v \in B, \max(\tilde{W}^v) \leq u) du}{\int_0^\infty e^{-t} \mathbb{P}(\max(\tilde{W}^v) \leq t) dt}.$$

Hence, the distribution of Y^v decouples into the distribution of $W^v = \tilde{W}^v \mid \{v^\top Y \geq 0\}$ and an independent standard exponential random variable. □

Proof of Proposition 3.3. The distribution of Y^v can be expressed by restricting the exponent measure Λ to the half-space $\mathcal{H}^v$. By plugging in the expression for Λ from (2.2), we then get, for any Borel set $A \subset \mathbb{R}^d$,

$$\begin{aligned}
\mathbb{P}(Y^v \in A) = \frac{\Lambda(A \cap \mathcal{H}^v)}{\Lambda(\mathcal{H}^v)} &= \frac{1}{\Lambda(\mathcal{H}^v)} \int_{\mathbb{R}} e^{-x} \mathbb{P}(U + \mathbf{1}x \in A, v^\top U + x > 0) dx \\
&= \frac{1}{\Lambda(\mathcal{H}^v)} \int_{\mathbb{R}} e^{-x} \mathbb{E}(\mathbb{1}\{U + \mathbf{1}x \in A, x > -v^\top U\}) dx \\
&= \frac{1}{\Lambda(\mathcal{H}^v)} \mathbb{E}\Big(\int_{-v^\top U}^\infty e^{-x} \mathbb{1}\{U + \mathbf{1}x \in A\} dx\Big). \qquad \text{(C.1)}
\end{aligned}$$

Next, we use the exponentially tilted random vector U^v to construct a random vector $\tilde{Y}^v$ as $\tilde{Y}^v := \mathbf{1}E + U^v - \mathbf{1}v^\top U^v$, with an independent standard exponential random variable E, which has

distribution

$$\begin{aligned}
\mathbb{P}(\tilde{Y}^v \in A) &= \int_0^\infty e^{-x}\mathbb{P}(U^v + \mathbf{1}x - \mathbf{1}v^\top U^v \in A)dx \\
&= \frac{1}{\mathbb{E}(\exp(v^\top U))}\int_0^\infty e^{-x}\mathbb{E}(\exp(v^\top U)\mathbb{1}\{U + \mathbf{1}x - \mathbf{1}v^\top U \in A\})dx \\
&= \frac{1}{\mathbb{E}(\exp(v^\top U))}\mathbb{E}(\int_0^\infty e^{-(x-v^\top U)}\mathbb{1}\{U + \mathbf{1}(x - v^\top U) \in A\}dx) \\
&= \frac{1}{\mathbb{E}(\exp(v^\top U))}\mathbb{E}(\int_{-v^\top U}^\infty e^{-x}\mathbb{1}\{U + \mathbf{1}x \in A\}dx).
\end{aligned}$$

Up to the normalizing constant, this is the same as (C.1). Since both are probability distributions with the same support, their normalizing constants must be equal, i.e.,

$$\Lambda(\mathcal{H}^v) = \mathbb{E}(\exp(v^\top U)), \tag{C.2}$$

and we have the stochastic representation

$$Y^v \overset{d}{=} \mathbf{1}E + U^v - \mathbf{1}v^\top U^v,$$

which yields the result after applying the projection P_v to both sides. The density expression follows directly from the definition of exponential tilting. □

Proof of Corollary 3.4. We check the exponent measure Λ^{W^v} obtained from (2.2) when plugging in $W^v + \mathbf{1}c$ for some $c \in \mathbb{R}$ as the generator U. Using the representation $W^v \overset{d}{=} P_v U^v$ from Proposition 3.3, we obtain, for any Borel set $A \subset \mathbb{R}^d$,

