Linking representations for multivariate extremes via a limit set

Natalia Nolde[1] and Jennifer L. Wadsworth[2]

[1]Department of Statistics, University of British Columbia, Canada

[2]Department of Mathematics and Statistics, Lancaster University, UK

August 13, 2021

Abstract

The study of multivariate extremes is dominated by multivariate regular variation, although it is well known that this approach does not provide adequate distinction between random vectors whose components are not always simultaneously large. Various alternative dependence measures and representations have been proposed, with the most well-known being hidden regular variation and the conditional extreme value model. These varying depictions of extremal dependence arise through consideration of different parts of the multivariate domain, and particularly exploring what happens when extremes of one variable may grow at different rates to other variables. Thus far, these alternative representations have come from distinct sources and links between them are limited. In this work we elucidate many of the relevant connections through a geometrical approach. In particular, the shape of the limit set of scaled sample clouds in light-tailed margins is shown to provide a description of several different extremal dependence representations.

Key words: multivariate extreme value theory; conditional extremes; hidden regular variation; limit set; asymptotic (in)dependence.

1 Introduction

Multivariate extreme value theory is complicated by the lack of natural ordering in $\mathbb{R}^d$, and the infinite possibilities for the underlying set of dependence structures between random variables. Some of the earliest characterizations of multivariate extremes were inspired by consideration of the vector of normalized componentwise maxima. Let $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ with $X_j \sim F_j$, and consider a sample $\boldsymbol{X}_i = (X_{1i}, \ldots, X_{di})$, $i = 1, \ldots, n$, of independent copies of $\boldsymbol{X}$. For a fixed j, defining $M_{j,n} = \max_{1 \leqslant i \leqslant n}(X_{ji})$, the extremal types theorem tells us that if we can find sequences such that $(M_{j,n} - b_{j,n})/a_{j,n}$ converges to a non-degenerate distribution, then this is the generalized extreme value distribution. Moreover, the sequence $b_{j,n} \sim F_j^{-1}(1 - c/n)$, $n \to \infty$, i.e., is of the same order as the $1 - 1/n$

quantile. A natural multivariate extension is then to examine the distribution of the vector of componentwise maxima, $(\boldsymbol{M}_n - \boldsymbol{b}_n)/\boldsymbol{a}_n$. This is intrinsically tied up with the theory of multivariate regular variation, because it leads to examination of the joint behaviour of the random vector when all components are growing at the rate determined by their $1 - 1/n$ quantile. If all components were marginally standardized to focus only on the dependence, then all normalizations would be the same.

In normalizing all components by the same amount, we only consider the dependence structure in a single "direction" in $\mathbb{R}^d$. In some cases this turns out to provide a rich description of the extremal dependence: if the limiting distribution of componentwise maxima does not have independent components, then an infinite variety of dependence structures are possible, indexed by a moment-constrained measure on a $d-1$ dimensional unit sphere. However, when the limiting dependence structure is independence, or even when some pairs are independent, this representation fails to discriminate between qualitatively different underlying dependence structures. While consideration of componentwise maxima is not necessarily a common applied methodology these days, the legacy of this approach persists: statistical methods that assume multivariate regular variation, such as multivariate generalized Pareto distributions, are still very popular in practice (e.g., Engelke and Hitz, 2020). A recent theoretical treatment of multivariate regular variation is given in Kulik and Soulier (2020).

Various other representations for multivariate extremes have emerged that analyze the structure of the dependence when some variables are growing at different rates to others. These include the so-called conditional extreme value model (Heffernan and Tawn, 2004; Heffernan and Resnick, 2007), whereby the components of $\boldsymbol{X}$ are normalized according to how they grow with a single component, X_j say. Related work examines behaviour in relation to an arbitrary linear functional of $\boldsymbol{X}$ (Balkema and Embrechts, 2007). The conditional representation allows consideration of those regions where some or all variables grow at a lesser rate than X_j if this is the region where the observations tend to lie. In other words, the limit theory is suited to giving a more detailed description of a broader range of underlying dependence structures. Another representation that explicitly considers different growth rates is that of Wadsworth and Tawn (2013). They focus particularly on characterizing joint survival probabilities under certain classes of inhomogeneous normalization; this was found to reveal additional structure that is not evident when applying a common scaling. More recently, Simpson et al. (2020) have examined certain types of unequal scaling with a view to classifying the strength of dependence in any sub group of variables of $\boldsymbol{X}$.

An alternative approach to adding detail to the extremal dependence structure focuses not on different scaling orders, but rather on second order effects when applying a common scaling. This idea was introduced by Ledford and Tawn (1996), and falls under the broader umbrella of hidden regular variation (Resnick, 2002). Various manuscripts have focused on analogizing concepts from standard multivariate regular variation to the case of hidden regular variation (e.g., Ramos and Ledford, 2009), but this approach still only focuses on a restricted region of the multivariate space where all variables are large simultaneously. For this reason, although higher-dimensional analogues exist, they are often not

practically useful for dimension $d > 2$.

Another manner of examining the extremal behaviour of $\boldsymbol{X}$ is to consider normalizing the variables such that they converge onto a limit set (e.g., Davis et al., 1988; Balkema and Nolde, 2010), described by a so-called gauge function. This requires light-tailed margins, which may occur naturally or through a transformation. If the margins are standardized to a common light-tailed form, then the shape of the limit set is revealing about the extremal dependence structure of the random variables, exposing in which directions we expect to see more observations.

Although various connections have been made in the literature, many of these representations remain somewhat disjointed. For example, there is no obvious connection between the conditional extremes methodology and the representation of Ledford and Tawn (1996, 1997), and whilst Wadsworth and Tawn (2013) provided a modest connection to conditional extremes, many open questions remain. In this paper we reveal several hitherto unknown connections that can be made through the shape of the limit set and its corresponding gauge function, when it exists, and provide a step towards unifying the treatment of multivariate extremes.

We next provide further elaboration and definition of the different representations of extremal dependence. For some definitions, it is convenient to have a standardized marginal form; we focus mainly on standard Pareto or standard exponential margins with notation $\boldsymbol{X}_P$ and $\boldsymbol{X}_E$, respectively. As mentioned above, working with common margins highlights dependence features. In Section 2 we recall the formulations of various representations for multivariate extremes, and provide a thorough background to the concepts of limit sets and their gauge functions, proving a useful new result on marginalization. Section 3 details connections linking conditional extremes, the representation of Wadsworth and Tawn (2013), Ledford and Tawn (1996), and that of Simpson et al. (2020). We provide illustrative examples in Section 4 and conclude in Section 5.

2   Background and definitions

2.1   Multivariate regular variation

A measurable function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at infinity (respectively, zero) with index $\rho \in \mathbb{R}$ if, for any $x > 0$, $f(tx)/f(t) \to x^\rho$, as $t \to \infty$ (or, respectively, $t \to 0$). We write $f \in \mathrm{RV}_\rho^\infty$ or $f \in \mathrm{RV}_\rho^0$, omitting the superscript in generic cases. If $f \in \mathrm{RV}_0$, then it is called slowly varying.

The random vector $\boldsymbol{X}$ is multivariate regularly varying on the cone $\mathbb{E} = [0, \infty]^d \backslash \{\boldsymbol{0}\}$, with index $\alpha > 0$, if for any relatively compact $B \subset \mathbb{E}$,

$$t\mathbb{P}(\boldsymbol{X}/b(t) \in B) \to \nu(B), \qquad t \to \infty, \tag{2.1}$$

with $\nu(\partial B) = 0$, $b(t) \in \mathrm{RV}_{1/\alpha}^\infty$, and the limit measure ν homogeneous of order $-\alpha$; see, e.g., Resnick

(2007), Section 6.1.4. The parts of $\mathbb{E}$ where ν places mass reveal the broad scale extremal dependence structure of $\boldsymbol{X}$. Specifically, note that we have the disjoint union $\mathbb{E} = \bigcup_C \mathbb{E}_C$, where

$$\mathbb{E}_C = \{\boldsymbol{x} \in \mathbb{E} : x_j > 0, j \in C; x_i = 0, i \notin C\} =: (0, \infty]^C \times \{0\}^{D \backslash C}, \tag{2.2}$$

and the union is over all possible subsets $C \subseteq D = \{1, \ldots, d\}$, excluding the empty set. If $\nu(\mathbb{E}_C) > 0$ then the variables indexed by C can take their most extreme values simultaneously, whilst those indexed by $D \backslash C$ are non-extreme.

The definition of multivariate regular variation in equation (2.1) requires tail equivalence of the margins. In practice, it is rare to find variables that have regularly varying tails with common indices, and multivariate regular variation is a dependence assumption placed on standardized variables. Without loss of generality, therefore, we henceforth consider $\boldsymbol{X} = \boldsymbol{X}_P$ with standard Pareto(1) margins in which case $\alpha = 1$ and $b(t) = t$.

Frequently, the set B in (2.1) is taken as $[\boldsymbol{0}, \boldsymbol{x}]^c = \mathbb{E} \backslash [\boldsymbol{0}, \boldsymbol{x}]$, leading to the *exponent function*,

$$V(\boldsymbol{x}) = \nu([\boldsymbol{0}, \boldsymbol{x}]^c). \tag{2.3}$$

Suppose that derivatives of V exist almost everywhere; this is the case for popular parametric models, such as the multivariate logistic (Gumbel, 1960), Hüsler–Reiss (Hüsler and Reiss, 1989) and asymmetric logistic distributions (Tawn, 1990). Let $\partial^{|C|}/\partial \boldsymbol{x}_C = \prod_{i \in C} \partial / \partial x_i$. If the quantity $\lim_{x_j \to 0, j \notin C} \partial^{|C|} V(\boldsymbol{x}) / \partial \boldsymbol{x}_C$ is non-zero, then the group of variables indexed by C places mass on $\mathbb{E}_C$ (Coles and Tawn, 1991).

Multivariate regular variation is often phrased in terms of a radial-angular decomposition. If (2.1) holds, then for $r \geqslant 1$,

$$\mathbb{P}(\boldsymbol{X}_P / \|\boldsymbol{X}_P\| \in A, \|\boldsymbol{X}_P\| > tr) / \mathbb{P}(\|\boldsymbol{X}_P\| > t) \to H(A) r^{-1}, \qquad t \to \infty,$$

where $A \subset \mathcal{S} = \{\boldsymbol{w} \in \mathbb{R}_+^d : \|\boldsymbol{w}\| = 1\}$ and $\|\cdot\|$ is any norm. That is, the radial variable $R = \|\boldsymbol{X}_P\|$ and the angular variable $\boldsymbol{W} = \boldsymbol{X}_P / \|\boldsymbol{X}_P\|$ are independent in the limit, with $R \sim \text{Pareto}(1)$ and $\boldsymbol{W} \in \mathcal{S}$ following distribution H. The support of the so-called spectral measure H can also be partitioned in a similar manner to $\mathbb{E}$. Letting

$$\mathbb{A}_C = \{\boldsymbol{w} \in \mathcal{S} : w_j > 0, j \in C; w_i = 0, i \notin C\},$$

we have $\mathcal{S} = \bigcup_C \mathbb{A}_C$. The measure ν places mass on $\mathbb{E}_C$ if and only if H places mass on $\mathbb{A}_C$.

2.2 Hidden regular variation

Hidden regular variation arises when: (i) there is multivariate regular variation on a cone (say $\mathbb{E}$), but the mass concentrates on a subcone $\tilde{\mathbb{E}} \subset \mathbb{E}$, and (ii) there is multivariate regular variation on the subcone $\mathbb{E}' \subseteq \mathbb{E} \backslash \tilde{\mathbb{E}}$ with a scaling function of smaller order than on the full cone. Suppose that (2.1) holds, and

ν concentrates on $\tilde{\mathbb{E}}$, in the sense that $\nu(\mathbb{E}\backslash\tilde{\mathbb{E}}) = 0$. For measurable $B \subset \mathbb{E}'$, we have hidden regular variation on $\mathbb{E}'$ if

$$t\mathbb{P}(\boldsymbol{X}_P/c(t) \in B) \to \nu'(B), \qquad t \to \infty, \qquad c(t) = o(t), \quad c(t) \in \mathrm{RV}_\zeta^\infty, \quad \zeta \in (0,1], \qquad (2.4)$$

with $\nu'(\partial B) = 0$ and the limit measure ν' homogeneous of order $-1/\zeta$ (Resnick, 2007, Section 9.4.1).

The most common cone to consider is $\mathbb{E}' = (0,\infty]^d$. This leads to the *residual tail dependence coefficient*, $\eta_D \in (0,1]$ (Ledford and Tawn, 1996). That is, suppose that (2.4) holds on $(0,\infty]^d$, then the regular variation index $\zeta = \eta_D$. The residual tail dependence coefficient for the subset $C \subset D$ is found through considering cones of the form

$$\mathbb{E}'_C = \{\boldsymbol{x} \in \mathbb{E} : x_j > 0, j \in C; x_i \in [0,\infty], i \notin C\} =: (0,\infty]^C \times [0,\infty]^{D\backslash C},$$

for which $\zeta = \eta_C$.

2.3   Different scaling orders

2.3.1   Coefficients $\tau_C(\delta)$

Simpson et al. (2020) sought to examine the extremal dependence structure of a random vector through determination of the cones $\mathbb{E}_C$ for which $\nu(\mathbb{E}_C) > 0$. Direct consideration of (hidden) regular variation conditions on these cones is impeded by the fact that $\mathbb{P}(\boldsymbol{X}_P/b(t) \in B) = 0$ for all $B \subset \mathbb{E}_C$, $C \neq D$, since no components of $\boldsymbol{X}_P/b(t)$ are equal to zero for $t < \infty$. Simpson et al. (2020) circumvent this issue by assuming that if $\nu(\mathbb{E}_C) > 0$, then there exists $\delta < 1$ such that

$$\lim_{t\to\infty} t\mathbb{P}(\min_{i\in C} X_{P,i} > xt, \max_{j\in D\backslash C} X_{P,j} \leqslant yt^\delta) = \lim_{t\to\infty} t\mathbb{P}(\min_{i\in C} X_{P,i}/t > x, \max_{j\in D\backslash C} X_{P,j}/t \leqslant yt^{\delta-1}) > 0, \qquad x,y > 0.$$
$$(2.5)$$

As such, under normalization by t, components of the random vector indexed by C remain positive, whereas those indexed by $D\backslash C$ concentrate at zero. Note that if assumption (2.5) holds for some $\delta < 1$ then it also holds for all $\delta' \in [\delta, 1]$. Simpson et al. (2020) expanded assumption (2.5) to

$$\mathbb{P}(\min_{i\in C} X_{P,i} > xt, \max_{j\in D\backslash C} X_{P,j} \leqslant yt^\delta) \in \mathrm{RV}_{-1/\tau_C(\delta)}^\infty, \qquad \delta \in [0,1], \qquad (2.6)$$

where (2.6) is viewed as a function of t, and the regular variation coefficients $\tau_C(\delta) \in (0,1]$. For a fixed δ, $\tau_C(\delta) < 1$ implies either that $\nu(\mathbb{E}_C) = 0$, or that $\nu(\mathbb{E}_C) > 0$, but that δ is too small for (2.5) to hold; see Simpson et al. (2020) for further details. Considering the coefficients $\tau_C(\delta)$ over all C and $\delta \in [0,1]$ provides information about the cones on which ν concentrates.