$$\begin{aligned}
\Lambda^{W^v}(A) &= \int_{\mathbb{R}} e^{-x}\mathbb{P}(W^v + \mathbf{1}c + \mathbf{1}x \in A)dx \\
&= \int_{\mathbb{R}} e^{-x}\mathbb{P}(U^v + \mathbf{1}(c + x - v^\top U^v) \in A)dx \\
&= \int_{\mathbb{R}} e^{-x}\frac{1}{\mathbb{E}(\exp(v^\top U))}\mathbb{E}(\exp(v^\top U)\mathbb{1}\{U + \mathbf{1}(c + x - v^\top U) \in A\})dx \\
&= \frac{1}{\mathbb{E}(\exp(v^\top U))}\mathbb{E}(\int_{\mathbb{R}} e^{-(x-v^\top U)}\mathbb{1}\{U + \mathbf{1}(c + x - v^\top U) \in A\}dx) \\
&= \frac{1}{\mathbb{E}(\exp(v^\top U))}\mathbb{E}(\int_{\mathbb{R}} e^{-x+c}\mathbb{1}\{U + \mathbf{1}x \in A\}dx) \\
&= \frac{\exp(c)}{\mathbb{E}(\exp(v^\top U))}\Lambda(A).
\end{aligned}$$

In the third equality we use the definition of U^v and in the fifth equality we use the change of variable $x \mapsto x + v^\top U - c$. For $c = \log(\mathbb{E}(\exp(v^\top U)))$ we obtain $\Lambda^{W^v} \equiv \Lambda$. □

Proof of Corollary 3.5. The proportionality of f_{Y^v} and λ follows from the fact that Y^v is defined as the restriction of Y to the half-space $\mathcal{H}^v$, while the distribution of Y is equal to Λ restricted to $\mathcal{L}$ and normalized to a probability measure, where $\mathcal{H}^v \subset \mathcal{L}$. The value of the proportionality constant $\Lambda(\mathcal{H}^v) = \mathbb{E}(\exp(v^\top U))$ is given in (C.2). The second expression follows by rearranging the first one for λ, using different vectors u and v. □

Proof of Lemma 3.6. Consider U^u and U^v as in Proposition 3.3. Since consecutive exponential tilting is equivalent to summing the tilting vectors, we have that U^v is the exponentially tilted version of U^u with tilting vector $v-u$. Let $A^0 = A^v + \mathbf{1}\mathbb{R}$ and note that $P_u^{-1}(A^u) = P_v^{-1}(A^v) = A^0$. Then, for the extremal functions W^u and W^v, we have

$$\begin{aligned}\mathbb{P}(W^v \in A^v) &= \mathbb{P}(U^v \in A^0) \\ &= \frac{1}{\mathbb{E}(\exp((v-u)^\top U^u))}\mathbb{E}(\exp((v-u)^\top U^u)\mathbb{1}\{U^u \in A^0\}) \\ &= \frac{1}{\mathbb{E}(\exp(v^\top W^u))}\mathbb{E}(\exp(v^\top W^u)\mathbb{1}\{W^u \in A^u\}),\end{aligned}$$

using in the last equality that

$$v^\top W^u = v^\top P_u U^u = v^\top (U^u - \mathbf{1}u^\top U^u) = (v-u)^\top U^u. \tag{C.3}$$

To compute the density of W^v, consider $\tilde{W}^u = P_u W^v$, which satisfies

$$\begin{aligned}\mathbb{P}(\tilde{W}^u \in A^u) &= \mathbb{P}(W^v \in A^v) \\ &= \frac{1}{\mathbb{E}(\exp(v^\top W^u))}\mathbb{E}(\exp(v^\top W^u)\mathbb{1}\{W^u \in A^u\}).\end{aligned}$$

This is an exponentially tilted version of W^u with tilting vector v and density

$$f_{\tilde{W}^u}(x) = \frac{1}{\mathbb{E}(\exp(v^\top W^u))}\exp(v^\top x) f_{W^u}(x), \qquad x \in u^\perp.$$

The distribution of W^v is then obtained by projecting $\tilde{W}^u$ back to $v^\perp$, which is a linear transformation with inverse P_u and Jacobian $\|v\|/\|u\|$ (see Lemma B.3). Its density on $v^\perp$ is therefore given by

$$f_{W^v}(y) = \frac{\|u\|}{\|v\|} f_{\tilde{W}^u}(P_u y) = \frac{\|u\|}{\|v\|}\frac{1}{\mathbb{E}(\exp(v^\top W^u))}\exp(v^\top P_u y) f_{W^u}(P_u y),$$

which is the stated density after simplifying $v^\top P_u y = v^\top y - v^\top \mathbf{1}u^\top y = -u^\top y$. □

Proof of Proposition 3.8. The proof goes by showing that the statement is true for all $u \in \Delta_{d-1}$ if it is true for one $v \in \Delta_{d-1}$, and then verifying the statement for the specific choice $v = d^{-1}\mathbf{1}$. To apply Corollary 3.5, we consider $y \in \mathbb{R}^d$ satisfying $v^\top y > 0$ and $u^\top y > 0$. Since f_{Y^u} must satisfy homogeneity along the $\mathbf{1}$-direction (cf. Section 2.2) and any $y \in \mathcal{H}^u$ can be shifted by a multiple of $\mathbf{1}$ to satisfy $v^\top y > 1$, the density expression also holds for $y \in \mathcal{H}^u$ with $v^\top y \le 0$.

Assume that the density expression from the result holds for some $v \in \Delta_{d-1}$, and apply

Corollary 3.5 and Lemma 3.6 to obtain

$$\begin{aligned}
f_{Y^u}(y) &= \frac{\mathbb{E}(\exp(v^\top U))}{\mathbb{E}(\exp(u^\top U))} f_{Y^v}(y) \\
&= \frac{\mathbb{E}(\exp(v^\top U))}{\mathbb{E}(\exp(u^\top U))} f_{W^v}(P_v y)\|v\| \exp(-v^\top y) \\
&= \frac{\mathbb{E}(\exp(v^\top U))}{\mathbb{E}(\exp(u^\top U))} \frac{\|u\|}{\|v\|} \frac{1}{\mathbb{E}(\exp(v^\top W^u))} \exp(-u^\top P_v y) f_{W^u}(P_u y)\|v\| \exp(-v^\top y) \\
&= \frac{\mathbb{E}(\exp(v^\top U))}{\mathbb{E}(\exp(u^\top U))\mathbb{E}(\exp(v^\top W^u))} \|u\| \exp(-u^\top y + u^\top \mathbf{1} v^\top y) \exp(-v^\top y) f_{W^u}(P_u y) \\
&= \frac{\mathbb{E}(\exp(v^\top U))}{\mathbb{E}(\exp(u^\top U))\mathbb{E}(\exp(v^\top W^u))} \|u\| \exp(-u^\top y) f_{W^u}(P_u y). \qquad \text{(C.4)}
\end{aligned}$$

To simplify the initial fraction in this expression, we use (C.3) and the definition of U^u as the exponentially tilted version of U with tilting vector u, yielding

$$\begin{aligned}
\mathbb{E}(\exp(v^\top W^u)) &= \mathbb{E}(\exp((v-u)^\top U^u)) \\
&= \frac{1}{\mathbb{E}(\exp(u^\top U))} \mathbb{E}(\exp(u^\top U) \exp((v-u)^\top U)) \\
&= \frac{1}{\mathbb{E}(\exp(u^\top U))} \mathbb{E}(\exp(v^\top U)).
\end{aligned}$$

The initial fraction in (C.4) therefore simplifies to 1, leaving

$$f_{Y^u}(y) = \|u\| \exp(-u^\top y) f_{W^u}(P_u y),$$

which is the stated density.