2.3.2   Angular dependence function $\lambda(\boldsymbol{\omega})$

Wadsworth and Tawn (2013) detailed a representation for the tail of $\boldsymbol{X}_P$ where the scaling functions were of different order in each component. They focussed principally on a sequence of univariate regular

variation conditions, characterizing

$$\mathbb{P}(\boldsymbol{X}_P > t^{\boldsymbol{\omega}}) = \ell(t; \boldsymbol{\omega}) t^{-\lambda(\boldsymbol{\omega})}, \qquad \boldsymbol{\omega} \in \mathcal{S}_\Sigma = \left\{ \boldsymbol{\omega} \in [0,1]^d : \sum_{j=1}^d \omega_j = 1 \right\}, \tag{2.7}$$

where $\ell(t; \boldsymbol{\omega}) \in \mathrm{RV}_0^\infty$ for each $\boldsymbol{\omega}$ and $\lambda : \mathcal{S}_\Sigma \to [0,1]$. Equivalently, $\mathbb{P}(\boldsymbol{X}_E > \boldsymbol{\omega} v) = \ell(e^v; \boldsymbol{\omega}) e^{-\lambda(\boldsymbol{\omega}) v}$. When all components of $\boldsymbol{\omega}$ are equal to $1/d$, connection with hidden regular variation on the cone $\mathbb{E}_D$ is restored, and we have $\eta_D = d\lambda(1/d, \ldots, 1/d)$. When the subcone $\mathbb{E}_D$ of $\mathbb{E}$ is charged with mass in limit (2.1), then $\lambda(\boldsymbol{\omega}) = \max_{1 \leqslant j \leqslant d} \omega_j$. One can equally focus on sub-vectors indexed by C to define $\lambda_C(\boldsymbol{\omega})$ for $\boldsymbol{\omega}$ in a $(|C| - 1)$-dimensional simplex; we continue to have $\eta_C = |C|\lambda_C(1/|C|, \ldots, 1/|C|)$ and $\nu(\mathbb{E}_C) > 0$ implies $\lambda_C(\boldsymbol{\omega}) = \max_{1 \leqslant j \leqslant |C|} \omega_j$.

2.4 Conditional extremes

For conditional extreme value theory (Heffernan and Tawn, 2004; Heffernan and Resnick, 2007), we focus on $\boldsymbol{X}_E$. Let $\boldsymbol{X}_{E,-j}$ represent the vector $\boldsymbol{X}_E$ without the jth component. The basic assumption is that there exist functions $\boldsymbol{a}^j : \mathbb{R} \to \mathbb{R}^{d-1}$, $\boldsymbol{b}^j : \mathbb{R} \to \mathbb{R}_+^{d-1}$ and a non-degenerate distribution K^j on $\mathbb{R}^{d-1}$ with no mass at infinity, such that

$$\mathbb{P}\left(\frac{\boldsymbol{X}_{E,-j} - \boldsymbol{a}^j(X_{E,j})}{\boldsymbol{b}^j(X_{E,j})} \leqslant \boldsymbol{z}, X_{E,j} - t > x \mid X_{E,j} > t \right) \to K^j(\boldsymbol{z}) e^{-x}, \qquad t \to \infty. \tag{2.8}$$

Typically, such assumptions are made for each $j \in D$. The normalization functions satisfy some regularity conditions detailed in Heffernan and Resnick (2007), but as Heffernan and Resnick (2007) only standardize the marginal distribution of the conditioning variable (i.e, X_j), allowing different margins in other variables, these conditions do not strongly characterize the functions $\boldsymbol{a}^j$ and $\boldsymbol{b}^j$ as used in (2.8).

When joint densities exist, application of L'Hôpital's rule gives that convergence (2.8) is equivalent to

$$\mathbb{P}\left(\frac{\boldsymbol{X}_{E,-j} - \boldsymbol{a}^j(t)}{\boldsymbol{b}^j(t)} \leqslant \boldsymbol{z} \mid X_{E,j} = t \right) \to K^j(\boldsymbol{z}), \qquad t \to \infty.$$

We will further assume convergence of the full joint density

$$\frac{\partial^{d-1}}{\partial \boldsymbol{z}} \mathbb{P}\left(\frac{\boldsymbol{X}_{E,-j} - \boldsymbol{a}^j(t)}{\boldsymbol{b}^j(t)} \leqslant \boldsymbol{z} \mid X_{E,j} = t \right) \to \frac{\partial^{d-1}}{\partial \boldsymbol{z}} K^j(\boldsymbol{z}) =: k^j(\boldsymbol{z}), \qquad t \to \infty, \tag{2.9}$$

which is the practical assumption needed for undertaking likelihood-based statistical inference using this model.

Connected to this approach is work in Balkema and Embrechts (2007), who study asymptotic behaviour of a suitably normalized random vector $\boldsymbol{X}$ conditional on lying in tH, where H is a half-space not containing the origin and $t \to \infty$. The distribution of $\boldsymbol{X}$ is assumed to have a light-tailed density whose level sets are homothetic, convex and have a smooth boundary. In this setting, with H taken to be the vertical half-space $\{\boldsymbol{x} \in \mathbb{R}^d : x_d > 1\}$, the limit is the so-called Gauss-exponential distribution with density $\exp\{-\boldsymbol{u}^T\boldsymbol{u}/2 - v\}/(2\pi)^{(d-1)/2}$, $\boldsymbol{u} \in \mathbb{R}^{d-1}$, $v > 0$.
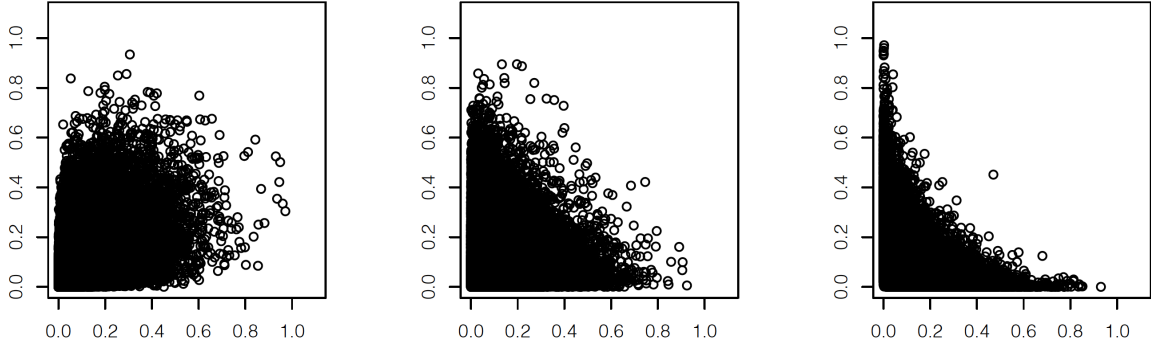
Figure 1: Sample clouds of $n = 10^5$ points simulated from meta-Gaussian distributions with standard exponential margins and copula correlation parameter $\rho = 0.5$ (left panel), $\rho = 0$ (middle panel) and $\rho = -0.5$ (right panel). The samples are scaled by the factor $r_n = \log n$. See Examples 4.1.1 and 4.1.2 for details.

2.5 Limit sets

2.5.1 Background

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent and identically distributed random vectors in $\mathbb{R}^d$. A random set $N_n = \{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$ represents a scaled n-point sample cloud. We consider situations in which there exists a scaling sequence $r_n > 0$, $r_n \to \infty$ such that scaled sample clouds N_n converges onto a deterministic set, containing at least two points. Figure 1 illustrates examples of sample clouds for which a limit set exists. Let $\mathcal{K}_d$ denote the family of non-empty compact subsets of $\mathbb{R}^d$, and $d_H(\cdot, \cdot)$ denote the Hausdorff distance between two sets (Matheron, 1975). A sequence of random sets N_n in $\mathcal{K}_d$ converges in probability onto a *limit set* $G \in \mathcal{K}_d$ if $d_H(N_n, G) \xrightarrow{\mathbb{P}} 0$ for $n \to \infty$. The following result gives convenient criteria for showing convergence in probability onto a limit set; see Balkema et al. (2010).

Proposition 1. *Random samples on $\mathbb{R}^d$ scaled by r_n converge in probability onto a deterministic set G in $\mathcal{K}_d$ if and only if*

(i) $n\mathbb{P}(\boldsymbol{X}/r_n \in U^c) \to 0$ *for any open set U containing G;*

(ii) $n\mathbb{P}(\boldsymbol{X}/r_n \in \{\boldsymbol{x} + \epsilon B\}) \to \infty$ *for all $\boldsymbol{x} \in G$ and any $\epsilon > 0$, where B is the Euclidean unit ball.*

Limit sets under various assumptions on the underlying distribution have been derived in Geffroy (1958, 1959); Fisher (1969); Davis et al. (1988); Balkema et al. (2010). Kinoshita and Resnick (1991) give a complete characterization of the possible limit sets as well as describe the class of distribution functions for which sample clouds can be scaled to converge (almost surely) onto a limit set. Furthermore, convergence in probability onto a limit set is implied by the tail large deviation principle studied in de Valk

(2016b,a).

Kinoshita and Resnick (1991) showed that if sample clouds can be scaled to converge onto a limit set almost surely, then the limit set is compact and star-shaped. A set G in $\mathbb{R}^d$ is star-shaped if $\boldsymbol{x} \in G$ implies $t\boldsymbol{x} \in G$ for all $t \in [0, 1]$. For a set $G \in \mathcal{K}_d$, if the line segment $\boldsymbol{0} + t\boldsymbol{x}$, $t \in [0, 1)$ is contained in the interior of G for every $\boldsymbol{x} \in G$, then G can be characterized by a continuous *gauge function*:

$$g(\boldsymbol{x}) = \inf\{t \geqslant 0 : \boldsymbol{x} \in tG\}, \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

A gauge function satisfies homogeneity: $g(t\boldsymbol{x}) = tg(\boldsymbol{x})$ for all $t > 0$, and the set G can be recovered from its gauge function via $G = \{\boldsymbol{x} \in \mathbb{R}^d : g(\boldsymbol{x}) \leqslant 1\}$. Examples of a gauge function include a norm $\|\cdot\|$ on $\mathbb{R}^d$, in which case $G = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \leqslant 1\}$ is the unit ball in that norm.

The shape of the limit set conveys information about extremal dependence properties of the underlying distribution. In particular, Balkema and Nolde (2010) make a connection between the shape of the limit set and asymptotic independence, whilst Nolde (2014) links its shape to the coefficient of residual tail dependence. We emphasize that the shape of the limit set depends on the choice of marginal distributions, as well as dependence structure. For example, if the components of (X_1, X_2) are independent with common marginal distribution, then $G = \{(x, y) \in \mathbb{R}_+^2 : x + y \leqslant 1\}$ if the margins are exponential; $G = \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leqslant 1\}$ if the margins are Laplace; and $G = \{(x, y) \in \mathbb{R}_+^2 : (x^\beta + y^\beta)^{1/\beta} \leqslant 1\}$ if the margins are Weibull with shape $\beta > 0$. In contrast, if the margins are exponential but G takes the latter form, this implies some dependence between the components.

2.5.2 Conditions for convergence onto a limit set

Proposition 1 provides necessary and sufficient conditions for convergence onto the limit set G, but these conditions are not particularly helpful for determining the form of G in practice.

In the following proposition, we state a criterion in terms of the joint probability density for convergence of suitably scaled random samples onto a limit set. This result is an adaptation of Proposition 3.7 in Balkema and Nolde (2010). The marginal tails of the underlying distribution are assumed to be asymptotically equal to a von Mises function. A function of the form $e^{-\psi}$ is said to be a *von Mises function* if ψ is a C^2 function with a positive derivative such that $(1/\psi'(x))' \to 0$ for $x \to \infty$. This condition on the margins says that they are light-tailed and lie in the maximum domain of attraction of the Gumbel distribution, i.e., for a random sample from such a univariate distribution, coordinate-wise maxima can be normalized to converge weakly to the Gumbel distribution (Resnick, 1987, Proposition 1.1).

Proposition 2. *Let the random vector $\boldsymbol{X}$ on $[0, \infty)^d$ have marginal distribution functions asymptotically equal to a von Mises function: $1 - F_j(x) \sim e^{-\psi_j(x)}$ for $\psi_j(x) \sim \psi(x)$, $x \to \infty$ $(j = 1, \ldots, d)$ and a joint probability density f satisfying:*

$$\frac{-\log f(t\boldsymbol{x}_t)}{\psi(t)} \to g_*(\boldsymbol{x}), \qquad t \to \infty, \qquad \boldsymbol{x}_t \to \boldsymbol{x}, \qquad \boldsymbol{x} \in [0, \infty)^d \tag{2.10}$$

for a continuous function g_* on $[0, \infty)^d$, which is positive outside a bounded set. Then a sequence of scaled random samples $N_n = \{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$ from f converges in probability onto a limit set G with $G = \{\boldsymbol{x} \in [0, \infty)^d : g_*(\boldsymbol{x}) \leqslant 1\}$. The scaling sequence r_n can be chosen as $\psi(r_n) \sim \log n$. Moreover, $\max G = (1, \ldots, 1)$.

Proof The mean measure of N_n is given by $n\mathbb{P}(\boldsymbol{X}/r_n \in \cdot)$ with intensity $h_n(\boldsymbol{x}) = nr_n^d f(r_n \boldsymbol{x})$. We show the convergence of that mean measure onto G, implying convergence of scaled samples N_n; see Balkema et al. (2010), Proposition 2.3. By (2.10) and the choice of r_n, we have

$$-\log f(r_n\boldsymbol{x}_n)/\log n \sim -\log f(r_n\boldsymbol{x}_n)/\psi(r_n) \to g_*(\boldsymbol{x}), \quad n \to \infty, \quad \boldsymbol{x}_n \to \boldsymbol{x}. \tag{2.11}$$

Continuous convergence in (2.11) with g_* continuous implies uniform convergence on compact sets. Hence, g_* is bounded on compact sets. For $G = \{g_*(\boldsymbol{x}) \leqslant 1\}$, we have $g_*(\boldsymbol{x}) < 1$ on the interior of G and $g_*(\boldsymbol{x}) > 1$ on the complement of G. Furthermore, applying L'Hôpital's rule and Lemma 1.2(a) in Resnick (1987), we have

$$\log r_n/\psi(r_n) \sim (1/\psi'(r_n))/r_n \to 0, \qquad r_n \to \infty.$$

Combining these results, we see that $-\log h_n(\boldsymbol{x}_n) \sim (g_*(\boldsymbol{x}_n) - 1)\log n$, which diverges to $-\infty$ on the interior of G and to $+\infty$ outside of G. This implies that

$$h_n(\boldsymbol{x}_n) \to \begin{cases} \infty, & \boldsymbol{x} \in G^o, \\ 0, & \boldsymbol{x} \in G^c, \end{cases}$$

giving convergence (in probability) of N_n onto limit set G.

The form of the margins $1 - F_j(x) \sim e^{-\psi_j(x)}$ with $\psi_j(x) \sim \psi(x) \to \infty$ gives $-\log(1 - F_j(x)) \sim \psi(x)$; i.e.,

$$-\log(1 - F_j(r_n)) \sim \psi(r_n) \sim \log n, \qquad n \to \infty.$$

This choice of r_n implies that the coordinate-wise maxima scaled by r_n converge in probability to 1 (Gnedenko (1943); de Haan (1970)), so that $\max G = (1, \ldots, 1)$. □

Remark 1. Condition (2.10) implies that $-\log f$ is multivariate regularly varying on $[0, \infty)^d$. Such densities are referred to as Weibull-like. The limit function g_* is homogeneous of some positive order k: $g_*(t\boldsymbol{x}) = t^k g_*(\boldsymbol{x})$ for all $t > 0$. The gauge function g of the limit set G can thus be obtained from g_* by setting $g(\boldsymbol{x}) = g_*^{1/k}(\boldsymbol{x})$.

When the margins are standard exponential, $\psi(t) = t$. Hence, for the random vector $\boldsymbol{X}_E$ with a Lebesgue density f_E on $\mathbb{R}_+^d$, condition (2.10) is equivalent to

$$-\log f_E(t\boldsymbol{x}_t)/t \to g_*(\boldsymbol{x}), \qquad t \to \infty, \qquad \boldsymbol{x}_t \to \boldsymbol{x}, \qquad \boldsymbol{x} \in [0, \infty)^d \tag{2.12}$$

with the limit function g_* equal to the gauge function g.