Lastly, we verify the "base case" $v = d^{-1}\mathbf{1}$. The densities of $W^{\mathbf{1}}$ and $E\mathbf{1}$ can be multiplied directly, since the spaces they are supported on are orthogonal (see Figure 4). Note that the multiplication of E with the vector $\mathbf{1}$ has inverse $d^{-1}\mathbf{1}^\top(\mathbf{1}E) = E$, with Jacobian $d^{-1} = \|d^{-1}\mathbf{1}\|$, yielding

$$f_{Y^{\mathbf{1}}}(y) = f_{W^{\mathbf{1}}}(P_{\mathbf{1}} y)\|d^{-1}\mathbf{1}\| \exp(-d^{-1}\mathbf{1}^\top y). \qquad \square$$

Proof of Corollary 3.10. We have

$$Y^v \overset{d}{=} W^v + \mathbf{1}E \overset{d}{=} U^v + \mathbf{1}(E - v^\top U^v)$$

and hence

$$Y_i^v - Y_j^v \overset{d}{=} W_i^v - W_j^v \overset{d}{=} U_i^v - U_j^v,$$

which implies that the variances of the three differences are equal. $\square$

Proof of Example 3.11. From the proof of Dombry et al. (2016, Proposition 6), we know that the logistic distribution with parameter $\theta \in (0,1)$ has generator $U = (U_1, \ldots, U_d)$ where U_i are independent with Gumbel distribution with scale θ and location $-\log(\Gamma(1-\theta))$.

Following Proposition 3.3, the multivariate exponential tilting by $\exp(v^\top U)$, applied to their product density, is then equivalent to a univariate exponential tilting of each U_i by $\exp(v_i U_i)$, i.e.,

$$\exp(v^\top U) \prod_{i=1}^{d} f_i(x_i) = \prod_{i=1}^{d} \exp(v_i x_i) f_i(x_i). \tag{C.5}$$

The variance of an exponentially tilted random variable can be computed as the second derivative of the cumulant-generating function of the original variable, evaluated at the tilting parameter v_i. For the considered Gumbel random variables with scale θ this computation yields

$$\mathrm{Var}(U_i^v) = \theta^2 \psi^{(1)}(1 - v_i\theta).$$

By Corollary 3.10 and the independence of the components U_i^v, the variogram of Y^v is then given by

$$\begin{aligned}\Gamma_{ij}^v &= \mathrm{Var}(U_i^v - U_j^v) \\ &= \mathrm{Var}(U_i^v) + \mathrm{Var}(U_j^v) \\ &= \theta^2\psi^{(1)}(1 - v_i\theta) + \theta^2\psi^{(1)}(1 - v_j\theta).\end{aligned}$$

□

Proof of Example 3.12. From Corradini and Strokorb (2024, Theorem/Definition 2.3) we know the following generator U for the Dirichlet model with parameters $\alpha_1, \dots, \alpha_d$:

$$U = \log(\tilde{U}_1, \dots, \tilde{U}_d),$$

where $\tilde{U}_i$ are independent $\mathrm{Gamma}(\alpha_i, 1/\alpha_i)$ random variables. The ith entry of U then has univariate (exp-Gamma) density

$$f_i(x) = \frac{1}{\Gamma(\alpha_i)(1/\alpha_i)^{\alpha_i}} \exp(\alpha_i x) \cdot \exp(-\alpha_i \exp x).$$

Applying the exponential tilting to their product density is equivalent to a univariate exponential tilting of each U_i by v_i (cf. (C.5)). For the density of the tilted random vector U^v we then have

$$\mathbb{P}(U^v \in dx) \propto \prod_{i=1}^{d} \exp(v_i x_i) \cdot \exp(\alpha_i x_i) \cdot \exp(-\alpha_i \exp x_i),$$

which is again a product of exp-Gamma densities, now with parameters $\alpha_i + v_i$ and $1/\alpha_i$.

Using the fact that the variance of an exponentially tilted random variable can be computed as the second derivative of the cumulant-generating function of the original variable, evaluated at the tilting parameter v_i, and an expression for the cumulant-generating function of an exp-Gamma random variable from Halliwell (2021), we obtain

$$\mathrm{Var}(U_i^v) = \psi^{(1)}(\alpha_i + v_i).$$

Computing the variogram of Y^v using Corollary 3.10 and the independence of the components U_i^v then yields

$$\begin{aligned}\Gamma_{ij}^v &= \mathrm{Var}(U_i^v - U_j^v) \\ &= \mathrm{Var}(U_i^v) + \mathrm{Var}(U_j^v) \\ &= \psi^{(1)}(\alpha_i + v_i) + \psi^{(1)}(\alpha_j + v_j).\end{aligned}$$

□

Proof of Example 3.13. The generator of the Hüsler–Reiss distribution with parameter matrix Γ can be chosen as any multivariate normal distribution with covariance matrix Σ and mean vector $\mu = -\frac{1}{2}d_\Sigma$ whose variogram (2.5) is Γ (Engelke and Hitz, 2020, Section 4.3). It is well-known that exponential tilting of a normal random vector corresponds to shifting the mean and leaving the covariance matrix unchanged. Therefore, we have

$$U^v \sim \mathcal{N}(\mu + \Sigma v, \Sigma),$$

and subsequent multiplication with P_v yields the stated distribution

$$\begin{aligned} W^v &\sim \mathcal{N}(P_v\mu + P_v\Sigma v, P_v\Sigma P_v^\top) \\ &\sim \mathcal{N}(\mu_v, \Sigma^v). \end{aligned}$$

In order to confirm the expression for μ_v and Σ^v in terms of Γ, we use the identities $P_v\mathbf{1} = \mathbf{0}$ and $\mathbf{1}^\top v = 1$ to simplify

$$\begin{aligned} -\tfrac{1}{2}P_v\Gamma P_v^\top &= -\tfrac{1}{2}P_v(d_\Sigma\mathbf{1}^\top + \mathbf{1}d_\Sigma^\top - 2\Sigma)P_v^\top \\ &= -\tfrac{1}{2}P_v(-2\Sigma)P_v^\top \\ &= P_v\Sigma P_v^\top, \\ -\tfrac{1}{2}P_v\Gamma v &= -\tfrac{1}{2}P_v(d_\Sigma\mathbf{1}^\top + \mathbf{1}d_\Sigma^\top - 2\Sigma)v \\ &= -\tfrac{1}{2}P_v(d_\Sigma - 2\Sigma v) \\ &= P_v\mu + P_v\Sigma v. \end{aligned}$$

To confirm the stated variogram expression, observe that

$$(\Sigma^v)_{ij} = e_i^\top P_v\Sigma P_v^\top e_j = \Sigma_{ij} - (\Sigma v)_i - (\Sigma v)_j + v^\top\Sigma v,$$

and hence

$$\Gamma^v_{ij} = (\Sigma^v)_{ii} + (\Sigma^v)_{jj} - 2(\Sigma^v)_{ij} = \Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij} = \Gamma_{ij},$$

since all expressions involving v cancel out. □

C.2 Proofs from Section 4

Proof of Lemma 4.1. It is to be shown that $\mathbb{E}(\exp(v^\top U)) = \exp(-\frac{1}{4}v^\top\Gamma v)$, for a valid generator U of the Hüsler–Reiss distribution with parameter matrix Γ. From Example 3.13, we know that a normally distributed random vector U with covariance matrix Σ and mean vector $-\frac{1}{2}d_\Sigma$ is such a generator provided it has variogram matrix Γ, i.e., it satisfies

$$\Gamma = d_\Sigma\mathbf{1}^\top + \mathbf{1}d_\Sigma^\top - 2\Sigma,$$

where d_Σ is the vector of diagonal entries of Σ. For such a random vector U, we have

$$\begin{aligned} \exp(-\tfrac{1}{4}v^\top\Gamma v) &= \exp(-\tfrac{1}{4}v^\top(d_\Sigma\mathbf{1}^\top + \mathbf{1}d_\Sigma^\top - 2\Sigma)v) \\ &= \exp(-\tfrac{1}{2}v^\top d_\Sigma + \tfrac{1}{2}v^\top\Sigma v) \\ &= \mathbb{E}(\exp(v^\top U)), \end{aligned}$$

using $\mathbf{1}^\top v = 1$ and the moment generating function of the multivariate normal distribution. □