Whilst the assumption of a Lebesgue density might appear strict, it is a common feature in statistical practice of extreme value analysis. The assumption permits simple elucidation of the connection between different representations for multivariate extremes. Furthermore, many statistical models, including elliptical distributions and vine copulas (Joe, 1996; Bedford and Cooke, 2001, 2002), are specified most readily in terms of their densities.

Convergence at the density level such as in (2.10) may not always hold. The condition requires the limit function and hence the gauge function of the limit set to be continuous, excluding limit sets for which rays from the origin cross the boundary in more than one point. We provide an example of such a situation in Section 4; see Example 4.1.2. A less restrictive set of sufficient conditions for convergence of sample clouds onto a limit set can be obtained using the survival function. The following proposition is Theorem 2.1 in Davis et al. (1988), with a minor reformulation in terms of scaling.

Proposition 3. *Suppose that the random vector $\boldsymbol{X}$ has support on $[0, \infty)^d$, the margins are asymptotically equal to a von Mises function: $1 - F_j(x) \sim e^{-\psi(x)}$ for $x \to \infty$ $(j = 1, \ldots, d)$, and the joint survival function satisfies*

$$\frac{-\log \mathbb{P}(\boldsymbol{X} \geqslant t\boldsymbol{x})}{\psi(t)} \to g_*(\boldsymbol{x}), \qquad t \to \infty, \qquad \boldsymbol{x} \in [0, \infty)^d \backslash \{0\}. \tag{2.13}$$

Further assume that g_* is strictly increasing, such that $g_*(\boldsymbol{x}) < g_*(\boldsymbol{y})$ if $\boldsymbol{x} \leqslant \boldsymbol{y}$ and $\boldsymbol{x} \neq \boldsymbol{y}$. Then for r_n satisfying $\psi(r_n) \sim \log n$, the sample cloud $N_n = \{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$ converges onto $G = \{\boldsymbol{x} \in [0, \infty)^d : g_*(\boldsymbol{x}) \leqslant 1\}$.

2.5.3 Marginalization

When $d > 2$, a key question is the marginalization from dimension d to dimension $m < d$. We prove below that, as long as the minimum over each coordinate of g is well-defined, then the gauge function determining the limit set in m dimensions is found through minimizing over the coordinates to be marginalized.

A continuous map h from the vector space V into the vector space $\widetilde{V}$ is *positive-homogeneous* if $h(r\boldsymbol{x}) = rh(\boldsymbol{x})$ for all $\boldsymbol{x} \in V$ and all $r > 0$. If $\widetilde{V} = \mathbb{R}^m$, the map h is determined by the m coordinate maps $h_j : V \to \mathbb{R}$, $j = 1, \ldots, m$ and in this case it suffices that these maps are continuous and positive-homogeneous.

Convergence onto a limit set is preserved under linear transformations (e.g., Lemma 4.1 in Nolde (2014)) and more generally under continuous positive-homogeneous maps with the same scaling sequences (Theorem 1.9 in Balkema and Nolde (2020)). A consequence of the latter result, referred to as the Mapping Theorem, is that projections of sample clouds onto lower-dimensional sub-spaces also converge onto a limit set.

Proposition 4. *Let N_n be an n-point sample cloud from a distribution of random vector $\boldsymbol{X}$ on $\mathbb{R}^d$. Assume N_n converges in probability, as $n \to \infty$, onto a limit set $G = \{\boldsymbol{x} \in \mathbb{R}^d : g(\boldsymbol{x}) \leqslant 1\}$ for a gauge*

function g. Let $\widetilde{\boldsymbol{X}} = (X_i)_{i \in I_m}$ denote an m-dimensional marginal of $\boldsymbol{X}$, where $I_m \subset I = \{1, \dots, d\}$ is an index set with $|I_m| = m$. Sample clouds from $\widetilde{\boldsymbol{X}}$ also converge, with the same scaling, and the limit set $\tilde{G} = P_m(G) = \{\boldsymbol{y} \in \mathbb{R}^m : \tilde{g}(\boldsymbol{y}) \leqslant 1\}$, where P_m is a projection map onto the coordinates of $\widetilde{\boldsymbol{X}}$ and

$$\tilde{g}(\boldsymbol{y}) = \min_{\{x_i : i \in I \backslash I_m\}} g(\boldsymbol{x}), \qquad \boldsymbol{x} = (x_1, \dots, x_d), \qquad \boldsymbol{y} = (x_i)_{i \in I_m}.$$

Proof Consider the bivariate case first with $\widetilde{\boldsymbol{X}} = X_2$. Sample clouds from X_2 converge onto the limit set $\tilde{G} \subset \mathbb{R}$, which is the projection of G onto the x_1-coordinate axis, by the Mapping Theorem. The projection is determined by the tangent to the level curve $\{\boldsymbol{x} \in \mathbb{R}^2 : g(\boldsymbol{x}) = 1\}$ orthogonal to the x_1-coordinate axis. Similarly, level curves of the gauge function $\tilde{g}$ of the set $\tilde{G}$ are determined by tangents to the level curves $\{\boldsymbol{x} \in \mathbb{R}^2 : g(\boldsymbol{x}) = c\}$ for $c \in [0, 1]$ orthogonal to the x_1-coordinate axis. These projections correspond to x_1 values which minimize $g(x_1, x_2)$. Sequentially minimizing over each of the coordinates to be marginalized gives the result. ▫

An illustration of this result is given in Section 4.2.

3 Linking representations for extremes to the limit set

For simplicity of presentation, in what follows we standardize to consider exponential margins for the light-tailed case. This choice is convenient when there is positive association in the extremes, but hides structure related to negative dependence. We comment further on this case in Section 5. Owing to the standardized marginals, it makes sense to refer to the limit set, rather than a limit set.

Connections between multivariate and hidden regular variation are well established, with the latter requiring the former for proper definition. Some connection between regular variation and conditional extremes was made in Heffernan and Resnick (2007) and Das and Resnick (2011), although they did not specify to exponential-tailed margins. The shape of the limit set has been linked to the asymptotic (in)dependence structure of a random vector (Balkema and Nolde, 2010, 2012). Asymptotic independence is related to the position of mass from convergence (2.1) on $\mathbb{E}$, but regular variation and the existence of a limit set in suitable margins are different conditions and one need not imply the other. Nolde (2014) links the limit set G to the coefficient of residual tail dependence, η_D.

In this section we present some new connections between the shape of the limit set, when it exists, and normalizing functions in conditional extreme value theory, the residual tail dependence coefficient, the function $\lambda(\boldsymbol{\omega})$ and the coefficients $\tau_C(\delta)$.

3.1 Conditional extremes

For the conditional extreme value model, the form of the normalizing functions $\boldsymbol{a}^j, \boldsymbol{b}^j$ is determined by the pairwise dependencies between $(X_{E,i}, X_{E,j})$, $i \in D \backslash j$. The two-dimensional marginalization of any

d-dimensional gauge function is given by Proposition 4, and we simply denote this by g here.

Proposition 5. *Suppose that for $\boldsymbol{X}_E = (X_{E,1}, X_{E,2})$ convergence (2.9) and assumption (2.12) hold, where the domain of K^j includes $(0, \infty)$. Define $\alpha_j = \lim_{x \to \infty} a^j(x)/x$, $j = 1, 2$. Then*

(i) *$g(1, \alpha_1) = 1$, $g(\alpha_2, 1) = 1$.*

(ii) *Suppose that $-\log f_E(t\boldsymbol{x}_t)/t = g(\boldsymbol{x}_t) + v(t)$, with $v(t) \in \mathrm{RV}_{-1}^{\infty}$ or $v(t) = o(1/t)$, and $a^j(t) = \alpha_j t + B^j(t)$ with either $B^j(t)/b^j(t) \in \mathrm{RV}_0^{\infty}$, or $B^j(t) = o(b^j(t))$. For $\beta_1, \beta_2 \leqslant 1$, if $g(1, \alpha_1 + \cdot) - 1 \in \mathrm{RV}_{1/(1-\beta_1)}^0$, then $b^1(x) \in \mathrm{RV}_{\beta_1}^{\infty}$; similarly if $g(\alpha_2 + \cdot, 1) - 1 \in \mathrm{RV}_{1/(1-\beta_2)}^0$, then $b^2(x) \in \mathrm{RV}_{\beta_2}^{\infty}$.*

(iii) *If there are multiple values α satisfying $g(1, \alpha) = 1$, then α_1 is the maximum such α, and likewise for α_2.*

Before the proof of Proposition 5, we give some geometric intuition. Figure 2 presents several examples of the unit level set of possible gauge functions, illustrating the shape of the limit set, for two-dimensional random vectors with exponential margins. On each figure, the slope of the red line indicates the value of α_1; i.e., the equation of the red line is $y = \alpha_1 x$. Intuitively, conditional extreme value theory poses the question: "given that variable X is growing, how does variable Y grow as a function of X?". We can now see that this is neatly described by the shape of the limit set: to first order, the values of Y occurring with large X are determined by the direction for which X is growing at its maximum rate. The necessity of a scale normalization in the conditional extreme value limit depends on the local curvature and particularly the rate at which $g(\alpha_1 + u, 1)$ approaches 1 as $u \to 0$. For cases (i), (iv), (v) and (vi) of Figure 2, the function approaches zero linearly in u: as a consequence $b^j(t) \in \mathrm{RV}_0^{\infty}$. For case (ii) the order of decay is u^2 and so $b^j(t) \in \mathrm{RV}_{1/2}^{\infty}$, whilst for (iii) the order is $u^{1/\theta}$ so $b^j(t) \in \mathrm{RV}_{1-\theta}^{\infty}$.

The class of distributions represented by gauge function (vi) (bottom left) can be thought of as those arising from a mixture of distributions with gauge functions (i) and (iii), up to differences in parameter values. In such an example, there are two normalizations that would lead to a non-degenerate limit in (2.8), but ruling out mass at infinity produces the unique choice $\alpha_1 = \alpha_2 = 1$, $\beta_1 = \beta_2 = 0$. If instead we chose to rule out mass at $-\infty$, then we would have $\alpha_1 = \alpha_2 = 0$ and $\beta_1 = \beta_2 = 1 - \theta$.

Proof of Proposition 5. In all cases we just prove one statement as the other follows analogously.
(i) By assumption (2.12), $(X_{E,1}, X_{E,2})$ have a joint density f_E and so conditional extremes convergence (2.9) can be expressed as

$$b^1(t) f_E(t, b^1(t)z + a^1(t)) e^t \to k^1(z) =: e^{-h^1(z)}, \qquad t \to \infty, \quad z \in [\lim_{t \to \infty} -a^1(t)/b^1(t), \infty)$$

with $k^1 = e^{-h^1}$ a density. Taking logs, we have

$$-\log f_E(t, b^1(t)z + a^1(t)) - t - \log b^1(t) \to h^1(z), \qquad t \to \infty. \tag{3.14}$$

Now use assumption (2.12) with $\boldsymbol{x}_t = (1, x_t) = (1, a^1(t)/t + zb^1(t)/t)$. That is,

$$-\log f_E(t, b^1(t)z + a^1(t)) = tg(1, x)[1 + o(1)] = tg(1, x_t)[1 + o(1)], \tag{3.15}$$

with $x = \lim_{t \to \infty} a^1(t)/t + zb^1(t)/t$. As the support of K^1 includes $(0, \infty)$, $h^1(z) < \infty$ for all $z \in (0, \infty)$, and combining (3.14) and (3.15) we have

$$g(1, x_t)[1 + o(1)] = 1 + h^1(z)/t + \log b^1(t)/t + o(1/t). \tag{3.16}$$

Suppose that $b^1(t)/t \to \gamma > 0$. Then $x_t \to \alpha_1 + \gamma z$, and taking $t \to \infty$ in (3.16) leads to $g(1, \alpha_1 + \gamma z) = 1$ for any z. But since the coordinatewise supremum of G is $(1, 1)$, $g(x, y) \geqslant \max(x, y)$ which would entail $z \leqslant (1 - \alpha_1)/\gamma$. No such upper bound applies, so we conclude $\gamma = 0$, i.e., $b^1(t) = o(t)$. Now taking limits in (3.16) leads to $g(1, \alpha_1) = 1$.

(ii) Let $g(1, \alpha_1 + u) - 1 =: r(u) \in \mathrm{RV}_\rho^0$, $\rho > 0$. We also have from (3.16) $g(1, \alpha_1 + b^1(t)/t + B^1(t)/t) - 1 = h^1(1)/t + \log b^1(t)/t - v(t) + o(1/t)$, so that the function $b^1(t)$ is a solution to the equation

$$r(b^1(t)/t + B^1(t)/t) = h^1(1)/t + \log b^1(t)/t - v(t) + o(1/t). \tag{3.17}$$

Equation (3.17) admits a solution if b^1 is regularly varying at infinity. A rearrangement provides that

$$b^1(t) = tr^{-1}(h^1(1)/t + \log b^1(t)/t - v(t) + o(1/t))[1 + B^1(t)/b^1(t)]^{-1};$$

if b^1 is regularly varying then $\log b^1(t)/t \in \mathrm{RV}_{-1}^\infty$, so that using the fact that $v(t) \in \mathrm{RV}_{-1}^\infty$ or $v(t) = o(1/t)$, combined with $r^{-1} \in \mathrm{RV}_{1/\rho}^0$, yields $b^1(t) \in \mathrm{RV}_{1-1/\rho}^\infty$. We now argue that such a solution is unique in this context. We know that the normalization functions a^1, b^1 lead to a non-degenerate distribution K^1 that places no mass at infinity. By the convergence to types theorem (Leadbetter et al. (1983, p.7), see also part (iii) of this proof), any other function $\tilde{b}^1$ leading to a non-degenerate limit with no mass at infinity must satisfy $\tilde{b}^1(t) \sim db^1(t)$, $t \to \infty$, for some $d > 0$, so that $\tilde{b}^1 \in \mathrm{RV}_{1-1/\rho}^\infty$ also. Finally, setting $\beta_1 = 1 - 1/\rho$ gives $b^1 \in \mathrm{RV}_{\beta_1}^\infty$.