Proof of Lemma 4.2. With $\Sigma = P_{\mathbf{1}}(-\frac{1}{2}\Gamma)P_{\mathbf{1}}$, let $X \in \mathbb{R}^{d\times(d-1)}$ be such that $\Sigma = XX^\top$. Let $v \in \mathbb{R}^d$ be such that $\mathbf{1}^\top v = 1$, and denote $\delta := v - d^{-1}\mathbf{1} \perp \mathbf{1}$. Then $\Sigma^v = P_v \Sigma P_v^\top = P_v X X^\top P_v^\top$, and

$$\begin{aligned} |\Sigma^v|_+ &= |P_v X X^\top P_v^\top|_+ \\ &= |P_v^\top P_v X X^\top|_+. \end{aligned} \tag{C.6}$$

For the first two factors in (C.6), we have

$$\begin{aligned} M^v &:= P_v^\top P_v \\ &= (\mathrm{I} - v\mathbf{1}^\top)(\mathrm{I} - \mathbf{1}v^\top) \\ &= \mathrm{I} - \mathbf{1}v^\top - v\mathbf{1}^\top + v\mathbf{1}^\top\mathbf{1}v^\top \\ &= \mathrm{I} - d^{-1}\mathbf{1}\mathbf{1}^\top + d\delta\delta^\top, \end{aligned}$$

which has the following eigenvectors:

$$\begin{aligned} M^v\mathbf{1} &= \mathbf{1} - \mathbf{1} + \mathbf{0} = \mathbf{0}, \\ M^v\delta &= \delta - \mathbf{0} + d\delta\delta^\top\delta = (1 + d\delta^\top\delta)\delta, \\ M^v w &= w + \mathbf{0} + \mathbf{0} = w, \quad \text{for } w \perp \mathbf{1}, \delta. \end{aligned}$$

Hence, the eigenvalues of M^v are 0, $1+d\delta^\top\delta$, and 1 with multiplicity $d-2$, yielding pseudo-determinant

$$|M^v|_+ = 1 + d\delta^\top\delta = dv^\top v = d\|v\|^2.$$

The last two factors in (C.6) give $XX^\top = \Sigma$. Since both M^v and Σ are symmetric with identical kernel span$\{\mathbf{1}\}$, we have $|M^v\Sigma|_+ = |M^v|_+|\Sigma|_+$. This can be seen by considering a basis change with an orthogonal basis containing $\mathbf{1}$, which transforms both matrices into block diagonal matrices with a zero block of size 1, and invertible blocks of size $d-1$, to which the multiplicativity of the regular determinant applies.

Putting together the above considerations for v, w, with $\mathbf{1}^\top w = \mathbf{1}^\top v = 1$, we get

$$|\Sigma^v|_+ = d\|v\|^2 \cdot |\Sigma|_+$$

and hence

$$\frac{|\Sigma^v|_+}{|\Sigma^w|_+} = \frac{\|v\|^2}{\|w\|^2}.$$

□

Proof of Proposition 4.3. First, we establish that $B := P_v^\top(\Sigma^v)^+ P_v$ is a pseudoinverse of Σ by verifying the conditions of Lemma S.1.4 in Hentschel et al. (2025). Note that

$$\begin{aligned} &\mathbf{1}^\top B = \mathbf{1}^\top(\mathrm{I} - v\mathbf{1}^\top)(\Sigma^v)^+(\mathrm{I} - \mathbf{1}v^\top) = \mathbf{0} \\ \Rightarrow \quad &\mathrm{Im}(B) \subseteq \mathbf{1}^\perp = \mathrm{Im}(\Sigma). \end{aligned}$$

Using $\Sigma = P_{\mathbf{1}}\Sigma^v P_{\mathbf{1}}$ and a series of simplifications of projection matrices, we have

$$\begin{aligned} \Sigma B &= P_{\mathbf{1}}\Sigma^v P_{\mathbf{1}} P_v^\top(\Sigma^v)^+ P_v \\ &= P_{\mathbf{1}}\Sigma^v(\Sigma^v)^+ P_v \\ &= P_{\mathbf{1}}(\mathrm{I} - vv^\top/\|v\|^2)P_v \\ &= P_{\mathbf{1}}. \end{aligned}$$

Hence, the conditions of the Lemma are satisfied, and we have $B = \Sigma^+ = \Theta$, yielding

$$\|P_v x\|^2_{(\Sigma^v)^+} = (P_v x)^\top (\Sigma^v)^+ (P_v x) = x^\top B x = \|x\|^2_\Theta.$$

Furthermore, Lemma 4.2 yields $|\Sigma^v|_+ = d\|v\|^2 |\Sigma|_+$, and a basic property of pseudo-determinants and pseudoinverses is $|\Sigma|_+^{-1} = |\Sigma^+|_+ = |\Theta|_+$, implying

$$\sqrt{|\Sigma^v|_+^{-1}} \cdot \|v\| = \sqrt{d^{-1}\|v\|^{-2}|\Sigma|_+^{-1}} \cdot \|v\| = \sqrt{d^{-1}|\Theta|_+}.$$

Combining these two identities with the expression for the density of Y^v in (4.1), we get

$$\begin{aligned} f_{Y^v}(y) &= \sqrt{(2\pi)^{-(d-1)}|\Sigma^v|_+^{-1}} \exp(-\tfrac{1}{2}\|P_v y - \mu_v\|^2_{(\Sigma^v)^+}) \|v\| \exp(-v^\top y) \\ &= \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+} \exp(-\tfrac{1}{2}\|y - \mu_v\|_\Theta) \exp(-v^\top y), \end{aligned}$$

with μ_v as defined in Example 3.13. Note that $\Theta\mu_v = \Theta P_{\mathbf{1}}\mu_v = \Theta\tilde{\mu}_v$, which is why μ_v can be replaced by $P_{\mathbf{1}}\mu_v = \tilde{\mu}_v$ in the expression above. To express μ_v and hence $\tilde{\mu}_v$ independently of the choice of Σ, observe that

$$P_v(-\tfrac{1}{2}\Gamma)v = P_v(-\tfrac{1}{2}d_\Sigma \mathbf{1}^\top - \tfrac{1}{2}\mathbf{1}d_\Sigma^\top + \Sigma)v = P_v(-\tfrac{1}{2}d_\Sigma + \Sigma v) = \mu_v.$$

To obtain the expression for the exponent measure density λ, we use Corollary 3.5 and Lemma 4.1 to get for all $y \in \mathcal{H}^v$ that

$$\begin{aligned} \lambda(y) &= f_{Y^v}(y)\, \mathbb{E}(\exp(v^\top U)) \\ &= \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+} \exp(-\tfrac{1}{4}v^\top \Gamma v) \exp(-\tfrac{1}{2}\|y - \mu_v\|^2_\Theta) \exp(-v^\top y). \end{aligned}$$

Since λ must be homogeneous, this expression also holds for all $y \in \mathbb{R}^d$. □

Proof of Corollary 4.6. As $v_0^\top \mathbf{1} = 1$, we can plug in v_0 in the density expression in Proposition 4.3. This yields

$$\lambda(y;\Theta) = \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+} \exp(-\tfrac{1}{4}v_0^\top \Gamma v_0) \cdot \exp(-v_0^\top y) \cdot \exp(-\tfrac{1}{2}\|y - \mu_{v_0}\|^2_\Theta).$$

We have $\frac{1}{4}v_0^\top \Gamma v_0 = t_0^2 \mathbf{1}^\top \Gamma^{-1}\Gamma\Gamma^{-1}\mathbf{1} = \frac{1}{2}t_0$ and $\mu_{v_0} = P_{\mathbf{1}}(-\frac{1}{2}\Gamma)v_0 = -t_0 P\Gamma\Gamma^{-1}\mathbf{1} = \mathbf{0}$, which yields

$$\lambda(y;\Theta) = \sqrt{d^{-1}(2\pi)^{-(d-1)}|\Theta|_+} \exp(-\tfrac{1}{2}t_0) \cdot \exp(-v_0^\top y) \cdot \exp(-\tfrac{1}{2}\|y\|^2_\Theta).$$