(iii) Suppose that

$$\mathbb{P}\left(\frac{X_{E,2} - a^1(t)}{b^1(t)} \leqslant z \,\Big|\, X_{E,1} > t\right) \to K^1(z), \quad \mathbb{P}\left(\frac{X_{E,2} - \tilde{a}^1(t)}{\tilde{b}^1(t)} \leqslant z \,\Big|\, X_{E,1} > t\right) \to \tilde{K}^1(z),$$

where neither K^1 nor $\tilde{K}^1$ has mass at $+\infty$. Then by the convergence to types theorem, $\tilde{a}^1(t) = a^1(t) + cb^1(t) + o(b^1(t))$ and $\tilde{b}^1(t) = db^1(t) + o(b^1(t))$, for some $d > 0$, and $\tilde{K}^1(z) = K^1(z/d + c)$. As such, $\tilde{a}^1(t)/t \sim a^1(t)/t \sim \alpha_1$. We conclude that if there was a non-degenerate $\tilde{K}^1$ limit for which $\tilde{a}^1(t)/t \sim \tilde{\alpha}_1 > \alpha_1$ then K^1 must place mass at $+\infty$; since by assumption it does not, then α_1 is the maximum value satisfying $g(1, \alpha_1) = 1$.   □

For distributions whose sample clouds converge onto a limit set described by a gauge function with piecewise-continuous partial derivatives possessing finite left and right limits, further detail can be given about β_j.
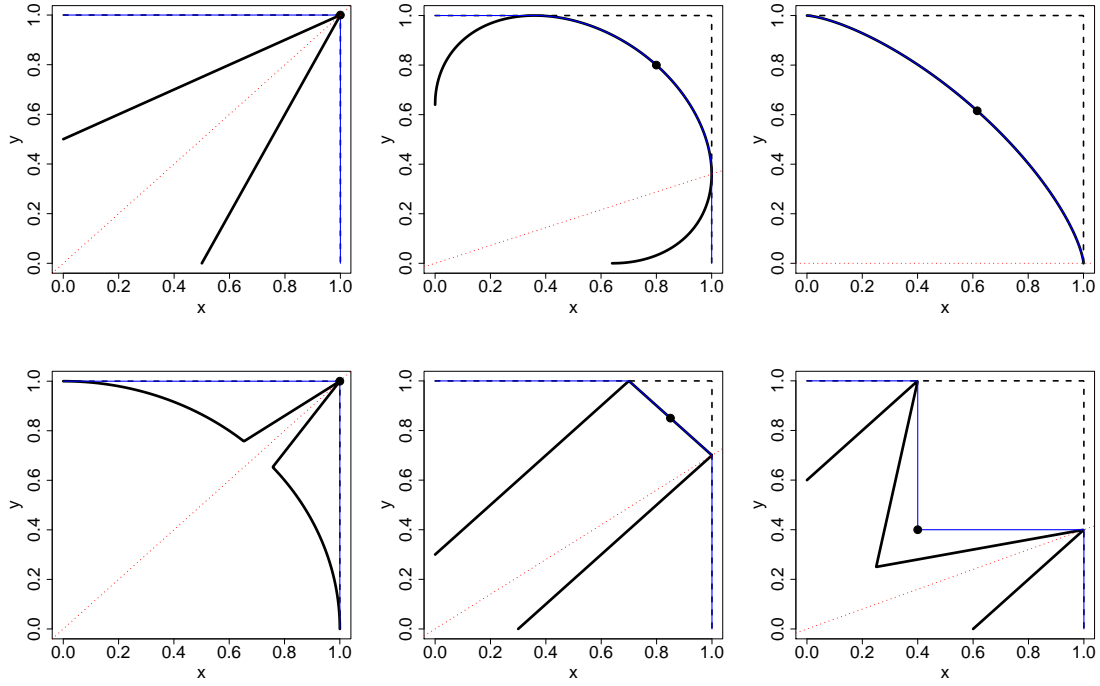
Figure 2: Unit level sets of six possible gauge functions for bivariate random vectors with exponential margins. The limit set G is the set bounded by these level sets and the axes. In each case the red line is $y = \alpha_1 x$, the blue lines represent $\lambda(\omega, 1 - \omega)$ and dots $\eta_{1,2}$ (see Section 3.2). Dashed lines represent the boundary $\max(x, y) = 1$. Clockwise from top left, the gauge functions represented are: (i) $\max(x, y)/\theta + (1 - 1/\theta) \min(x, y)$; (ii) $(x + y - 2\theta\sqrt{xy})/(1 - \theta^2)$; (iii) $(x^{1/\theta} + y^{1/\theta})^\theta$; (iv) $\max\{(x - y)/\theta, (y - x)/\theta, \min(x - \mu y, y - \mu x)/(1 - \theta - \mu)\}$; (v) $\max((x - y)/\theta, (y - x)/\theta, (x + y)/(2 - \theta))$; (vi) $\min\{\max(x, y)/\theta_1 + (1 - 1/\theta_1) \min(x, y), (x^{1/\theta_2} + y^{1/\theta_2})^{\theta_2}\}$. In each case $\theta \in (0, 1)$; in some cases the endpoints are permitted as well. For case (iv), $\theta + \mu < 1$.

Proposition 6. *Let G be a limit set whose gauge function g has piecewise-continuous partial derivatives $g_1(x,y) = \partial g(x,y)/\partial x$, $g_2(x,y) = \partial g(x,y)/\partial y$ possessing finite left and right limits, and for which the conditions of Proposition 5 hold. Then (i) $\beta_1 \geqslant 0$ if $g_2(1,(\alpha_1)_+) = 0$; (ii) $\beta_1 = 0$ if $0 < g_2(1,(\alpha_1)_+) < \infty$. Further, if $\alpha_1 > 0$ then $0 \leqslant g_2(1,(\alpha_1)_+) < \infty$, so that $\beta_1 \geqslant 0$. Analogous statements hold for α_2, β_2.*

Proof. Consider the partial derivative

$$g_2(1,(\alpha_1)_+) = \lim_{u \to 0^+} \frac{g(1,\alpha_1 + u) - g(1,\alpha_1)}{u} \geqslant 0,$$

as $g(1,\alpha_1 + u) \geqslant g(1,\alpha_1)$. We note $g(1,\alpha_1) = 1$, such that $g(1,\alpha_1 + u) - 1 \sim u g_2(1,(\alpha_1)_+)$, $u \to 0^+$. Since this is regularly varying with index $1/(1 - \beta_1)$ by assumption, $g_2(1,(\alpha_1)_+) = 0$ implies $g(1,\alpha_1 + u) - 1 = o(u)$, hence $1/(1 - \beta_1) \geqslant 1$, and $0 < g_2(1,(\alpha_1)_+) < \infty$ implies $1/(1 - \beta_1) = 1$, so (i) and (ii) follow. If g is differentiable at the point $(1,\alpha_1)$, then since $g(1,y) \geqslant 1$, $g_2(1,(\alpha_1)_+) = g_2(1,(\alpha_1)_-) = 0$ and (i) holds. Otherwise, in a neighbourhood of $(1,\alpha_1)$, we can express

$$g(x,y) = \begin{cases} \hat{g}(x,y), & y \leqslant \alpha_1 x \\ \tilde{g}(x,y), & y \geqslant \alpha_1 x, \end{cases}$$

where the homogeneous functions $\tilde{g}$ and $\hat{g}$ have continuous partial derivatives at $(1,\alpha_1)$. Euler's homogeneous function theorem gives $1 = \tilde{g}_1(1,\alpha_1) + \tilde{g}_2(1,\alpha_1)\alpha_1 = g_1(1_-,\alpha_1) + \alpha_1 g_2(1,(\alpha_1)_+)$ so that for $\alpha_1 > 0$, $g_2(1,(\alpha_1)_+) < \infty$, and hence (i) or (ii) hold.   □

We remark on links with existing work on conditional extreme value limits for variables with a polar-type representation, whereby $(X_1, X_2) = R(W_1, W_2)$ for $R > 0$ and (W_1, W_2) constrained by some functional dependence. Abdous et al. (2005), Fougères and Soulier (2010) and Seifert (2014) consider a type of conditional extremes limit for certain such polar constructions, where in the light-tailed case, the shape of the constraint on (W_1, W_2) feeds into the normalization and limit distribution. However, limit sets are sensitive to marginal choice, and because the above papers do not consider conditional extreme value limits in standardized exponential-tailed margins, further connections are limited.

3.2 Different scaling orders: $\lambda(\boldsymbol{\omega})$

We now focus on the connection with $\lambda(\boldsymbol{\omega})$, as defined in Section 2.3. When $\boldsymbol{\omega} = (1/d,\ldots,1/d)$, this yields the link with the residual tail dependence coefficient η_D, which has already been considered in Nolde (2014). Define the region

$$R_{\boldsymbol{\omega}} = \left(\frac{\omega_1}{\max(\omega_1,\ldots,\omega_d)}, \infty\right] \times \cdots \times \left(\frac{\omega_d}{\max(\omega_1,\ldots,\omega_d)}, \infty\right].$$

Proposition 7. *Suppose that the sample cloud $N_n = \{\boldsymbol{X}_E^1/\log n, \ldots, \boldsymbol{X}_E^n/\log n\}$ converges onto a limit set G, and that for each $\boldsymbol{\omega} \in \mathcal{S}_{\Sigma}$, equation (2.7) holds. Then*

$$\lambda(\boldsymbol{\omega}) = \max(\boldsymbol{\omega}) \times r_{\boldsymbol{\omega}}^{-1},$$

where

$$r_{\boldsymbol{\omega}} = \min\{r \in [0,1] : rR_{\boldsymbol{\omega}} \cap G = \varnothing\}.$$

Corollary 1 (Nolde (2014))**.**

$$1/\eta_D = d\lambda(1/d,\ldots,1/d) = r_{(1/d,\ldots,1/d)}^{-1} = \left[\min\left\{r \in [0,1] : (r,\infty]^d \cap G = \varnothing\right\}\right]^{-1}.$$

Proof of Proposition 7. The proof follows very similar lines to Proposition 2.1 of Nolde (2014). Firstly note that $\lambda(\boldsymbol{\omega}) = \kappa(\boldsymbol{\omega})$, where $\kappa : [0,\infty)^d\backslash\{\mathbf{0}\} \to (0,\infty)$ is a 1-homogeneous function defined by

$$\mathbb{P}(\boldsymbol{X}_E > \boldsymbol{\beta}t) = \ell(e^t;\boldsymbol{\beta})e^{-t\kappa(\boldsymbol{\beta})}, \qquad \ell(\cdot;\boldsymbol{\beta}) \in \mathrm{RV}_0^\infty \text{ for every } \boldsymbol{\beta} \in [0,\infty)^d\backslash\{\mathbf{0}\}.$$

As a consequence,

$$\frac{\lambda(\boldsymbol{\omega})}{\max(\boldsymbol{\omega})} = \lim_{t\to\infty} -\log\mathbb{P}\{\boldsymbol{X}_E > t\boldsymbol{\omega}/\max(\boldsymbol{\omega})\}/t. \tag{3.18}$$

Without loss of generality, suppose that $\max(\boldsymbol{\omega}) = \omega_d$, so that $R_{\boldsymbol{\omega}} = (\omega_1/\omega_d,\infty] \times \cdots \times (\omega_{d-1}/\omega_d,\infty] \times (1,\infty]$. Because of the convergence of the sample cloud onto G, we have by Proposition 1 that for any $\epsilon > 0$ and large enough t

$$\mathbb{P}(\boldsymbol{X}_E \in te^\epsilon r_{\boldsymbol{\omega}}R_{\boldsymbol{\omega}}) \leqslant \mathbb{P}(\boldsymbol{X}_E \in tR_{(0,\ldots,0,1)}) = e^{-t} \leqslant \mathbb{P}(\boldsymbol{X}_E \in te^{-\epsilon}r_{\boldsymbol{\omega}}R_{\boldsymbol{\omega}}),$$

implying $-\log\mathbb{P}(\boldsymbol{X}_E \in tr_{\boldsymbol{\omega}}R_{\boldsymbol{\omega}}) \sim t$. Therefore $-\log\mathbb{P}(\boldsymbol{X}_E \in tR_{\boldsymbol{\omega}}) \sim tr_{\boldsymbol{\omega}}^{-1}$, and combining with equation (3.18) gives the result.

$\square$

Figure 3 illustrates some of the concepts used in the proof of Proposition 7 when $d = 2$ and $\boldsymbol{\omega} = (\omega, 1 - \omega)$.

The blue lines in Figure 2 represent $\lambda(\boldsymbol{\omega})$, depicting the unit level set of $\lambda(\omega, 1 - \omega)/\max(\omega, 1 - \omega)$, and the dots illustrate the value of $r_{1/2} = \eta_{1,2}$. We can now see clearly how, in two dimensions, different dependence features are picked out by the conditional extremes representation and hidden regular variation based on $\eta_{1,2}$. Often, values of $\eta_{1,2} > 1/2$ or $\alpha > 0$ are associated with positive extremal dependence. From example (iv) of Figure 2 (bottom right), we observe $\eta_{1,2} < 1/2$ but $\alpha > 0$. We have that Y does grow with X (and vice versa) but only at a specific rate. On the other hand, joint extremes, where (X, Y) take similar values, are rare, occurring less frequently than under independence.

From example (iv) we can also see that one of the conclusions following Proposition 2.1 in Nolde (2014) is not true: the point $(r_{1/2}, r_{1/2})$ need not lie on the boundary of G, meaning that we do not necessarily have $\eta_D = 1/g(\mathbf{1})$, although we can deduce the bound $\eta_D \geqslant 1/g(\mathbf{1})$. Similarly, there are occasions when $g(r_{\boldsymbol{\omega}}\boldsymbol{\omega}/\max(\boldsymbol{\omega})) = 1$, implying $\lambda(\boldsymbol{\omega}) = g(\boldsymbol{\omega})$, but clearly this is not always true. In Proposition 8, we resolve when this is the case by representing $r_{\boldsymbol{\omega}}$ in terms of g.
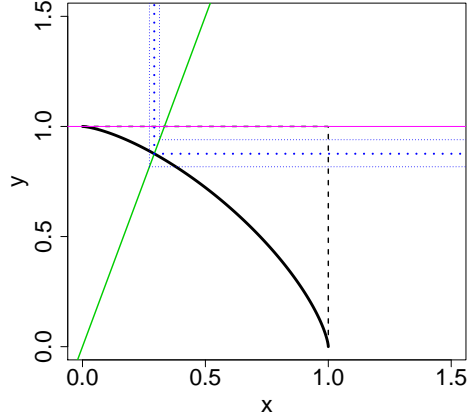
Figure 3: Illustration of the concepts used in the proof of Proposition 7: the green line represents the ray $x = \{\omega/(1-\omega)\}y$, $\omega < 1/2$. The region above the purple line represents $R_{(0,1)}$, the region to the north-east of the thick dotted blue lines represents $r_\omega R_\omega$, whilst the two sets of thin dotted blue lines illustrate the regions $e^{-\epsilon}r_\omega R_\omega$ and $e^\epsilon r_\omega R_\omega$. The ratio of the distance from the origin to where the green line intersects the boundary of the limit set G, and where the green line intersects the boundary $\max(x,y) = 1$, is equal to r_ω.

Define $B_{\boldsymbol{\omega}}$ to be the boundary of the region $R_{\boldsymbol{\omega}}$, i.e.,

$$B_{\boldsymbol{\omega}} = \bigcup_{i=1}^{d}\{\boldsymbol{x} \in \mathbb{R}_+^d : x_i = \omega_i/\max(\boldsymbol{\omega}), x_j \geqslant \omega_j/\max(\boldsymbol{\omega}), j \neq i\}.$$

Proposition 8. *Assume the conditions of Proposition 7. Then*

$$r_{\boldsymbol{\omega}} = \left[\min_{\boldsymbol{y} \in B_{\boldsymbol{\omega}}} g(\boldsymbol{y})\right]^{-1}, \quad \text{and hence} \quad \lambda(\boldsymbol{\omega}) = \max(\boldsymbol{\omega}) \times \min_{\boldsymbol{y} \in B_{\boldsymbol{\omega}}} g(\boldsymbol{y}).$$

From Proposition 8, we observe that $\lambda(\boldsymbol{\omega}) = g(\boldsymbol{\omega})$ if $\arg\min_{\boldsymbol{y} \in B_{\boldsymbol{\omega}}} g(\boldsymbol{y}) = \boldsymbol{\omega}/\max(\boldsymbol{\omega})$, i.e., the vertex of the set $B_{\boldsymbol{\omega}}$. The proof of Proposition 8 is deferred until after Proposition 11, for which the proof is very similar.

Remark 2. We note that $\min_{\boldsymbol{y} \in B_{\boldsymbol{\omega}}} g(\boldsymbol{y}) = \min_{\boldsymbol{y} \in R_{\boldsymbol{\omega}}} g(\boldsymbol{y})$.

3.3 Coefficients $\tau_C(\delta)$

3.3.1 Connections to limit set G

In two dimensions, the coefficients $\tau_1(\delta)$ and $\tau_2(\delta)$ provide a somewhat complementary concept to the function $\lambda(\boldsymbol{\omega})$. Rather than considering the impact of the limit set G on the shape of the function defined by both variables exceeding thresholds growing at different rates, we are considering what is occurring

when one variable exceeds a growing threshold and the other is upper bounded by a certain lesser growth rate. The left and centre panels in Figure 4 provide an illustration of $\lambda(\boldsymbol{\omega})$ and $\tau_j(\delta)$ in two dimensions.

Define the region $R_{C,\delta} = (1,\infty]^C \times [0,\delta]^{D\setminus C} = \{\boldsymbol{x} : x_i \in (1,\infty], i \in C, x_j \in [0,\delta], j \in D\setminus C\}$, so that, for example, when $d = 3$, $R_{\{1,3\},\delta} = (1,\infty] \times [0,\delta] \times (1,\infty]$.

Proposition 9. *Suppose that the sample cloud $N_n = \{\boldsymbol{X}_E^1/\log n, \ldots, \boldsymbol{X}_E^n/\log n\}$ converges onto a limit set G, and that the assumption in equation (2.6) holds. For $\delta \in [0,1]$, and $C \subset D$,*

$$\tau_C(\delta) = r_{C,\delta} = \min\{r \in [0,1] : rR_{C,\delta} \cap G = \varnothing\}.$$

The coefficient $\tau_D = \eta_D$, and does not depend on δ.

Proof. The coefficient τ_D describes the order of hidden regular variation on the cone $(0,\infty]^d$, which is precisely the same as η_D. For $\tau_C(\delta)$, $|C| < d$, we consider the function of t

$$\mathbb{P}(\min_{i \in C} X_{P,i} > tx, \max_{j \in D\setminus C} X_{P,j} \leqslant yt^\delta) \in \mathrm{RV}^\infty_{-1/\tau_C(\delta)}, \qquad 0 < x, y < \infty.$$

Take $x = y = 1$. Then

$$\begin{aligned}
\tau_C(\delta) &= \lim_{t\to\infty} \frac{-\log \mathbb{P}(X_{P,1} > t)}{-\log \mathbb{P}(\min_{i \in C} X_{P,i} > t, \max_{j \in D\setminus C} X_{P,j} \leqslant t^\delta)} \\
&= \lim_{t\to\infty} \frac{-\log \mathbb{P}(X_{E,1} > t)}{-\log \mathbb{P}(\min_{i \in C} X_{E,i} > t, \max_{j \in D\setminus C} X_{E,j} \leqslant \delta t)},
\end{aligned} \tag{3.19}$$

where the denominator in (3.19) can be expressed $-\log \mathbb{P}(\boldsymbol{X}_E \in tR_\delta)$. As in the proof of Proposition 7, the convergence onto the limit set and exponential margins enables us to conclude that $-\log \mathbb{P}(\boldsymbol{X}_E \in tr_{C,\delta}R_\delta) \sim t$, and hence $-\log \mathbb{P}(\boldsymbol{X}_E \in tR_\delta) \sim tr_{C,\delta}^{-1}$. Combining with (3.19) gives $\tau_C(\delta) = r_{C,\delta}$. □

In the two-dimensional case, it is possible to express $\tau_j(\delta)$ simply in terms of the gauge function. For higher dimensions, we refer to Proposition 11.

Proposition 10. *Assume the conditions of Proposition 9. When $d = 2$, $\tau_1(\delta) = [\min_{\gamma \in [0,\delta]} g(1,\gamma)]^{-1}$ and $\tau_2(\delta) = [\min_{\gamma \in [0,\delta]} g(\gamma,1)]^{-1}$.*

Proof. For $\gamma \in [0,1]$, the points $(1/g(1,\gamma), \gamma/g(1,\gamma))$ lie on the curve $\{(x,y) \in [0,1]^2 : g(x,y) = 1, x \geqslant y\}$. The value $r_{1,\delta}$ is the maximum value of $1/g(1,\gamma)$ for $\gamma \in [0,\delta]$, hence $\tau_1(\delta) = [\min_{\gamma \in [0,\delta]} g(1,\gamma)]^{-1}$. A symmetric argument applies to $\tau_2(\delta)$. □

The right panel of Figure 4 provides an illustration: in blue the value of δ is such that $\tau_1(\delta) < 1$; in red the value of δ is such that $\tau_2(\delta) = 1$. Further detail on this example is given in Section 4.1.5.

The question arises: does Proposition 10 still hold for $d > 2, |C| = 1$? Let g_C denote the gauge function for the limit set of $(X_{E,j} : j \in C)$. By Proposition 4, we know that $g_{i,j}(x_i, x_j) = \min_{\boldsymbol{x}_{-i,-j} \in [0,\infty)^{d-2}} g_D(\boldsymbol{x})$. As such, equality will hold if $\arg\min_{\boldsymbol{x}_{-i,-j} \in [0,\infty)^{d-2}} g_D(\boldsymbol{x}) \in [0,\delta]^{d-2}$. Note that the dimension does
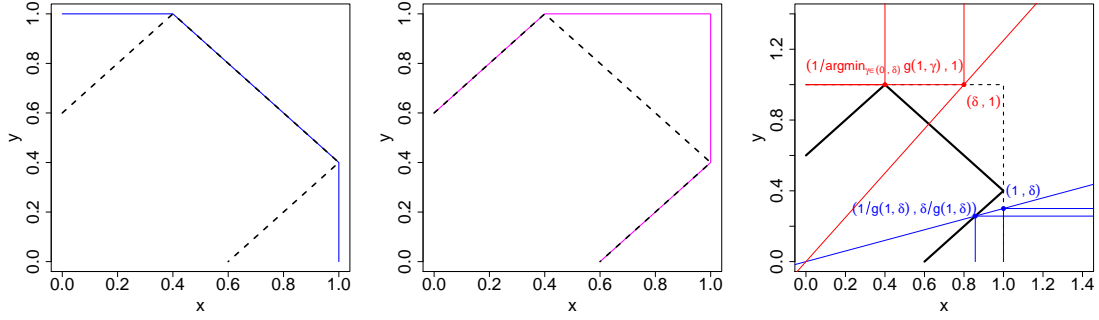
Figure 4: Illustration of $\lambda(\boldsymbol{\omega})$ and $\tau_j(\delta)$ for the gauge function $g(x,y) = \max\{(x-y)/\theta, (y-x)/\theta, (x+y)/(2-\theta)\}$. The left panel illustrates $\lambda(\omega, 1-\omega)$ in blue. The centre panel illustrates $\tau_j(\delta)$ in purple: $\tau_1(\delta), \delta \in [0,1]$ is represented by the values below the main diagonal, whilst $\tau_2(\delta)$ is represented by those values above. The set G is added on both panels with dashed lines. The right panel illustrates $\tau_1(\delta)$ and $\tau_2(\delta)$ in terms of the gauge function.

indeed play a key role here: when looking at $\tau_j(\delta)$ for a d-dimensional problem, we are looking at the situation where $d-1$ coordinates are upper bounded by a growth rate determined by δ. In contrast, when marginalizing and looking at $\tau_j(\delta)$ for a 2-dimensional problem, the $d-2$ coordinates that we have marginalized over are unrestricted and so can represent small or large values. As such, the answer to our question is negative in general.

Proposition 11 details the precise value of $\tau_C(\delta)$ in terms of g for any dimension d. In a similar spirit to Section 3.2, define the boundary of the region $R_{C,\delta}$ as

$$B_{C,\delta} = B_{C,\delta}^1 \cup B_{C,\delta}^\delta,$$

where

$$B_{C,\delta}^1 = \bigcup_{i \in C}\{\boldsymbol{x} \in \mathbb{R}_+^d : x_i = 1, x_j \geqslant 1 \; \forall j \in C\backslash i, \; x_k \leqslant \delta \; \forall k \in D\backslash C\}$$

$$B_{C,\delta}^\delta = \bigcup_{i \in D\backslash C}\{\boldsymbol{x} \in \mathbb{R}_+^d : x_i = \delta, x_j \leqslant \delta \; \forall j \in (D\backslash C)\backslash i, \; x_k \geqslant 1 \; \forall k \in C\},$$

so, for example, when $d = 3$,

$$B_{\{1,3\},\delta} = \{\boldsymbol{x} \in \mathbb{R}_+^3 : x_1 = 1, x_2 \leqslant \delta, x_3 \geqslant 1\} \cup \{\boldsymbol{x} \in \mathbb{R}_+^3 : x_1 \geqslant 1, x_2 = \delta, x_3 \geqslant 1\} \cup \{\boldsymbol{x} \in \mathbb{R}_+^3 : x_1 \geqslant 1, x_2 \leqslant \delta, x_3 = 1\}.$$

For $C = D$, $R_D = (1,\infty]^d$, and $B_D = \{\boldsymbol{x} : \min(\boldsymbol{x}) = 1\}$.

Proposition 11. *Assume the conditions of Proposition 9. For any $C \subseteq D$,*

$$\tau_C(\delta) = \left[\min_{\boldsymbol{y} \in B_{C,\delta}} g(\boldsymbol{y})\right]^{-1} = \left[\min_{\boldsymbol{y} \in B_{C,\delta}^1} g(\boldsymbol{y})\right]^{-1}.$$

Proof. The vertex of the region $R_{C,\delta}$, or its boundary $B_{C,\delta}$, which has components 1 on the coordinates indexed by C, and δ in the other coordinates, lies on $\mathcal{S}_\vee := \{\boldsymbol{x} \in \mathbb{R}_+^d : \max(\boldsymbol{x}) = 1\}$. The region $G \subseteq \mathcal{S}_\vee$, and because the coordinatewise supremum of G is $\boldsymbol{1}$, the boundary of G intersects with $\mathcal{S}_\vee$. Now consider scaling the region $R_{C,\delta}$ by $r_{C,\delta} \in (0,1]$ until it intersects with G. The point of intersection must lie on the boundary of the scaled region $r_{C,\delta} R_{C,\delta}$, i.e., on $r_{C,\delta} B_{C,\delta}$, and on the boundary of G, $\{\boldsymbol{x} \in \mathbb{R}_+^d : g(\boldsymbol{x}) = 1\}$. Therefore, there exists $\boldsymbol{x}^\star \in B_{C,\delta}$ such that $g(r_{C,\delta} \boldsymbol{x}^\star) = 1$, which is rearranged to give $\tau_C(\delta) = r_{C,\delta} = 1/g(\boldsymbol{x}^\star)$. Furthermore, we must have that such a point $\boldsymbol{x}^\star = \arg\min_{\boldsymbol{y} \in B_{C,\delta}} g(\boldsymbol{y})$, otherwise there exists some $\boldsymbol{x}' \in B_{C,\delta}$ such that $g(\boldsymbol{x}') < g(\boldsymbol{x}^\star)$ and so $g(r_{C,\delta} \boldsymbol{x}') < 1$, meaning that $r_{C,\delta} \neq \min\{r \in (0,1] : r R_{C,\delta} \cap G = \varnothing\}$. We conclude that $\boldsymbol{x}^\star = \arg\min_{\boldsymbol{y} \in B_{C,\delta}} g(\boldsymbol{y})$, so $\tau_C(\delta) = 1/\min_{\boldsymbol{y} \in B_{C,\delta}} g(\boldsymbol{y})$.

To show that $\arg\min_{\boldsymbol{y} \in B_{C,\delta}} g(\boldsymbol{y}) \in B_{C,\delta}^1$, let $\bar{\boldsymbol{x}} = \arg\min_{\boldsymbol{y} \in B_{C,\delta}^1} g(\boldsymbol{y})$, $\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{y} \in B_{C,\delta}^\delta} g(\boldsymbol{y})$, and let $\tilde{x}_l = \min_{k \in C} \tilde{x}_k \geqslant 1$. Then $g(\tilde{\boldsymbol{x}}) = \tilde{x}_l g(\tilde{\boldsymbol{x}}/\tilde{x}_l)$, but $\tilde{\boldsymbol{x}}/\tilde{x}_l \in B_{C,\delta}^1$, so $g(\tilde{\boldsymbol{x}}/\tilde{x}_l) \geqslant g(\bar{\boldsymbol{x}})$ and hence $\tilde{x}_l g(\tilde{\boldsymbol{x}}/\tilde{x}_l) \geqslant g(\bar{\boldsymbol{x}})$ as $\tilde{x}_l \geqslant 1$. □

When $d = 2$ and $|C| = 1$, we note that $B_{\{j\},\delta}^1 = \{\boldsymbol{x} : x_j = 1, x_i \leqslant \delta\}$, which gives the equality in Proposition 10.

Proof of Proposition 8. The proof follows exactly as for the first equality in Proposition 11, replacing $R_{C,\delta}, B_{C,\delta}$ and $r_{C,\delta}$ with $R_{\boldsymbol{\omega}}, B_{\boldsymbol{\omega}}$ and $r_{\boldsymbol{\omega}}$. □

3.3.2 Estimation of coefficients $\tau_C(\delta)$

When $C = D$, equation (2.6) yields $\mathbb{P}(\min_{i \in D} X_{P,i} > t) \in \mathrm{RV}_{-1/\tau_D}^\infty$, implying that τ_D can be estimated as the reciprocal of the tail index of the so-called structure variable $\min_{i \in D} X_{P,i}$. This is identical to estimating the residual tail dependence coefficient η_D, for which the Hill estimator is commonly employed. However, for C with $|C| < d$, we assume $\mathbb{P}(\min_{i \in C} X_{P,i} > t, \max_{j \in D \backslash C} X_{P,j} < t^\delta) \in \mathrm{RV}_{-1/\tau_C(\delta)}^\infty$, but this representation does not lend itself immediately to an estimation strategy, as there is no longer a simple structure variable for which $1/\tau_C(\delta)$ is the tail index.

In order to allow estimation, Simpson et al. (2020) considered $\mathbb{P}(\min_{i \in C} X_{P,i} > t, \max_{j \in D \backslash C} X_{P,j} < (\min_{i \in C} X_{P,i})^\delta)$, but they only offered empirical evidence that the assumed index of regular variation for this probability was the same as in equation (2.6). We now prove this to be the case.

Define $R_{C,\delta}^x = \{\boldsymbol{x} \in \mathbb{R}_+^d : x_i > 1, i \in C, \ x_j \leqslant \delta \min_{l \in C} x_l, j \in D \backslash C\}$, and $B_{C,\delta}^x = B_{C,\delta}^{x,1} \cup B_{C,\delta}^{x,\delta}$ to be its boundary, where

$$B_{C,\delta}^{x,1} = B_{C,\delta}^1 = \bigcup_{i \in C} \{\boldsymbol{x} \in \mathbb{R}_+^d : x_i = 1, x_j \geqslant 1 \ \forall j \in C \backslash i, \ x_k \leqslant \delta \min_{l \in C} x_l = \delta \ \forall k \in D \backslash C\}$$

$$B_{C,\delta}^{x,\delta} = \bigcup_{i \in D \backslash C} \{\boldsymbol{x} \in \mathbb{R}_+^d : x_i = \delta \min_{l \in C} x_l, \ x_j \leqslant \delta \min_{l \in C} x_l \ \forall j \in (D \backslash C) \backslash i, \ x_k \geqslant 1 \ \forall k \in C\},$$

and we specifically note the equality $B_{C,\delta}^{x,1} = B_{C,\delta}^1$.

Proposition 12. *Assume the conditions of Proposition 9. If* $\mathbb{P}(\min_{i \in C} X_{P,i} > t, \max_{j \in D \setminus C} X_{P,j} < t^\delta) \in$ $\mathrm{RV}^\infty_{-1/\tau_C(\delta)}$, *and* $\mathbb{P}(\min_{i \in C} X_{P,i} > t, \max_{j \in D \setminus C} X_{P,j} < (\min_{i \in C} X_{P,i})^\delta) \in \mathrm{RV}^\infty_{-1/\tilde{\tau}_C(\delta)}$, *then* $\tilde{\tau}_C(\delta) = \tau_C(\delta)$.