It follows that for $v_0 \geq \mathbf{0}$ we have $Y^{v_0} = W^{v_0} + E\mathbf{1}$ with $W^{v_0} \sim \mathcal{N}(\mathbf{0}, \Sigma^{v_0})$, where

$$\begin{aligned} \Sigma^{v_0} &= P_{v_0}\Sigma P_{v_0}^\top \\ &= P_{v_0}P_{\mathbf{1}}(-\tfrac{1}{2}\Gamma)P_{\mathbf{1}}^\top P_{v_0}^\top \\ &= P_{v_0}(-\tfrac{1}{2}\Gamma)P_{v_0}^\top \\ &= -\tfrac{1}{2}\Gamma + \tfrac{1}{2}\mathbf{1}v_0^\top\Gamma + \tfrac{1}{2}\Gamma v_0\mathbf{1}^\top - \tfrac{1}{2}\mathbf{1}v_0^\top\Gamma v_0\mathbf{1}^\top \\ &= -\tfrac{1}{2}\Gamma + t_0\mathbf{1}\mathbf{1}^\top + t_0\mathbf{1}\mathbf{1}^\top - t_0\mathbf{1}\mathbf{1}^\top \\ &= -\tfrac{1}{2}\Gamma + t_0\mathbf{1}\mathbf{1}^\top. \end{aligned}$$

□

Proof of Corollary 4.7. Let t_0 and v_0 be the resistance radius and curvature with respect to a conditionally negative definite variogram matrix Γ. It follows from Devriendt (2022, Proposition 6.14) that the minimum of $\Lambda(\mathcal{H}^v) = -\frac{1}{2}v^\top \Gamma v$ over all $v \in \mathbb{R}^d$ with $v^\top \mathbf{1} = 1$ is $-t_0$, uniquely achieved at $v = v_0$. Furthermore, because Γ has zero diagonal and positive entries otherwise, the quadratic form $v^\top \Gamma v$ is nonnegative for $v \in \Delta_{d-1}$ and only vanishes when v is a canonical unit vector. This gives the second inequality. □

References

Bolin, D., Braunsteins, P., Engelke, S., and Huser, R. (2025). Intrinsic Whittle–Matérn fields and sparse spatial extremes. https://arxiv.org/abs/2512.23395.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 53(2):377–392.

Corradini, M. and Strokorb, K. (2024). Stochastic ordering in multivariate extremes. *Extremes*, 27(3):357–396.

Devriendt, K. (2022). *Graph geometry from effective resistances*. PhD thesis, University of Oxford.

Devriendt, K., Echave-Sustaeta Rodríguez, I., and Röttger, F. (2026). Extremal conditional independence for Hüsler-Reiss distributions via modular functions. https://arxiv.org/abs/2601.21931.

Dombry, C., Engelke, S., and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103:303–317.

Dombry, C., Eyi-Minko, F., and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124.

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64(1):25–46.

Einmahl, J. H. J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.*, 37(5B):2953–2989.

Engelke, S., Gnecco, N., and Röttger, F. (2025). Extremes of structural causal models. https://arxiv.org/abs/2503.06536.

Engelke, S., Hentschel, M., Lalancette, M., and Röttger, F. (2024a). Graphical models for multivariate extremes. https://arxiv.org/abs/2402.02187.

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. (2024b). *graphicalExtremes: Statistical Methodology for Graphical Extreme Value Models*. R package version 0.3.3.

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. (2022). Structure learning for extremal tree models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):2055–2087.

Gower, J. (1982). Euclidean distance geometry. *Math. Scientist*, 7:1–14.

Halliwell, L. J. (2021). The Log-Gamma Distribution and Non-Normal Error. *Variance*, 13(2):173–189.

Hentschel, M., Engelke, S., and Segers, J. (2025). Statistical inference for Hüsler—Reiss graphical models through matrix completions. *Journal of the American Statistical Association*, 120(550):909–921.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Prob. Letters*, 7(4):283–286.

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

Resnick, S. I. (2008). *Extreme Values, Regular Variation, and Point Processes*, volume 4. Springer Science & Business Media.

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.

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

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

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

Wan, P. (2026). Characterizing extremal dependence on a hyperplane. *Biometrika*, 113(2):asag015.