Proof. Define $r^x_{C,\delta} = \min\left\{ r \in [0,1] : rR^x_{C,\delta} \cap G = \varnothing \right\}$, where for $r > 0$, $rR^x_{C,\delta} = \{\boldsymbol{x} \in \mathbb{R}^d_+ : x_i > r, i \in C, \ x_j \leqslant \delta \min_{l \in C} x_l, j \in D \setminus C\}$. Similarly to Propositions 7 and 9,

$$-\log \mathbb{P}(\boldsymbol{X}_E \in tr^x_{C,\delta} R^x_{C,\delta}) = -\log \mathbb{P}(\min_{i \in C} X_{E,i} > tr^x_{C,\delta}, \max_{j \in D \setminus C} X_{E,j} < \delta \min_{i \in C} X_{E,i}) \sim t,$$

and we conclude $\tilde{\tau}_C(\delta) = r^x_{C,\delta}$. As in Proposition 11, we have $r^x_{C,\delta} = \min_{\boldsymbol{y} \in B^x_{C,\delta}} g(\boldsymbol{y})$. Noting again that $\arg\min_{\boldsymbol{y} \in B^x_{C,\delta}} g(\boldsymbol{y}) \in B^{x,1}_{C,\delta} = B^1_{C,\delta}$ shows that $r^x_{C,\delta} = r_{C,\delta} = \tau_C(\delta)$. □

4 Examples

We illustrate several of the findings of Section 3 with some concrete examples. In Section 4.1 we focus on the intuitive and geometrically simple case $d = 2$; in Section 4.2, we examine some three-dimensional examples for which visualization is still possible but more intricate. Additional examples are given in the arXiv version of this article.

Proposition 2 implies that on $\mathbb{R}^d_+$, the same limit set G as in exponential margins will arise for any marginal choice with $\psi_j(x) \sim x$, $x \to \infty$, provided $e^{-\psi_j(x)}$ is a von Mises function. In some of the examples below, it is convenient to establish a limit set and its gauge function using this observation rather than transforming to exactly exponential margins.

Models with convenient dependence properties are often constructed through judicious combinations of random vectors with known dependence structures; see, for example, Engelke et al. (2019) for a detailed study of so-called random scale or random location constructions. In Section 4.3, we use our results to elucidate the shape of the limit set when independent exponential-tailed variables are mixed additively. The spatial dependence model of Huser and Wadsworth (2019) provides a case study.

4.1 Examples and illustrations for $d = 2$

All of the examples considered in this section are symmetric, so, for the conditional extremes representation and coefficients $\tau_j(\delta)$, we only consider one case, omitting the subscript on the quantities α_j and β_j. Table 1 summarizes the dependence information from various bivariate distributions described in Sections 4.1.1–4.1.4, as well as the arXiv version.

4.1.1 Meta-Gaussian distribution: nonnegative correlation

Starting with a Gaussian bivariate random vector and transforming its margins to standard exponential, we obtain a meta-Gaussian distribution with exponential margins. Such a distribution inherits the copula

Copula	$g(x,y)$	$\lambda(\omega, 1-\omega)$	$\eta_{1,2}$	$\tau_1(\delta)$	α, β
Gaussian $\rho \geqslant 0$	$\frac{x+y-2\rho(xy)^{1/2}}{1-\rho^2}$	$\begin{cases} \max(\omega, 1-\omega), & \frac{\min(\omega,1-\omega)}{\max(\omega,1-\omega)} \leqslant \rho^2 \\ \frac{1-2\rho(\omega(1-\omega))^{1/2}}{1-\rho^2}, & \frac{\min(\omega,1-\omega)}{\max(\omega,1-\omega)} \geqslant \rho^2 \end{cases}$	$\frac{1+\rho}{2}$	$\begin{cases} 1 & \delta \geqslant \rho^2 \\ \frac{1-\rho^2}{1+\delta-2\rho\delta^{1/2}} & \delta < \rho^2 \end{cases}$	$\alpha = \rho^2$ $\beta = 1/2$
Logistic GP	$\frac{1}{\theta}\max(x,y) + (1-\frac{1}{\theta})\min(x,y)$	$\max(\omega, 1-\omega)$	1	$[\theta^{-1} + (1-\theta^{-1}\delta)]^{-1}$	$\alpha = 1$ $\beta = 0$
Inverted logistic	$(x^{1/\theta} + y^{1/\theta})^\theta$	$g(\omega, 1-\omega)$	$2^{-\theta}$	1	$\alpha = 0$ $\beta = 1-\theta$
Hüsler–Reiss GP	$\begin{cases} \infty, & x \neq y \\ x, & x = y \end{cases}$	$\max(\omega, 1-\omega)$	1	$\begin{cases} 0, & \delta < 1 \\ 1, & \delta = 1 \end{cases}$	$\alpha = 1$ β undetermined
Inverted Hüsler–Reiss	$x\Phi\left(\frac{\lambda}{2} + \frac{\log\left(\frac{x}{y}\right)}{\lambda}\right) + y\Phi\left(\frac{\lambda}{2} + \frac{\log\left(\frac{y}{x}\right)}{\lambda}\right)$	$g(\omega, 1-\omega)$	$\left[2\Phi\left(\frac{\lambda}{2}\right)\right]^{-1}$	1	$\alpha = 0$ $\beta = 1$

Table 1: Summary of dependence measures across a range of bivariate examples

of the Gaussian distribution. For simplicity, we consider the case where the underlying Gaussian random vector has standard normal components with correlation ρ.

Then, for $\rho \geqslant 0$, the joint probability density f_E satisfies:

$$-\log f_E(tx, ty)/t = (x + y - 2\rho(xy)^{1/2})/(1 - \rho^2) + O(\log t/t), \quad t \to \infty, \quad x, y \geqslant 0,$$

so that $g(x, y) = (x + y - 2\rho(xy)^{1/2})/(1 - \rho^2)$. The convergence in (2.10) holds on $[0, \infty)^d$ and hence the limit set exists and is given by $\{\boldsymbol{x} \in [0, \infty)^d : g(\boldsymbol{x}) \leqslant 1\}$. This is example (ii) in Figure 2.

Conditional extremes: Setting $g(\alpha, 1) = 1$ leads to $(\alpha^{1/2} - \rho)^2 = 0$, i.e., $\alpha = \rho^2$. For β we have $g(\rho^2 + u, 1) - 1 = u^2/\{2\rho(1 - \rho^2)\} + O(u^3) \in \text{RV}_2^0$, hence $\beta = 1/2$.

Function $\lambda(\boldsymbol{\omega})$:   By Proposition 8, we need to find $1/r_\omega = \min_{\boldsymbol{y} \in B_\omega} g(x, y)$. If $\min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \leqslant \rho^2$, then $\min_{\boldsymbol{y} \in B_\omega} g(x, y) = 1$, with the minima occuring at the points $(1, \rho^2)$, $(\rho^2, 1)$. Otherwise, if $\min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \geqslant \rho^2$, then $\min_{\boldsymbol{y} \in B_\omega} g(x, y) = g(1, \min(\omega, 1 - \omega)/\max(\omega, 1 - \omega))$. Putting this together with Proposition 7, we find

$$\lambda(\omega, 1 - \omega) = \begin{cases} \max(\omega, 1 - \omega), & \min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \leqslant \rho^2 \\ g(\omega, 1 - \omega) = \frac{1 - 2\rho(\omega(1-\omega))^{1/2}}{1 - \rho^2}, & \min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \geqslant \rho^2. \end{cases}$$

This is the same form as given in Wadsworth and Tawn (2013). We therefore have $\eta_{1,2} = [2g(1/2, 1/2)]^{-1} = g(1, 1)^{-1} = (1 + \rho)/2$.

Coefficients $\tau_j(\delta)$:   From Proposition 10, we have $\tau_1(\delta) = [\min_{\gamma \in [0, \delta]} g(1, \gamma)]^{-1} = [g(1, \min(\delta, \rho^2))]^{-1}$. Therefore, $\tau_1(\delta) = 1$ if $\delta \geqslant \rho^2$, else $\tau_1(\delta) = (1 - \rho^2)/(1 + \delta - 2\rho\delta^{1/2}) < 1$. Note that these values are very laborious to calculate via Gaussian survival functions, and they were not given in Simpson et al. (2020).

4.1.2   Meta-Gaussian distribution: negative correlation

When $\rho < 0$, Proposition 2 cannot be applied as the continuous convergence condition (2.10) does not hold along the axes. Hence, we only gain a partial specification, when $x > 0, y > 0$, through this route. Instead, here we can apply Proposition 3 since the limit function g in (2.13) satisfies the monotonicity condition given immediately thereafter. This limit function is given by

$$g(x, y) = \begin{cases} (x + y - 2\rho(xy)^{1/2})/(1 - \rho^2), & x > 0, \ y > 0, \\ x, & y = 0, \\ y, & x = 0. \end{cases}$$

Figure 5 illustrates the limit sets G for the three cases $\rho > 0, \rho = 0$ and $\rho < 0$. In the latter case, large values of one variable tend to occur with small values of the other, which causes the limit set to
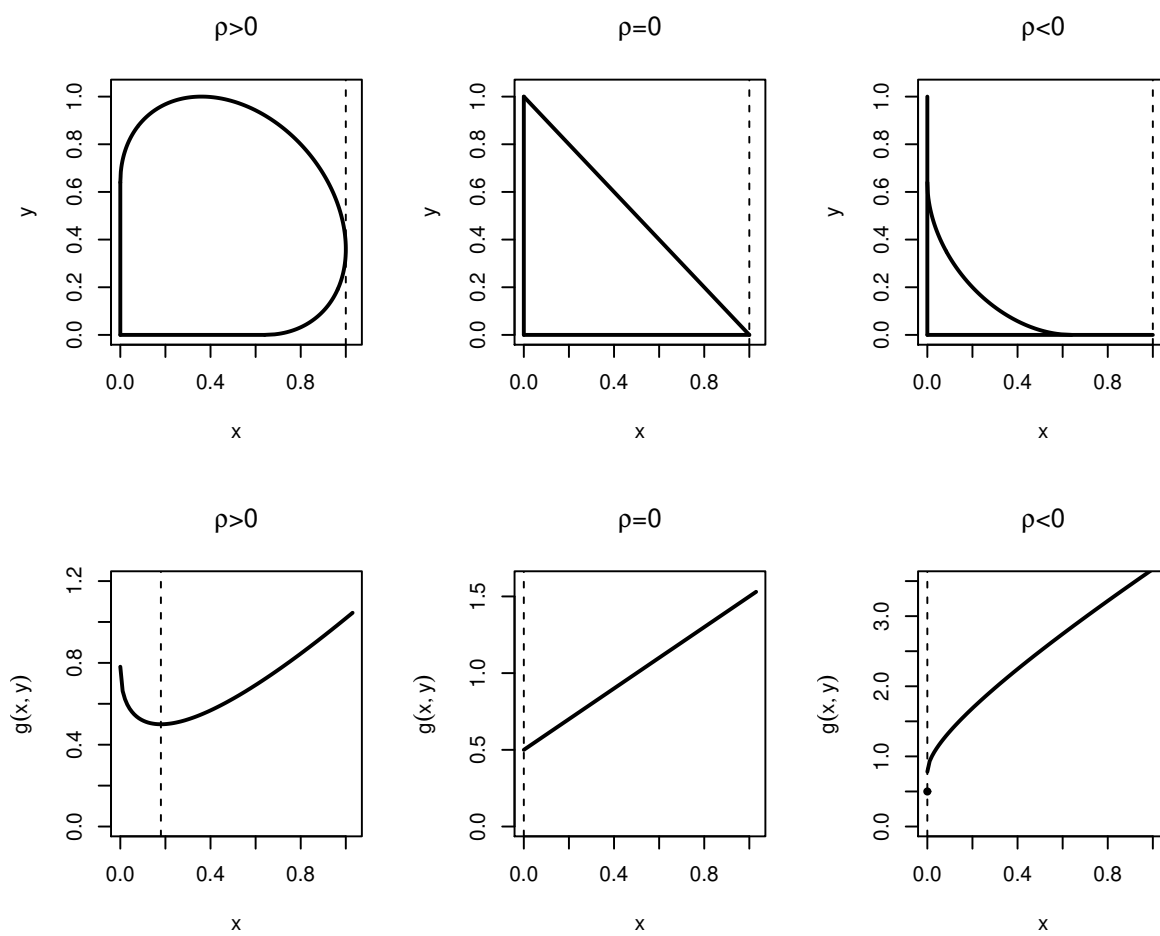
Figure 5: Top row: Limit sets $\{(x, y) : g(x, y) \leqslant 1\}$ for a bivariate meta-Gaussian distribution with exponential margins. Bottom row: corresponding plots of function $g(x, y)$ for a fixed value of y.

include lines along the axes and the function g is not continuous. Such difficulties can be alleviated by consideration of Laplace margins for distributions displaying negative dependence, which is discussed further in Section 5.

4.1.3 Logistic generalized Pareto copula

The logistic generalized Pareto distribution with conditionally exponential margins ($\mathbb{P}(X_{\tilde{E}} > x) = \mathbb{P}(X_{\tilde{E}} > 0)e^{-x}$, $x > 0$) and dependence parameter $\theta \in (0, 1)$ satisfies

$$f_{\tilde{E}}(x, y) = \theta^{-1} 2^{-\theta} e^{-(x+y)/\theta} (e^{-x/\theta} + e^{-y/\theta})^{\theta-2}$$

$$-\log f_{\tilde{E}}(tx, ty)/t = \theta^{-1} \max(x, y) + (1 - \theta^{-1}) \min(x, y) + O(1/t),$$

so the gauge function is $g(x, y) = \theta^{-1} \max(x, y) + (1 - \theta^{-1}) \min(x, y)$. This form of gauge function is found throughout several symmetric asymptotically dependent examples, such as those distributions whose spectral measure H places no mass on 0 and 1 and possess densities that are regularly varying at the endpoints 0, 1 such that $dH(w)/dw \in \mathrm{RV}^0_{1/\theta-2}$, $-dH(1-w)/dw \in \mathrm{RV}^0_{1/\theta-2}$. This is example (i) in Figure 2.

Conditional extremes: Solving for $g(\alpha, 1) = 1$, we obtain $\alpha = 1$, whilst $g(1 + u, 1) - 1 = u/\theta \in \mathrm{RV}^0_1$, hence $\beta = 0$.

Function $\lambda(\boldsymbol{\omega})$: We have that $\arg\min_{\boldsymbol{y} \in B_\omega} g(\boldsymbol{y}) = (1, 1)$, so $r_\omega = 1$ and $\lambda(\omega, 1 - \omega) = \max(\omega, 1 - \omega)$. Therefore $\eta_{1,2} = 1$.

Coefficients $\tau_j(\delta)$: $\tau_1(\delta) = [\min_{\gamma \in [0,\delta]} g(1, \gamma)]^{-1} = [g(1, \delta)]^{-1} = [\theta^{-1} + (1 - \theta^{-1}\delta)]^{-1}$. This matches the value calculated in the Supplementary Material of Simpson et al. (2020).

4.1.4 Inverted extreme value distribution

The inverted extreme value copula is the joint lower tail of an extreme value copula, translated to be the joint upper tail. That is, if (U_1, U_2) have an extreme value copula with uniform margins, then $(1 - U_1, 1 - U_2)$ have an inverted extreme value copula. In two dimensions, its density in exponential margins may be expressed as

$$f_E(x, y) = \{l_1(x, y)l_2(x, y) - l_{12}(x, y)\} \exp\{-l(x, y)\},$$

where $l(\boldsymbol{x}) = V(1/\boldsymbol{x})$, for V the exponent function in (2.3), is the 1-homogeneous stable tail dependence function (e.g., Beirlant et al., 2004, Ch.8) of the corresponding extreme value distribution, and $l_1(x, y) = \partial l(x, y)/\partial x$, etc. We thus have

$$-\log f_E(tx, ty)/t = l(x, y) + O(1/t), \qquad t \to \infty,$$

so $g(x, y) = l(x, y)$.

Conditional extremes: Stable tail dependence functions always satisfy $l(x, 0) = x, l(0, y) = y$ and so $g(1, 0) = g(0, 1) = 1$. Hence, if $\alpha = 0$ is the only solution to $g(\alpha, 1) = 1$, then $a(x)/x \sim 0$. An example of this is given by the inverted extreme value logistic copula, whereby $l(x, y) = (x^{1/\theta} + y^{1/\theta})^\theta, \theta \in (0, 1]$. This is example (iii) of Figure 2, for which we have $\alpha = 0$ and $\beta = 1 - \theta$.

Further examples in this class are given in the arXiv version.

Function $\lambda(\boldsymbol{\omega})$: Since $g(x, y) = l(x, y)$, and l is a convex function satisfying $l(x, 0) = x, l(0, y) = y$, $\arg\min_{\boldsymbol{y} \in B_\omega} g(\boldsymbol{y}) = (\omega, 1 - \omega)/\max(\omega, 1 - \omega)$. Hence, $\lambda(\omega, 1 - \omega) = g(\omega, 1 - \omega)$ in this case.

Coefficients $\tau_j(\delta)$: Since $g(1, 0) = 1$, we have $\tau_1(\delta) = 1$ for all $\delta \in [0, 1]$.

4.1.5 Density defined by g

If $g : \mathbb{R}_+^d \to \mathbb{R}_+$ is a gauge function describing a limit set G, then $f(\boldsymbol{x}) = e^{-g(\boldsymbol{x})}/(d!|G|)$ is a density (see Balkema and Nolde, 2010). In general, except for the case of $g(\boldsymbol{x}) = \sum_{i=1}^d x_i$, the margins are not exactly exponential, and may be heavier than exponential, for example in the case $g(\boldsymbol{x}) = \max_{1 \leqslant i \leqslant d}(x_i)$.

We consider the density defined by $g(x, y) = \max\{(x - y)/\theta, (y - x)/\theta, (x + y)/(2 - \theta)\}, \theta \in (0, 1]$: this is example (vi) in Figure 2, and illustrated in Figure 4. The marginal density is given by

$$[2e^{-x} - \theta e^{-x/\theta} - 2(1 - \theta)e^{-x/(1-\theta)}]/[4\theta - 3\theta^2].$$

Conditional extremes: Solving for $g(\alpha, 1) = 1$, we obtain $\alpha = 1 - \theta$, whilst $g(1 - \theta + u, 1) - 1 = u/(2 - \theta) \in \mathrm{RV}_1^0$, hence $\beta = 0$.

Function $\lambda(\boldsymbol{\omega})$: If $\min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \leqslant 1 - \theta$, then $\arg\min_{\boldsymbol{y} \in B_\omega} = (1, 1 - \theta)$, or $(1 - \theta, 1)$ and $r_\omega = 1$; otherwise, $\arg\min_{\boldsymbol{y} \in B_\omega} = (1, \omega/(1 - \omega))$ or $((1 - \omega)/\omega, 1)$, and $r_\omega = \{1 + \min(\omega, 1 - \omega)/\max(\omega, 1 - \omega)/(2 - \theta)\}$. As such

$$\lambda(\omega, 1 - \omega) = \begin{cases} \max(\omega, 1 - \omega), & \min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \leqslant 1 - \theta \\ g(\omega, 1 - \omega) = \frac{1}{2-\theta}, & \min(\omega, 1 - \omega)/\max(\omega, 1 - \omega) \geqslant 1 - \theta, \end{cases}$$

and the residual tail dependence coefficient $\eta_{1,2} = 1 - \theta/2$.

Coefficients $\tau_j(\delta)$: $\tau_1(\delta) = [\min_{\gamma \in [0,\delta]} g(1, \gamma)]^{-1} = [g(1, \min(\delta, 1 - \theta))]^{-1}$. Therefore, $\tau_1(\delta) = 1$ if $\delta \geqslant 1 - \theta$, else $\tau_1(\delta) = \theta/(1 - \delta) < 1$.

Further examples presented in the arXiv version include the Hüsler–Reiss generalized Pareto copula, and two boundary cases with $g(x, y) = \max(x, y)$, one displaying asymptotic dependence and one asymptotic independence. In the latter of these, we find that there is no conditional extreme value limit with positive support, but there is one with negative support. We comment that the results of Proposition 5 do indeed focus predominantly on the positive end of the support for limit distributions, but most known examples of conditional limits have support including $(0, \infty)$. A natural next step is to consider the implications relating to negative support. We particularly note the possibility that the order of regular variation of the two functions $g(1, \alpha_1 + u) - 1 \in \mathrm{RV}^0_{1/(1-\beta_1^+)}$ and $g(1, \alpha_1 - u) - 1 \in \mathrm{RV}^0_{1/(1-\beta_1^-)}$ need not be equal, though for each of our examples where both functions are regularly varying, $\beta^+ = \beta^-$. If $\beta^+ > \beta^-$, it seems likely that a limit distribution with positive support only would arise, and vice versa when $\beta^+ < \beta^-$.

4.2   Examples and illustrations for $d = 3$

In this section we give two examples, focusing on issues that arise for $d > 2$.

4.2.1   Gaussian copula

The general form of the gauge function for a meta-Gaussian distribution with standard exponential margins and correlation matrix Σ with non-negative entries is

$$g(\boldsymbol{x}) = (\boldsymbol{x}^{1/2})^\top \Sigma^{-1} \boldsymbol{x}^{1/2}.$$

Figure 6 displays the level set $g(\boldsymbol{x}) = 1$ when the Gaussian correlations in Σ are $\rho_{12} = 0.75, \rho_{13} = 0.25, \rho_{23} = 0.4$. The red dots on the level set are the points $(1, 1, \gamma)/g(1, 1, \gamma)$, $(1, \gamma, 1)/g(1, \gamma, 1)$ and $(\gamma, 1, 1)/g(\gamma, 1, 1)$ for $\gamma \in [0, 1]$. The figure also provides an illustration of $\tau_{2,3}(\delta)$ for $\delta = 0.2$ and $\delta = 0.8$: in each case the light blue line from the origin is $\gamma \times (\delta, 1, 1)$, $\gamma \in [0, 1]$, whilst the pink lines trace out the boundary $B_{\{2,3\},\delta}$ and $\tau_{2,3}(\delta)B_{\{2,3\},\delta}$. We see that when $\delta = 0.2$ (left panel), $\tau_{2,3}(0.2) = 1/g(0.2, 1, 1)$, i.e., $\min_{\boldsymbol{y} \in B_{\{2,3\},0.2}} g(\boldsymbol{y}) = g(0.2, 1, 1)$. However, when $\delta = 0.8$, $\min_{\boldsymbol{y} \in B_{\{2,3\},0.8}} g(\boldsymbol{y}) = g(\gamma^\star, 1, 1)$, for $\gamma^\star \in [0, 0.8]$, so $\tau_{2,3}(0.8) = 1/g(\gamma^\star, 1, 1)$. We note that the same value of $\tau_{2,3}(\delta)$ applies for any $\delta \geqslant \gamma^\star$: for this example, when $\delta \geqslant \gamma^\star \approx 0.51$, $\tau_{2,3}(\delta) = 0.7 = \eta_{2,3}$.

The reason that $\tau_{2,3}(\delta) = \eta_{2,3}$ for sufficiently large δ is because in this case $\arg\min_{x_1} g(x_1, 1, 1) = \gamma^\star$, meaning that the two-dimensional marginalization $g_{\{2,3\}}(1, 1) = g(\gamma^\star, 1, 1)$, and we further have that $g_{\{2,3\}}(1, 1) = \min_{\boldsymbol{y} \in B_{2,3}} g_{\{2,3\}}(\boldsymbol{y})$, so $\eta_{2,3} = 1/g_{2,3}(1, 1)$. In Section 4.2.2 we will illustrate a gauge function for which $\arg\min_{x_3} g(1, 1, x_3) > 1$, and consequently $\tau_{1,2}(\delta) < \eta_{1,2}$ for all $\delta \leqslant 1$.

The right panel of Figure 6 illustrates $\tau_1(\delta)$ for $\delta = 0.2$ and $\delta = 0.6$. When $\delta = 0.6$, the boundary $B_{1,\delta}$ already touches G, and so $\tau_1(0.6) = 1$. In this example, $\tau_1(\delta) = 1$ for any $\delta \geqslant 0.5625 = \rho_{12}^2$. As such, $\tau_1(0.2) < 1$ as illustrated in the figure. We comment that if we had marginalized over X_2, and were
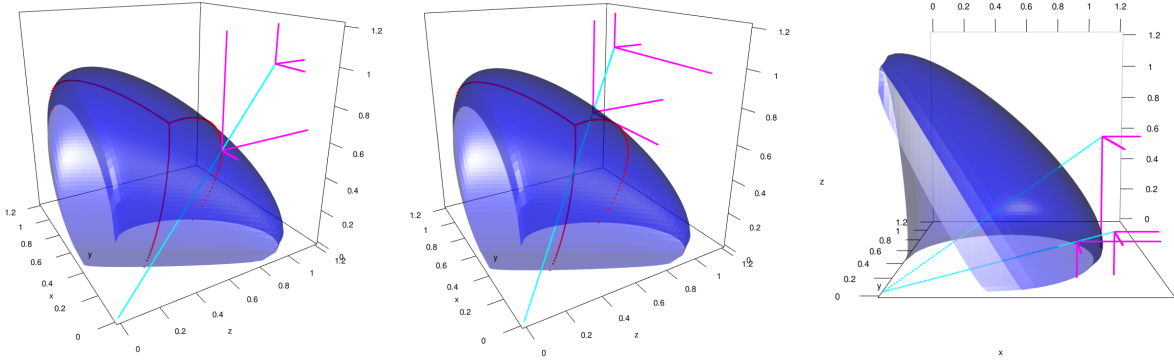
Figure 6: Level set $g(\boldsymbol{x}) = 1$ for a trivariate meta-Gaussian distribution with exponential margins. The left panel illustrates $\tau_{2,3}(0.2)$: the boundary set indicated by the pink lines is scaled along the blue trajectory until it touches G, which happens in this case at the corner point $(0.2, 1, 1)/g(0.2, 1, 1)$. The centre panel illustrates $\tau_{2,3}(0.8)$: the boundary set is again pulled back along the indicated trajectory until it touches G: in this case this does not occur at a corner point. The right panel illustrates $\tau_1(\delta)$ in a similar manner, for $\delta = 0.2, 0.6$.

looking at $\tau_1(\delta)$ for the variables (X_1, X_3), then we would have $\tau_1(\delta) = 1$ for any $\delta \geqslant 0.0625 = \rho_{13}^2$. This provides an illustration of the dimensionality of the problem interacting with $\tau_C(\delta)$, and is again related to the point at which the minimum point defining the lower-dimensional gauge function occurs.

4.2.2 Vine copula

Three-dimensional vine copulas are specified by three bivariate copulas: two in the "base layer", giving the dependence between, e.g., X_1, X_2 and X_2, X_3 and a further copula specifying the dependence between $X_1|X_2$ and $X_3|X_2$. Here we take the base copulas to be independence for (X_1, X_2), and the inverted Clayton copula with parameter $\beta > 0$ for (X_2, X_3). The final copula is taken as inverted Clayton with parameter $\gamma > 0$. The gauge function that arises in exponential margins is

$$g(\boldsymbol{x}) = (1 + \beta) \max(x_2, x_3) - \beta \min(x_2, x_3) - \gamma x_1 - (\gamma + 1)(\beta + 1)(\max(x_2, x_3) - x_2)$$
$$+ (2\gamma + 1) \max(x_1, (\beta + 1)(\max(x_2, x_3) - x_2)). \qquad (4.20)$$

Figure 7 displays the level set $g(\boldsymbol{x}) = 1$. In this figure we also give an illustration of a case where $\tau_C(1) < \eta_C$: in particular, for this example $\tau_{1,2}(1) < \eta_{1,2,3} = \eta_{1,2} = 1/2$. The purple lines represent the boundary of the region $\tau_{1,2}(1)R_{\{1,2\},1} = \tau_{1,2}(1, \infty]^2 \times [0, 1]$, while the green lines represent the boundary of the region $\eta_{1,2,3}(1, \infty]^3$. Theorem 1 of Simpson et al. (2020) tells us that $\eta_{1,2} = \max(\tau_{1,2}(1), \tau_{1,2,3})$, where $\tau_{1,2,3} = \eta_{1,2,3}$. Therefore $\tau_{1,2}(1) < \eta_{1,2}$ guarantees that $\eta_{1,2} = \eta_{1,2,3}$.

We also illustrate Proposition 4, minimizing (4.20) over x_3. If $x_2 > x_3$ then the minimum over x_3 occurs by setting $x_3 = x_2$ and is equal to $x_2 + (1+\gamma)x_1$. If $x_2 < x_3$ then owing to the final term we need to
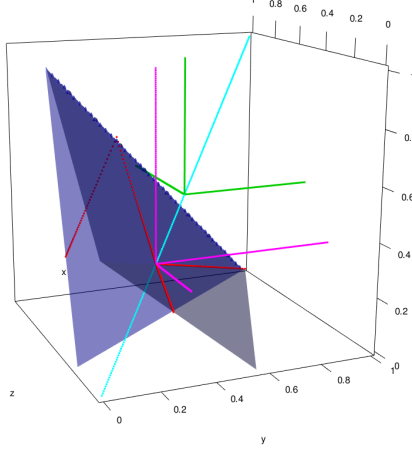
Figure 7: Level set $g(\boldsymbol{x}) = 1$ with g as in (4.20). The figure illustrates $\tau_{1,2}(1)$ and $\eta_{1,2,3} = \eta_{1,2} = 1/2$.

consider the cases $x_3 \lesseqgtr x_1/(1+\beta) + x_2$. In both cases, the mimimum is attained at $x_3 = x_1/(1+\beta) + x_2$, and is equal to $x_1 + x_2 < (1+\gamma)x_1 + x_2$. As such, $\min_{x_3} g(x_1, x_2, x_3) = x_1 + x_2$. This result is as expected since the bivariate margins of vine copulas that are directly specified in the base layer are equal to the specified copula: in this case, independence.

4.3   Mixing independent vectors

Here we exploit the results from previous sections to consider what happens when independent exponential random vectors are additively mixed such that the resulting vector still has exponential type tails. We consider as a case study the spatial model of Huser and Wadsworth (2019), which following a reparameterization can be expressed

$$\{X_{\tilde{E}}(s) = \gamma S_E + V_E(s) : s \in \mathcal{S} \subset \mathbb{R}^2\}, \qquad \gamma \in (0, \infty), \tag{4.21}$$

where $S_E \sim \text{Exp}(1)$ is independent of the spatial process V_E, which also possesses unit exponential margins and is asymptotically independent at all spatial lags $s_1 - s_2 \neq 0$. The process V_E is assumed to possess hidden regular variation, with residual tail dependence coefficient satisfying $\eta^V(s_1, s_2) < 1$ for all $s_1 \neq s_2$. The resulting process $X_{\tilde{E}}$ is asymptotically independent for $\gamma \in (0, 1]$ and asymptotically dependent for $\gamma > 1$; see also Engelke et al. (2019) for related results.

When $\gamma < 1$, $\mathbb{P}(X_{\tilde{E}}(s) > x) \sim e^{-x}/(1-\gamma)$. In this case, Huser and Wadsworth (2019) show that the

residual tail dependence coefficient for the process $X_{\tilde{E}}$ is given by

$$\eta^X = \begin{cases} \eta^V, & \gamma < \eta^V \\ \gamma, & \eta^V \leqslant \gamma \leqslant 1. \end{cases} \tag{4.22}$$

That is, the strength of the extremal dependence as measured by the residual tail dependence coefficient η^X is increasing in γ for $\gamma \geqslant \eta^V$. In contrast, Wadsworth and Tawn (2019) showed that under mild conditions, the process (4.21) has the same conditional extremes normalization as the process $V_E(s)$, with identical limit distribution when the scale normalizations $b_{s-s_0}(t) \to \infty$ as $t \to \infty$. Here, the subscript $s - s_0$ alludes to the fact that the conditioning event in (2.8) is $\{V_E(s_0) > t\}$ and we study the normalization at some other arbitrary location $s \in \mathcal{S}$. In combination, we see that the results of Huser and Wadsworth (2019) and Wadsworth and Tawn (2019) suggest that the addition of the variable γS_E to V_E affects the extremal dependence of $X_{\tilde{E}}$ differently for different extreme value representations. We elucidate these results further in the context of the limit sets and their gauge functions. A summary is provided here, with full derivations in the arXiv version.

Let us suppose that $\boldsymbol{S}_E \in \mathbb{R}_+^d$ has unit exponential margins, density f_{S_E}, and gauge function g_S, and is independent of $\boldsymbol{V}_E \in \mathbb{R}_+^d$, which has unit exponential margins, density f_{V_E}, and gauge function g_V. Let $\boldsymbol{Z}_E = (\boldsymbol{S}_E, \boldsymbol{V}_E) \in \mathbb{R}_+^{2d}$ be the concatenation of these vectors; this has exponential margins and gauge function $g_Z(\boldsymbol{z}) = g_S(z_1, \ldots, z_d) + g_V(z_{d+1}, \ldots, z_{2d})$.

Now consider the linear transformation of $\boldsymbol{Z}_E$ to

$$A\boldsymbol{Z}_E = (\gamma Z_{E,1} + Z_{E,d+1}, \ldots, \gamma Z_{E,d} + Z_{E,2d}, Z_{E,1}, \ldots, Z_{E,d}) = (\gamma \boldsymbol{S}_E + \boldsymbol{V}_E, \boldsymbol{S}_E) = (\boldsymbol{X}_{\tilde{E}}, \boldsymbol{S}_E),$$

where $A \in \mathbb{R}^{2d \times 2d}$ is the matrix describing this transformation. By Lemma 4.1 of Nolde (2014), the normalized sample cloud $\{A\boldsymbol{Z}_{E,i}/\log n : i = 1, \ldots, n\}$ converges onto the set AG, where $G = \{\boldsymbol{z} \in \mathbb{R}_+^{2d} : g_Z(\boldsymbol{z}) \leqslant 1\}$, so $AG = \{\boldsymbol{z} \in \mathbb{R}_+^{2d} : A^{-1}\boldsymbol{z} \in G\} = \{\boldsymbol{z} \in \mathbb{R}_+^{2d} : g_Z(A^{-1}\boldsymbol{z}) \leqslant 1\}$. Consequently, the gauge function of $A\boldsymbol{Z}_E$ is $g_Z(A^{-1}\boldsymbol{z})$, i.e., $g_Z(\boldsymbol{x}, \boldsymbol{s}) = g_S(\boldsymbol{s}) + g_V(\boldsymbol{x} - \gamma\boldsymbol{s})$, for $\boldsymbol{x} > \gamma\boldsymbol{s}$.

Next we apply Proposition 4 to the vector $A\boldsymbol{Z}$, marginalizing over the last d coordinates, which are equal to $\boldsymbol{S}_E$. This leaves us with the gauge function of $X_{\tilde{E}}$, denoted g_X, and given by

$$g_X(\boldsymbol{x}) = \min_{\boldsymbol{s} \in [\boldsymbol{0}, \boldsymbol{x}/\gamma]} g_S(s_1, \ldots, s_d) + g_V(x_1 - \gamma s_1, \ldots, x_d - \gamma s_d).$$

To illustrate the results of Huser and Wadsworth (2019) and Wadsworth and Tawn (2019) concerning model (4.21), we need to take $\boldsymbol{S}_E = S_E\boldsymbol{1}$, i.e., perfect dependence. Although such a vector does not have a d-dimensional Lebesgue density, convergence of the sample cloud based on the univariate random variable S_E onto the unit interval $[0, 1]$ implies that the limit set is $G_S = \{\boldsymbol{x} \in \mathbb{R}_+^d : x_1 = x_2 = \cdots = x_d = x, x \leqslant 1\}$. Such a set can be described by the gauge function

$$g_S(\boldsymbol{s}) = \begin{cases} \infty, & s_i \neq s_j \;\; \text{for any} \;\; i, j \\ s, & s_1 = \cdots = s_d = s. \end{cases}$$

As such, in this case, $g_X(\boldsymbol{x}) = \min_{s \in [0, \min(\boldsymbol{x})/\gamma]}\{s + g_V(\boldsymbol{x} - \gamma s)\}$.

Residual tail dependence   To find the residual tail dependence coefficient η^X, we require

$$\min_{\boldsymbol{x}:\min(\boldsymbol{x})=1} g_X(\boldsymbol{x}) = \min_{\boldsymbol{x}:\min(\boldsymbol{x})=1} \min_{s \in [0, \min(\boldsymbol{x})/\gamma]}\{s + g_V(\boldsymbol{x} - \gamma s)\}$$

$$= \min_{s \in [0, 1/\gamma]} \min_{\boldsymbol{x}:\min(\boldsymbol{x})=1}\{s + g_V(\boldsymbol{x} - \gamma s)\}.$$

For fixed s, consider $\min_{\boldsymbol{x}:\min(\boldsymbol{x})=1} g_V(\boldsymbol{x} - \gamma s) = \min_{\boldsymbol{z}:\min(\boldsymbol{z})=1-\gamma s} g_V(\boldsymbol{z}) = g_V(\boldsymbol{y}^\star \times (1 - \gamma s))$, where $\boldsymbol{y}^\star = \arg\min_{\boldsymbol{y}:\min(\boldsymbol{y})=1} g_V(\boldsymbol{y})$. As such

$$\min_{\boldsymbol{x}:\min(\boldsymbol{x})=1} g_X(\boldsymbol{x}) = \min_{s \in [0, 1/\gamma]}\{s + g_V(\boldsymbol{y}^\star)(1 - \gamma s)\} = \begin{cases} g_V(\boldsymbol{y}^\star), & \gamma < 1/g_V(\boldsymbol{y}^\star), \\ 1/\gamma, & \gamma \geqslant 1/g_V(\boldsymbol{y}^\star). \end{cases}$$

Recalling that $\eta^X = [\min_{\boldsymbol{x}:\min(\boldsymbol{x})=1} g_X(\boldsymbol{x})]^{-1}$ and $\eta^V = 1/g_V(\boldsymbol{y}^\star)$, this yields (4.22).

Conditional extremes   For the conditional extremes normalization, we now let g_V and g_X denote two-dimensional gauge functions. Suppose that α_V, β_V are such that $g_V(\alpha_V, 1) = 1$ and $g_V(\alpha_V + u, 1) - 1 \in \mathrm{RV}^0_{1/(1-\beta_V)}$. We have

$$1 = g_X(\alpha_X, 1) = \min_s\{s + g_V(\alpha_X - \gamma s, 1 - \gamma s)\}. \tag{4.23}$$

Suppose that the right hand side of (4.23) is minimized at $s^\star \geqslant 0$, i.e., $g_X(\alpha_X, 1) = s^\star + g_V(\alpha_X - \gamma s^\star, 1 - \gamma s^\star)$. Because $\alpha_X \leqslant 1$ and $g_V(v_1, v_2) \geqslant \max(v_1, v_2)$, this yields $1 = g_X(\alpha_X, 1) \geqslant 1 + (1 - \gamma)s^\star$, therefore we must have $s^\star = 0$ for $\gamma \in (0, 1)$. Consequently, $\alpha_X = \alpha_V = \alpha$.

Calculations for the scale normalization are more involved and can be found in the arXiv version. We find that for g_V differentiable at $(\alpha, 1)$, $g_X(\alpha + u, 1) - 1 \sim g_V(\alpha + u, 1) - 1$, $u \to 0$, whereas in the non-differentiable case we do not necessarily have this link but can deduce that the regular variation indices are $\beta_X = \beta_V = 0$.

Figure 8 displays examples of gauge functions g_V and g_X. We observe from this figure how, when γ is sufficiently large, the shape of g_V is modified to produce g_X. The modification is focussed around the diagonal, and explains visually why the residual tail dependence coefficient changes while the conditional extremes normalization does not. The left and right panels illustrate differentiable cases, and the centre panel non-differentiable.

5   Discussion

In this work we have demonstrated how several concepts of extremal dependence can be unified through the shape of the limit set G of the scaled sample cloud $N_n = \{\boldsymbol{X}_1/r_n, \ldots, \boldsymbol{X}_n/r_n\}$ arising for distributions
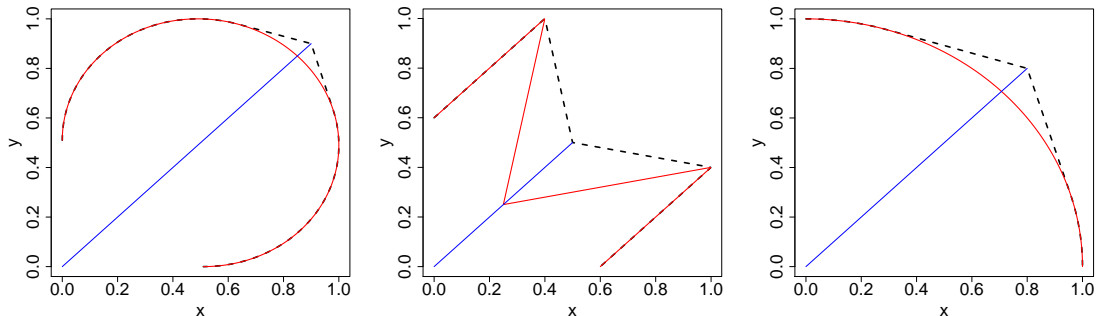
Figure 8: Solid red lines depict the level sets $g_V(\boldsymbol{x}) = 1$, where g_V is of the form (ii), (iii) and (iv) (L–R) from Figure 2. Dashed black lines depict the level sets $g_X(\boldsymbol{x}) = 1$. In each picture the blue solid line segment is from $(0,0)$ to (γ, γ), and denotes the limit set of the fully dependent random vector γS_E. From left to right, $\gamma = 0.9, 0.5, 0.8$.

with light-tailed margins. For concreteness our focus has been on exponential margins, but other choices can be useful. In the case of negative dependence between extremes — such that large values of one variable are most likely to occur with small values of another — the double exponential-tailed Laplace margins can be more enlightening. As an example, for the bivariate Gaussian copula with $\rho < 0$ we observed that the limit set G is described by a discontinuous gauge function g that cannot be established through the simple mechanism of Proposition 2. In Nolde (2014), the gauge function for this distribution in Laplace margins is calculated as

$$g(x, y) = \begin{cases} (|x| + |y| - 2\rho|xy|^{1/2})/(1 - \rho^2), & x, y \geqslant 0 \ \ \text{or} \ \ x, y \leqslant 0, \\ (|x| + |y| + 2\rho|xy|^{1/2})/(1 - \rho^2), & x \geqslant 0, y \leqslant 0 \ \ \text{or} \ \ x \leqslant 0, y \geqslant 0. \end{cases}$$

When $\rho < 0$, this yields $g(1, -\rho^2) = 1$, and $g(1, -\rho^2 + u) \in \mathrm{RV}_2^0$, so that extending Proposition 5, we would find that the conditional extremes normalizations are $a^j(t) \sim -\rho^2 t$ and $b^j(t) \in \mathrm{RV}_{1/2}^0$, as given in Keef et al. (2013).

The study of extremal dependence features through the limit set G is enlightening both for asymptotically dependent and asymptotically independent random vectors, particularly as it can be revealing for mixture structures where mass is placed on a variety of cones $\mathbb{E}_C$ as defined in (2.2). However, many traditional measures of dependence within the asymptotically dependent framework, which are typically functions of the exponent function V given in equation (2.3), or spectral measure H, are not revealed by limit set G. For example, it was noted in the example of Section 4.1.3 that the limit set described by the gauge function $g(x, y) = \theta^{-1} \max(x, y) + (1 - \theta^{-1}) \min(x, y)$ can arise for several different spectral measures, although clearly the parameter θ demonstrates some link between strength of dependence and shape of G.

Nonetheless, multivariate regular variation and associated limiting measures have been well-studied

in extreme value theory, but representations that allow greater discrimination between asymptotically independent or mixture structures much less so. The limit set elucidates many of these alternative dependence concepts and provides meaningful connections between them. We have not directly considered connections between the various dependence measures without reference to G, and we note that the limit set might not always exist. We leave such study to future work.

Acknowledgements

References

Abdous, B., Fougères, A.-L., and Ghoudi, K. (2005). Extreme behaviour for bivariate elliptical distributions. *Canadian Journal of Statistics*, 33(3):317–334.

Balkema, A.A., Embrechts, P., and Nolde, N. (2010). Meta densities and the shape of their sample clouds. *J. Multivariate Analysis*, 101:1738–1754.

Balkema, G. and Embrechts, P. (2007). *High Risk Scenarios and Extremes. A Geometric Approach.* European Mathematical Society, Zurich.

Balkema, G. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432.

Balkema, G. and Nolde, N. (2012). Asymptotic dependence for homothetic light-tailed densities. *Adv. Appl. Prob.*, 44.

Balkema, G. and Nolde, N. (2020). Samples with a limit shape, multivariate extremes, and risk. *Advances in Applied Probability*, 52:491–522.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268.

Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2004). *Statistics of extremes: theory and applications.* John Wiley & Sons.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):377–392.

Das, B. and Resnick, S. I. (2011). Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252.

Davis, R. A., Mulrow, E., and Resnick, S. I. (1988). Almost sure limit sets of random samples in $\mathbb{R}^d$. *Advances in applied probability*, 20(3):573–599.

de Haan, L. (1970). *On Regular Variation and Its Application to Weak Convergence of Sample Extremes*. CWI Tract 32, Amsterdam.

de Valk, C. (2016a). Approximation and estimation of very small probabilities of multivariate extreme events. *Extremes*, 19:687–717.

de Valk, C. (2016b). Approximation of high quantiles from intermediate quantiles. *Extremes*, 19:661–686.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes (with discussion). *Journal of the Royal Statistical Society, Series B*.

Engelke, S., Opitz, T., and Wadsworth, J. (2019). Extremal dependence of random scale constructions. *Extremes*, 22(4):623–666.

Fisher, L. (1969). Limiting sets and convex hulls of samples from product measures. *Ann. Math. Statist.*, 40:1824–1832.

Fougères, A.-L. and Soulier, P. (2010). Limit conditional distributions for bivariate vectors with polar representation. *Stochastic models*, 26(1):54–77.

Geffroy, J. (1958). Contribution à la theorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris*, VII:37–121.

Geffroy, J. (1959). Contribution à la theorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris*, VIII:3–65.

Gnedenko, B. (1943). Sur la distribution limité du terme d'une série aléatoire. *Ann. Math.*, 44:423–453.

Gumbel, E. J. (1960). Distributions des valeurs extremes en plusiers dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444.

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

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Kinoshita, K. and Resnick, S. I. (1991). Convergence of scaled random samples in $\mathbb{R}^d$. *The Annals of Probability*, pages 1640–1663.

Kulik, R. and Soulier, P. (2020). *Heavy-tailed time series*. Springer.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.

Matheron, G. (1975). *Random Sets and Integral Geometry*. John Wiley & Sons, New York.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):219–241.

Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Resnick, S. (2007). *Heavy-tail Phenomena. Probabilistic and Statistical Modeling*. Springer.

Seifert, M. I. (2014). On conditional extreme values of random vectors with polar representation. *Extremes*, 17(2):193.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Wadsworth, J. and Tawn, J. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5B):2689–2714.

Wadsworth, J. and Tawn, J. (2019). Higher-dimensional spatial extremes via single-site conditioning. *arXiv:1912.06560*.